An Energy-Efficient System with Timing-Reliable Error-Detection Sequentials

by

Yaoqiang 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, 2016

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

A new type of energy-efficient digital system that integrate Error Detection Sequential (EDS) and Dynamic Voltage Scaling (DVS) circuits has been developed [1, 2, 3, 4]. In these systems, EDS-monitored paths convert the Process, Voltage and Temperature (PVT) variations into timing variations. Nevertheless, the conversion can suffer from the reliability issue (extrinsic EDS-reliability). EDS circuits detect the unfavorable timing variations (so called "error") and guide DVS circuits to adjust the operating voltage to a proper lower level to save the energy. However, the error detection is generally susceptible to the metastability problem (intrinsic EDS-reliability) due to the synchronizer in EDS circuits. The Mean Time Between Failure (MTBF) due to metastability is exponentially related to the synchronizer delay.

This dissertation proposes a new EDS circuit deployment strategy to enhance the extrinsic EDS-reliability. This strategy requires neither buffer insertion nor an extra clock and is applicable for Field-Programmable Gate Array (FPGA) implementations. An FPGA-based Discrete Cosine Transform with EDS and DVS circuits deployed in this fashion demonstrates up to 16.5% energy savings over a conventional design at equivalent frequency setting and image quality, with a 0.8% logic element and 3.5% maximum frequency penalties.

Voltage-Boosted Synchronizers (VBSs) are proposed to improve the synchronizer delay under single low-voltage supply environments. A VBS consists of a Jamb latch and a switched-capacitor-based charge pump that provides a voltage boost to the Jamb Latch to speed up the metastability resolution. The charge pump can be either Clock-driven Voltage-Boosted Synchronizer (CVBS) or Metastability-driven Voltage-Boosted Synchronizer (MVBS). A new methodology for extracting the metastability parameters of synchronizers under changing biasing currents is proposed. For a 1-year MTBF specification, MVBS and CVBS show 2.0 to 2.7 and 5.1 to 9.8 times the delay improvement over the basic Jamb latch, respectively, without large power consumption. Optimization techniques including transistor sizing, Forward Body Biasing (FBB) and dynamic implementation are further applied. For a common MTBF specification at typical PVT conditions, the opti-

mized MVBS and CVBS show 2.97 to 7.57 and 4.14 to 8.13 times the delay improvement over the basic Jamb latch, respectively. In post-Layout simulations, MVBS and CVBS are 1.84 and 2.63 times faster than the basic Jamb latch, respectively.

Acknowledgements

I am extremely grateful to Professor Sachdev, my Ph.D advisor, for his support and guidance throughout this work. I would also like to thank Professor Aagaard, Professor Bishop and Professor Trefler for being on the Ph.D committee. Special thanks go to Professor Fei Yuan from the Ryerson University as the external examiner. Thank you for all your valuable feedbacks and suggestions.

I would also like to thank Professor David Nairn, Professor Siddharth Garg and Professor Adam Neale for their helpful support.

I would also like specially thank Pierce Chuang for being the best collaborator and friend during the academic project. I would also like to thank Qing Li for helping me solving layout issues and everyone else in the CMOS Design and Reliability Group at the University of Waterloo for their support.

I would like to acknowledge the financial support from the China Scholarship Council (CSC).

Finally, I would like to thank my family for their support and encouragement during my Ph.D study.

Dedication

To my beloved wife, son and mom.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ASIC Application-Specific Integrated Circuit

DSP digital signal processing

DCT Discrete Cosine Transform

CCI Cross-Coupled Inverter

CMOS Complementary Metal-Oxide-Semiconductor

DFS Dynamic Frequency Scaling

DVS Dynamic Voltage Scaling

EDS Error Detection Sequential

FPGA Field-Programmable Gate Array

MOS Metal-Oxide-Semiconductor

NMOS N-channel Metal-Oxide-Semiconductor

MTBF Mean Time Between Failure

PLL Phase-Locked Loop

PMOS P-channel Metal-Oxide-Semiconductor

PVT Process, Voltage and Temperature

TRC Tunable Replica Circuit

TSMC Taiwan Semiconductor Manufacturing Company

FOM Figure-of-Merit

VCO Voltage-Controlled Oscillator

PSNR Peak-Signal-Noise-Ratio

LSB Least-Significant Bit

VOS Voltage Over-Scaling

DFS Dynamic Frequency Scaling

STA Static Timing Analysis

LDMC Logic Delay Measurement Circuit

FBB Forward Body Biasing

PS Power Supply

LR Linear Regulator

LE Logic Element

PnR Place and Route

ISB Intermediate-Significant Bit

RTL Register-Transfer Level

SEU Single Event Upset

ORI Original Jamb Latch

GNDED Grounded Jamb Latch

GATED Gated Jamb Latch

VBS Voltage-Boosted Synchronizer

MVBS Metastability-driven Voltage-Boosted Synchronizer

CVBS Clock-driven Voltage-Boosted Synchronizer

Chapter 1

Introduction

1.1 Introduction

As CMOS technologies evolve into nanometer regions, performance, power and reliability are the three most critical problems in digital design. High performance saves computational time, low power economizes energy while strong reliability assures long-term quality. Nowadays, low power/energy design has become more and more important mainly for two reasons [8, 9]. First, digital circuit applications powered by batteries such as laptops, car devices and cellphones are becoming ubiquitous . The speed of these digital circuits has evolved to satisfy the average usage. However, their power/energy consumptions have gained more concerns due to the slow evolution of battery technologies[10]. Second, since the integration scale of transistors in systems such as data centers [11] is increasing, the power consumption shall be limited and/or heat removals shall be added.

However, timing variation due to the effect of PVT variations of digital systems has become a major obstacle for high-performance and low-power. In order to counteract the impact of these variations, a delay- or voltage-margin is often added to clock period or supply voltage, respectively. However, the added margin deteriorates the performance or increases the power consumption. Energy-efficient designs with EDS circuits have been developed as alternatives in which error signals quantifying the timing variation are used

for further responses such as timing error recovery or DVS. Nevertheless, this type of systems suffers from the timing-reliability problem due to EDS circuits (in this work the EDS-reliability is referred to this problem). The EDS-reliability directly affects the system reliability. The EDS-reliability of EDS circuits can be quantified using MTBF. Both the intrinsic and extrinsic EDS-reliability exist. The intrinsic EDS-reliability is caused by the metastability behavior of synchronizers in EDS circuits which perform as a classifier in the timing domain (essentially *synchronizer reliability*). The extrinsic EDS-reliability is defined as the avoidance of the actual timing errors when the EDS circuits are deployed to detect the errors (slack deficit or timing errors). The MTBF for the extrinsic EDS reliability is directly related to the EDS circuits deployment in the application systems.

1.2 Motivation

To improve the EDS reliability, often circuit parameters such as performance, area, and energy are sacrificed. The motivation of this research is to improve the EDS reliability with much improved circuit parameters. In particular, this thesis makes contributions in two areas: (1) the effectiveness of an EDS system for power savings is demonstrated in a FPGA with real life data. In this experiment a feedback loop was constructed with EDS. A DSP unit with Discrete Cosine Transform (DCT) was implemented for image processing. Experimental results show that substantial power saving can be achieved without compromising image quality. (2) synchronizers are often used in digital systems for a broad range of applications. In general, they help synchronize an asynchronous event to the synchronous system. A register or a latch is a simple synchronizer aligning data to an incoming clock. Synchronizers are also key elements in EDS systems determining error signals. Metastability can compromise synchronizers and EDS reliability; and it becomes an issue when the supply voltage is aggressively lowered to conserve power and energy. We are proposing VBSs which demonstrate better metastability performance.

1.3 Thesis Overview

The rest of the dissertation is organized as follows: Chapter 2 discusses the background of energy-efficient systems with EDS circuits and the EDS-reliability related issues. This chapter first introduces the effects of PVT variations and the conventional methods to combat these effects. EDS circuits and state-of-the-art digital systems with EDS circuits are studied. The metastability problem of synchronizers in EDS circuits is discussed. Chapter 3 proposes a new EDS deployment strategy to improve the extrinsic EDS-reliability. The proposed strategy is applied to an FPGA-based DCT unit with EDS and DVS circuits. Chapter 4 proposes VBSs to solve the performance bottleneck with a MTBF specification of the intrinsic EDS reliability, in an Application-Specific Integrated Circuit (ASIC) technology. A new methodology of metastability parameter extraction is also proposed. Simulation results of proposed synchronizers show the significant improvement of τ. Chapter 5 concludes the contributions and proposes the future work.

Chapter 2

Energy-Efficient Digital Design with EDS circuits

2.1 Introduction

Chapter 1 introduced the concept of the energy-efficient systems with EDS circuits and the motivation of timing-reliable EDS circuit design and deployment. This chapter first introduces the PVT variations and the conventional methods to counteract the effects of the variations. EDS circuits that detect the PVT variations are introduced and the detailed timing of EDS circuits is analyzed. In the next section, state-of-the-art digital systems with EDS circuits are introduced and the extrinsic reliability problem of EDS circuits is discussed. Later, the metastability problem of EDS circuits is discussed in detail, including basic concepts of metastability, metastable signal transformation, metastability failure mode, metastability mitigation and the methodologies for metastability parameter extraction. Finally conclusions are drawn.

2.2 Digital Circuits and Variability

2.2.1 Conventional Synchronous Design

As shown in Fig 2.1, the delay $t_{d,dp}$ of one stage of a data-path pipeline in a conventional synchronous design is given by

$$t_{d,dp} = t_{CQ} + t_{setup} + t_{logic} + t_{margin}, \tag{2.1}$$

where t_{CQ} and t_{setup} are the CLK-to-Q output time and the setup time of the sequential circuits (used as storage elements) $FF1$ and $FF2$ respectively, t_{logic} is the longest path delay of the combinational circuits. The margin t_{margin} is added to counteract the effect of the worst-case combination of PVT variations to ensure correct functionality. The t_{margin} restricts the clock frequency.

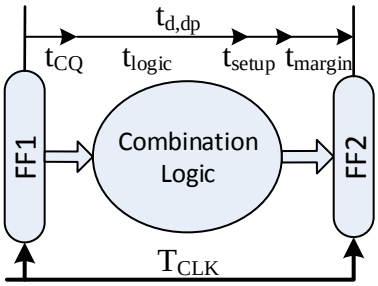


Figure 2.1: Conventional Design.

We have the first constraint for clock period T_{CLK} from the data-path:

$$T_{CLK} \geq t_{d,dp}. \tag{2.2}$$

2.2.2 Circuit Scaling and Variability

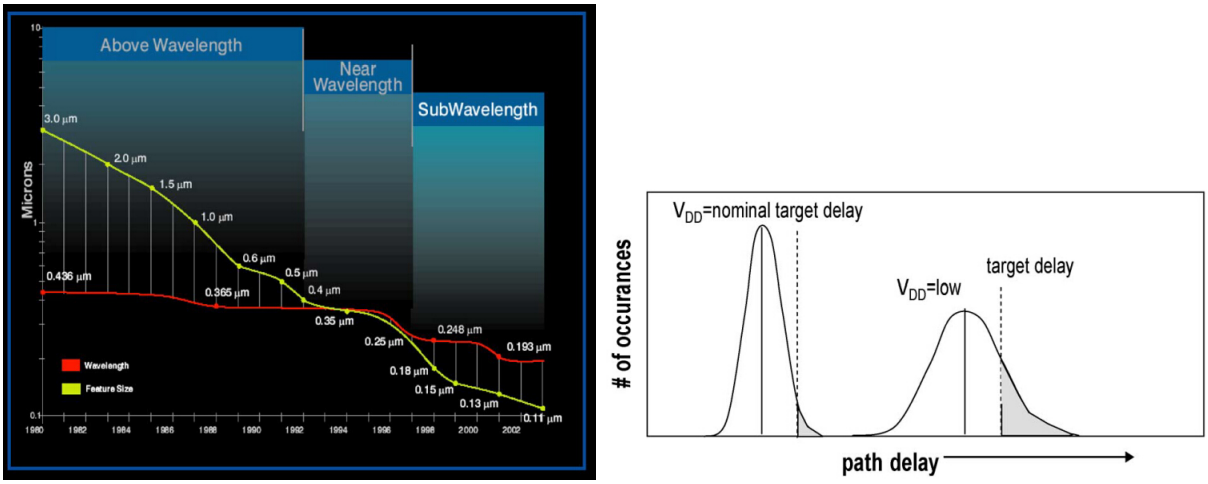
Low-power/energy design has become important for digital applications. However, it is faced with unavoidable obstacles such as PVT variations that occur globally or locally in chips, spread across CMOS product lifespans.

Due to the imperfect manufacturing, actual transistor sizes and routing delays vary among chips and inside a single chip. Aging effects [6] also requires an extra margin. As process technologies scale down into deep sub-micron levels, device feature sizes eventually become smaller than the optical wavelength of lithography process in advanced processes (c.f. Fig 2.2a)[6]. Thus, the process variations have become even larger than before.

Although power supplies are used to maintain stable external voltages, data switching activities can still induce significant internal voltage variations. Abrupt data switching activities can cause sudden current changes that induce IR droop and inductance voltage drop. In high data switching activities, a large current flows through resistances, generating heat from inside a chip. As supply voltages scale down to low levels, path delays of circuits vary more severely (c.f Fig 2.2b) [6].

Since CMOS integrations such as the emerging 3D integrated circuits and circuit performance are rapidly increasing, the power density has become a constrain for CMOS chip designs. However, packaging approaches for sufficient heat removal is still not economical [12]. As the junction temperature goes up, both the carrier's mobility and the driving current of the transistors decrease. Thus the speed of the circuit is degraded. Assuming the same ambient temperature, the junction-temperature variation is more and more exacerbated, leading to worse circuit delays.

In summary, as technologies further scale, the PVT variations have become more severe. Thus data-path delays vary in a larger scope, requiring a larger t_{margin}.



(a) (b)

Figure 2.2: (a) The trend of CMOS technologies; (b) a demonstration of path delay variations (from [6]).

2.3 Energy-Efficient Digital Systems with EDS Circuits

2.3.1 State-of-Art Approaches to Deal With Data-path Variability

[6] provides a summary of stage-of-art approaches to deal with Data-path variabilities, listed as follows according to the level of the variability awareness.

Transistor level optimizations, including transistor sizing and dual V_{TH}, have been proposed to both speed up data-path and reduce the timing uncertainty at the expense of the area and/or power consumption.

Critical instruction management. The critical instructions trigger the critical paths and can be executed in a different way as compared with other instructions, such as reducing operating frequency and using a dedicated unit to process them. Nevertheless, the latency is increased or an extra area is needed.

Post-silicon calibration. To counteract the effect of the process variations, in addition to adding a part of margin for process variations to all chips, conventional designs also use benchmarks to measure the speed of each taped-out chip. After the measurement, the process-variation margin can be reduced by applying adaptive voltage or frequency. The energy is saved or the speed is improved at the cost of calibration effort.

Sensor-based adaptive architecture technique. Here a sensor is the circuit that provide dynamic information. Examples include on-chip leakage sensors, the control voltage of Voltage-Controlled Oscillator (VCO) ([13]) in a Phase-Locked Loop (PLL) that can dynamically capture performance variations in a circuit, on-chip wearout detection circuit, temperature sensor and EDS circuits. Nevertheless, EDS circuits are relatively light and provide the finest information. This dissertation focus on EDS circuits only.

2.3.2 Variability and EDS circuits

PVT variations are unavoidable, however, their worst-case combination is assumed to be rare ([14]). Leveraging this characteristic, EDS circuits have been developed to detect the variations [1, 2, 3, 4]. As the name suggests, EDS circuits are special sequential circuits that detect errors. Here an error means unfavorable timing variations of the EDS-monitored data-paths, such as a slack deficit or an actual timing error. Nearly all state-of-the-art EDS circuits follow the architecture abstraction in Fig 2.3a. An EDS circuit like Fig 2.3b consists of a storage element (from D to Q), a synchronizer (from D to E), and an XOR gate. Fig 2.3d shows a simple deployment of EDS circuits. In this deployment, signals $S0$ and $S1$ are the outputs of precedent stages at $t = 0$ (Fig 2.4a) and propagate through the combinational logic (so called EDS-monitored paths) to port D at variable time that depends on paths and PVT variations. EDS-monitored paths map the effect of the PVT variations into timing variations and the EDS circuits act as binary classifiers to detect slack-deficit or timing errors.

2.3.3 EDS Circuits Timing

The brief timing of EDS circuits are shown in Fig 2.3c where D, Q and E represent the signals at port D, Q and E respectively). Various arrival time of The signal D causes diverse EDS output combinations of Q and E, triggering normal or abnormal system behaviors.

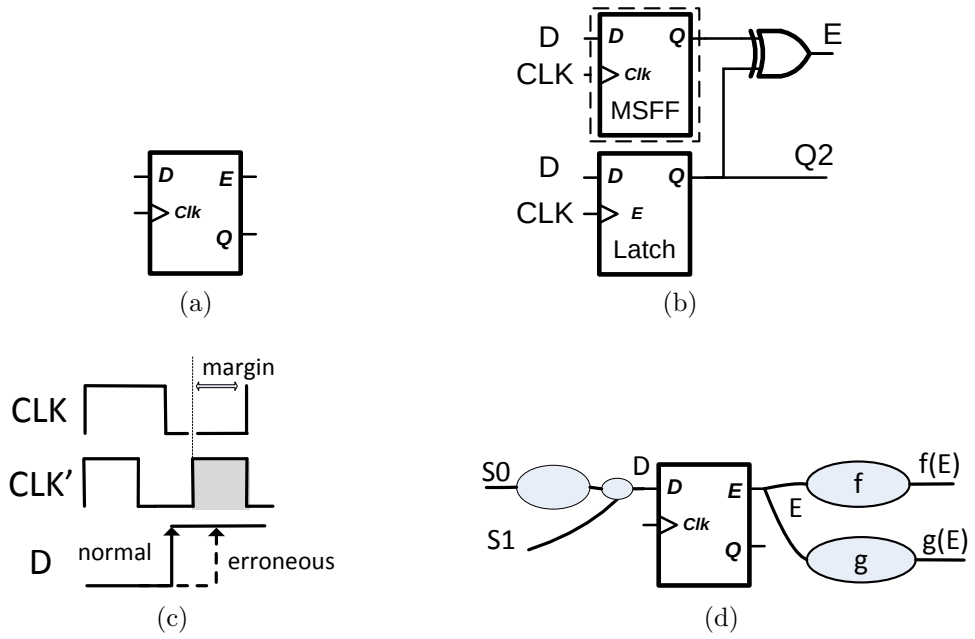


Figure 2.3: (a) EDS abstraction (b) An EDS circuit: Double Sampling with Time Borrowing (c) EDS brief timing. (d) EDS circuit deployment.

If D arrives between (β, ϕ) ("normal windows", Fig 2.4b), Q shows no timing errors ("false") and E indicates "negative". This case represents normal executions with the highest probability among all the output combinations.

If D arrives between (ϕ,$\phi+\beta$) ("detection window", Fig 2.4c), Q outputs a timing error ("true") while E indicates "positive". This "True positive" will not lead to system failure if the timing error is to be corrected.

If D arrives too early (from S1) between (ϕ,$\phi+\beta$) ("min-delay window", Fig 2.4c), Q is "false" and E reports "positive". During the recovery of "false positive", the data

of the errant instruction still corrupts the data of the earlier instruction in the latch of EDS circuits. Thus the min-delay constraint must be satisfied for all the EDS-monitored paths in these systems, which can be done by inserting buffers into the short ones of the EDS-monitored paths (such as S1).

If D and CLK transition simultaneously at t=ϕ, i.e., D falls into "the metastable window" δ (Fig 2.4d), E becomes metastable. The synchronizer in the EDS circuit is ambivalent in identifying the arrival time of D. Therefore, port E outputs an intermediate-level signal (Fig 2.4d), namely "synchronization metastability". To be distinct from timing errors in data paths, in this thesis metastability only denotes the "synchronization metastability" in the control paths.

2.3.4 EDS-Triggered System Responses

Further responses are applied accordingly to achieve energy efficiency or performance improvement and this will be discussed later. For example, the guided DVS circuits accordingly adjust the operating voltage to a proper lower level. Due to the quadratic relationship between dynamic power consumption and supply voltage, significant power saving can be achieved.

2.4 EDS Applications

2.4.1 Microprocessors with EDS

To reduce energy consumption, EDS circuits have been applied in microprocessors [1, 15, 16] to monitor critical paths. To ensure no system failures, the detected timing errors are then corrected by timing error recovery circuits implemented by a microprocessor's inherent mechanisms such as instruction stalling and jumping. The slow response circuits like DVS or Dynamic Frequency Scaling (DFS) then accordingly and properly adjust supply voltage V_{op} or clock frequency (for simplicity, this dissertation only considers DVS).

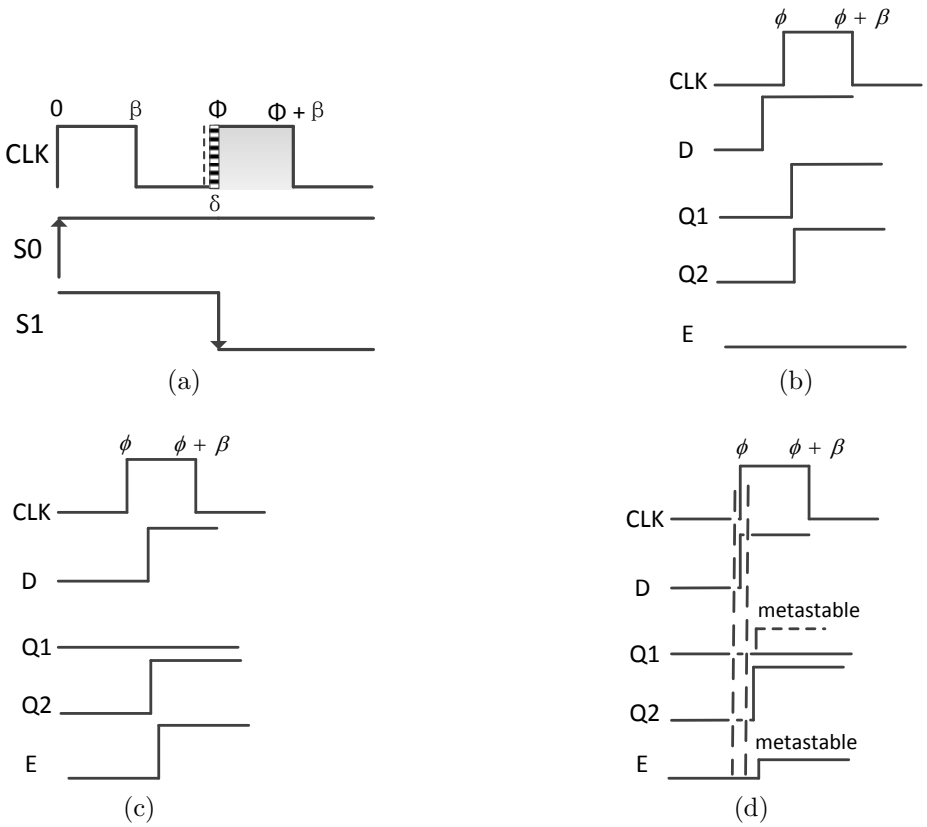


Figure 2.4: EDS detailed timing: (a) timing region division; (b) normal, (c) erroneous, (d) metastable timing.

The timing error recovery mechanisms have one primary advantage. By the virtue of the timing error recovery mechanism, a faster clock CLK' can be used instead of the CLK used in the conventional design for error-free operations at the same supply voltage. This is illustrated in Fig 2.3c. Thus, at the same clock frequency, the operating voltage V_{op} for microprocessors with EDS circuits can be automatically lowered by DVS circuits, which is illustrated by

$$V_{lower} \leq V_{op} \leq V_{upper}. \tag{2.3}$$

Here the voltage V_{upper} is the point where conventional designs can operates. The V_{lower} is the lowest safe limit for DVS in microprocessors with EDS. This limit can be obtained by

Static Timing Analysis (STA) [15]. The lower V_{op} leads to energy reduction. Nevertheless, extra circuits and error recovery actions can induce some energy penalty. If the energy reduction surpasses the energy penalty, the total energy is saved.

Many state-of-the-art microprocessors with EDS have been proposed in a variety of architectures, processes and supply voltages. Researchers reported 22% to 52% energy savings at equivalent performance in [15]. About 41% throughput gain at equivalent energy consumption was reported by [1]. Other examples of similar energy savings were reported in [17],[18],[19],[20], [21].

Nevertheless, the timing error recovery mechanisms have several disadvantages:

(1) Adding significant design complexity. Currently, timing error recovery mechanisms are built based on the inherent microprocessor control logic which usually is not available in DSP circuits.

(2) Adding indeterminate latency. This may not be acceptable for real-time DSP circuits. Large or small latency may or may not be a problem for many DSP circuits, however, indeterminate latency can be troublesome for the control logic of DSP circuits.

(3) Metastability problem. It mostly impacts fast responses such as timing error recovery circuits. [22] calculates P_{meta} to be 2e−30 under the supply voltage ranging 1.2 V−1.8 V in a 0.18 µm technology. However, it is unclear whether it is calculated under the worst-case or normal PVT conditions. Even so, recently, after analyzing the state-of-the-art academic microprocessors with EDS circuits, [23] believes the metastability problem to be one of possible reasons for why there has been no industrial microprocessor with EDS circuits. Nevertheless, a large amount of time for metastability resolution can be added to slow responses such as DVS. Thus, the metastability problem is less likely to crash the systems with only slow responses.

2.4.2 DSP Circuits with EDS

For real-time DSP circuits, it is preferable to apply the mentioned slack-deficit-detection-only techniques. The difficulty of these techniques mainly lies in PVT-to-timing converters.

Due to the lack of error correction and DVS limit, PVT-to-timing converters must satisfy speculative and accuracy requirements to avoid direct and indirect timing errors, respectively. The speculative requirement is that neither monitored nor non-monitored paths should generate unacceptable timing errors when slack-deficits are being detected. The accuracy requirement can be viewed from the two perspectives of spatial and temporal sampling: EDS-monitored paths should be placed close to the actual circuits to achieve high correlation or else a larger extra margin is needed to compensate for local variations; EDS-monitored paths should be regularly activated with realistic input data for accurate detection or else there is a possibility of misguided VOS (temporal sampling).

Targeting to satisfy the requirements with minimum penalties, various EDS circuit deployment strategies were proposed. In [1] Tunable Replica Circuits (TRCs) consist of a series of delay-tunable buffers with EDS circuits placed at the endpoints of TRCs. A TRC mimic the worst case delay of a pipeline system with a small timing margin. A limited number of TRC can be placed along with the actual circuits and activated in every clock cycle to detect global PVT fluctuations. In [2], it was shown that a small number of timing errors at LSB in some error-tolerant DSP circuits do not significantly degrade overall quality. Thus, the non-critical Least-Significant Bit (LSB) paths of a DCT unit are purposely extended by adding buffers as critical paths. EDS circuits are augmented to these extended paths. A test image Lena is used as the input data. PSNR is a typical measurement method for image processing. PSNR of the output of the proposed design is only degraded about 4 dB compared to that of the DCT design. Nevertheless, only simulation results are provided and no actual circuit was implemented.

Unlike ASIC, finely-tuned buffers are impractical in an FPGA [24, 3, 4]. Instead, [25] introduced a method namely LDMC. As in Fig. 2.5a, 128 inverters form a very-long chain. The input of the chain is the clock CLK. CLK also drives 128 flip-flops as EDS circuits to monitor outputs of each inverter. The 128 outputs of the flip-flops represent the delay of the chain and thus indicate the PVT status of the FPGA chip. Accordingly, a DVS circuit is driven to adjust the supply voltage.

Another way is to use fine-tuned clocks in FPGA. As in Fig. 2.5b from [24, 3, 4], EDS circuits (the shadow register) monitor some of the critical paths of FPGA-based DSP

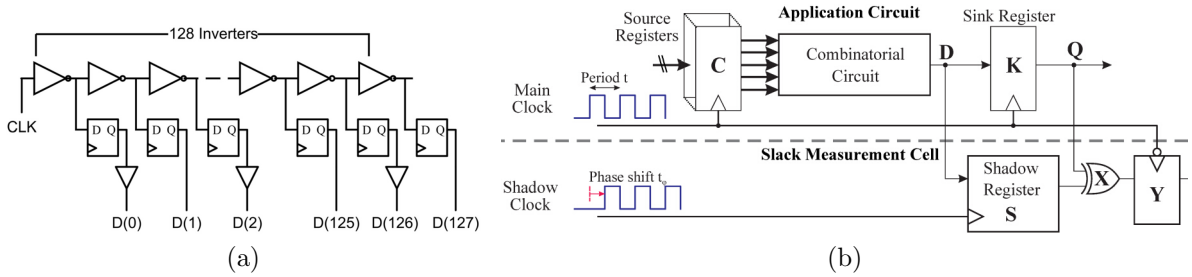


Figure 2.5: (a) LDMC (from [25]); (b) Slack measurement (from [24]).

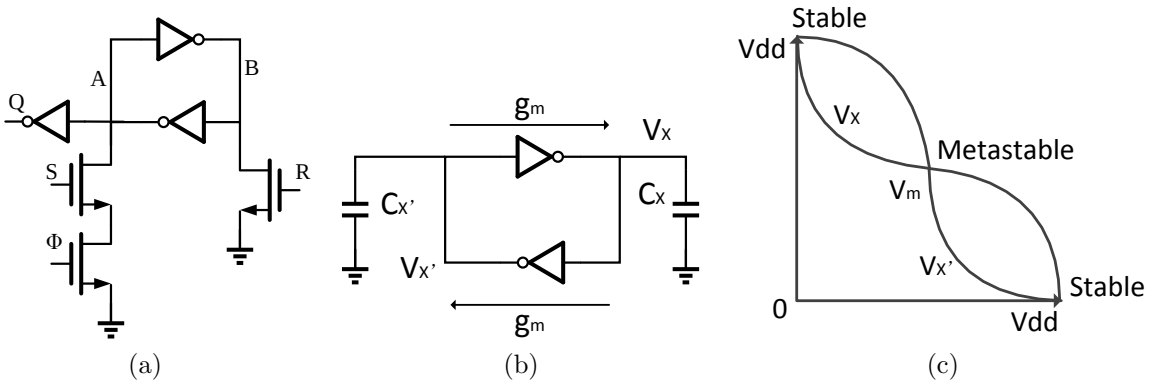
circuits. An extra phase-shifted clock (shadow clock) together with the main clock is used to drive these EDS circuits, precluding buffer insertions. Nevertheless this method can have two major disadvantages, an extra clock and validation with realistic data. The extra clock phase-shifting requires more resource, place-and-route effort and energy consumption. More importantly, the critical paths are not often activated with realistic data. Typically, a new FPGA design methodology should be validated from two perspectives, various hardware styles and extensive input data. [3, 4] provides numerous types of test hardware (even including the DCT that will be used for my work). However, I was unable to discover the exact input data set for the measurement in the papers. Nevertheless, the problem of unexercised paths is mentioned. It is also stated that various methods are being studied to alleviate this problem. However, no recent research is reported.

2.5 Synchronizer and Metastability Considerations for EDS Circuits

Synchronizers and their metastability are an elusive topic. A tutorial of metastability is provided in [26]. Further details of synchronizer knowledge can be found in the book *Synchronization and Arbitration in Digital Systems* [27]. In the following paragraphs, synchronizers and metastability will be described in brief.

Usually a latch (Fig 2.6a) (including the CCI Fig 2.6b) is a basic synchronizing unit.

Normally S fully charges the CCI to true stable states ($(V_X, V_X') = (V_{dd}, 0)$ or $(V_X, V_X') = (0, V_{dd})$, shown in Fig 2.6c). However, if S and Φ transition simultaneously, from the circuit perspective, the setup or hold time of the synchronizer will be violated, the CCI are only partially-charged and enter an intermediate level V_m shown in Fig 2.6c. Ideally the CCI stay at V_m forever, however, in the actual world even a very small noise can break the balance and force the CCI to settle down to true stable states. This phenomenon is called metastability.



(a) (b) (c)

Figure 2.6: A synchronizer (a) The ORI schematic (b) CCIs (c) voltage transfer curve.

2.5.1 Synchronization Parameters

MTBF due to Metastability

Metastability failure probability per operation, i.e., the probability that the metastability resolution time reaches or exceeds a given value t_r is defined as [28]

$$P_{meta} = f_c t_w e^{-\frac{t_r}{\tau}} \tag{2.4}$$

and MTBF due to metastability is [28]

$$MTBF = \frac{1}{P_{meta} f_d} = \frac{e^{\frac{t_r}{\tau}}}{T_w f_c f_d} \tag{2.5}$$

where t_w is the asymptotic width of the metastability window, τ is the resolution time constant, f_d and f_c are data and clock frequencies, respectively. *The metastability parameters such as τ and t_w describe the relationship between delay t_r and MTBF.* t_w is usually a small portion of clock period and $t_w f_c$ is calculated as 0.05 for a 90 µm technology at 0.3 V in [29].

Synchronizer Delay

For synchronizers with constant τ, t_r can be expressed as

$$t_r = N_r \cdot \tau \tag{2.6}$$

where N_r is a *normalized* unit-less value of the metastability resolution time. N_r is a dominant term for MTBF. Although MTBF, P_{meta} and N_r are all useful parameters for describing synchronizer reliability, researchers tend to use N_r.

The delay t_d of a synchronizer such as a Jamb latch [30] in Fig. 2.6a is approximated as [29]

$$t_d \simeq t_n + t_r \tag{2.7}$$

where t_n is the nominal delay of a synchronizer without violating the setup time requirement. t_n is given by

$$t_n = t_{setup} + t_{CQ}. \tag{2.8}$$

Thus

$$t_d \simeq t_n + N_r \cdot \tau = t_n + \tau \cdot ln(MTBF \cdot T_w \cdot f_c \cdot f_d) \tag{2.9}$$

So far we have another constraint for the clock period from the synchronization paths which is given by

$$T_{CLK} \geq t_d, \tag{2.10}$$

assuming synchronizer delay t_d is the speed bottleneck. Here provides a summary (Table. 2.1) of the timing parameters in this Chapter.

Table 2.1: Definitions of timing parameters.

Timing parameter	Definition
$t_{d,dp}$	the delay of one stage of a data-path pipeline
t_{CQ}	the CLK-to-Q output time of the sequential circuits
t_{setup}	the setup time of the sequential circuits
t_{logic}	the longest path delay of the combinational circuits
t_{margin}	the margin added to counteract the effect of PVT variations
t_r	the metastability resolution time
t_w	the asymptotic width of the metastability window
τ	the metastability resolution time constant
t_d	the delay of a synchronizer
t_n	the nominal delay of a synchronizer

Energy Consumption

The overall average energy consumption E_{total} can be estimated as

$$E_{total} = (1 - a\%)E_{idle} + a\%[(1 - b\%)E_{norm} + b\%E_{meta})] \tag{2.11}$$

where E_{idle}, E_{norm} and E_{meta} are the energy consumption during the idle status, a normal data activity and metastability, respectively. $a\% = \frac{f_d}{f_c}$ is data activity and $b\% = T_w f_c$ is metastability probability with data activities. For asynchronous communication systems or the final synchronizers in a resilient digital systems, $a\%$ is usually very small due to the hand-shake protocol or the small timing-error rate setting. For synchronizers used as data samplers such as shadow flip-flops in the Razor-style designs, $a\%$ is the same as data-path activities. $b\%$ is very small as calculated. Thus E_{total} is mainly determined by E_{idle} and E_{norm}. However, synchronizers are not always operated at the maximum clock frequency. Thus in this case, extra energy can be represented by the leakage power of the synchronizers.

2.5.2 Figure-of-Merit of Synchronizer Design

This dissertation classifies digital design parameters into three categories: system parameters, circuit external parameters and circuit internal parameters. System parameters such as MTBF and t_d are determined or designed by system designers. Circuit external parameters such as t_n, t_w and τ are visible to system designers and designed by circuit designers. Circuit internal parameters such as the capacitance C and the transconductance g_m are invisible to system designers and determined by circuit designers. An Figure-of-Merit (FOM) is chosen from circuit external parameters. It is the key prerequisite to the success of a circuit design. In other words, if targeting an improper FOM, fancy circuit designs may easily be introduced. However, these designs might not meet the requirements of actual systems where a circuit FOM should be evaluated. Hence, this work adopts the FOM of synchronizers that are well accepted and explored by mainstream research. Nevertheless, this could place a tougher task for innovative circuits design than using a self-proposed FOM. The FOM of synchronizers in this dissertation is obtained and reviewed from a system perspective as follows.

Firstly, MTBF due to metastability of synchronizers can directly constrain the overall system MTBF. This is because the overall system MTBF is upper-bounded by the smallest MTBF among various kinds of failure mechanisms. Secondly, the synchronizer delay t_d can possibly constrain the system speed T_{CLK} for a fixed system architecture or increase the latency of the system. Thirdly, the number of such synchronizers in typical system designs is usually only a few. Typical system designs use no or only a few synchronizers for peripherals, and these synchronizers consume a negligible percentage of power consumption. These extra synchronizing flip-flops in EDS circuits consume $< 0.9\%$ ([1]) to 5.7% ([17]) of the entire power. Asynchronous communication systems (sakurai [31]) can also leverage only a few synchronizers. [30], [28], [32],[33]). Thus, the delay of critical synchronizers is usually weighted heavier than their power and area. Synchronizer delay and MTBF of synchronizers due to metastability exclusively constrain each other and τ is the key coefficient to describe this constraining relationship. A common viewpoint is that in synchronizer designs, MTBF due to metastability is prioritized the highest and specified

first, while synchronizer delay is usually weighted heavier than energy consumption or area cost and targeted to be improved by methods such as reducing τ of the synchronizer. In simple words, τ is the main FOM of synchronizers. Notice that t_d and MTBF are NOT circuit parameters but system parameters. In other words, "high speed(performance)" or "reliable" may not be proper words to describe a synchronizer and "metastable-hardened" is used instead for describing synchronizers with better metastability parameters (τ and t_w).

Nevertheless, the circuit parameter τ alone is less intuitive. It is better illustrated and understood along with MTBF and t_d. The specification of the metastability-induced MTBF needs to be set first. Researchers [34] report a metastability-induced MTBF 10^{10} times greater than the targeted Single Event Upset (SEU) induced MTBF. Similarly, authors in [22] calculate $P_{meta} = 2e{-}30$ in its system, which is roughly converted into $N_r = 65$ (two stages). Conventional synchronizer design suggests a N_r around 29-40 for single stage synchronizing. This work applies a $N_r = 35$ specification for one synchronizing latch, indicating a 1-year MTBF for the latch if assuming $t_w f_c = 0.05$ [29] and $f_d = 10^9$ (assuming the worst case scenario).

EDS circuits inspire a deeper understanding of the roles that sequential circuits play in system designs, either storage elements or synchronizers. Storage elements do not suffer from the metastability problem since they are protected by timing margins even in the systems with EDS circuits. Thus neither τ nor T_w is a FOM of storage element. This is evident since the standard cell flip-flops show very poor τ. When used as storage elements, their number usually is large and the power(energy) consumption and the nominal delay (t_n) of sequential circuits usually constitute considerable percentages of the system's power(energy) and speed, respectively. Thus, these two parameters are similarly important and the Power(Energy) Delay Product can be used as the FOM of storage elements [35]. In summary, synchronizers and storage elements have a different FOM. Furthermore, not only system designers but also STA tools are able to identify the roles of sequential circuits as synchronizers or storage elements because properly installed storage elements (synchronizers) always satisfy (violate) the static timing constraint shown in Eq. (2.1). Due to all these reasons, it is better to design dedicate synchronizers than to design general sequential

circuits for both usages.

Impact of Scaling on Synchronizer FOM

As mentioned both synchronizer delay and data-path delay can bottleneck system speed, thus a comparison between them is necessary for system designers. however, this comparison is unknown to circuit designers and estimated by

$$k = \frac{\tau}{t_{FO4}} \tag{2.12}$$

where t_{FO4} is the delay of a inverter with a fan-out of 4. As technologies scales, though the absolute data-path delay and synchronizer delay (control-path) are both improved, the former gain more improvement than the latter. Similarly, as supply voltages scale down, the synchronizer delay worses more than the data-path delay does. Thus synchronizer delay t_d becomes more likely than data-path delay $t_{d,dp}$ to bottleneck system performance.

As mentioned in Section 2.2.2, technology scaling also enlarge PVT variations. This can impact not only data-path delays but also synchronizer delay.

2.5.3 Synchronizer Design Fundamentals

To minimize t_d, two techniques have been developed, including synchronization pipelining or improving metastability parameters. Similar to data-path pipelining, the synchronizer pipelining is to convey the metastable status of the current synchronizer to those of the subsequent synchronizers and continue to settle down. Thus it gives more metastability resolution time. This is shown in

$$t_r = t_{r,1} + t_{r,2} + \cdots + t_{r,n}, \tag{2.13}$$

where $t_{r,i}$ is the metastability resolution time of the ith synchronizer. This equation implies that each synchronizer in the synchronization pipeline is identically important, even though the metastability is initialized at the first stage. More importantly, for the same overall

metastability-induced MTBF requirement, the MTBF specification of each stage can be reduced, leading to a smaller t_d for each stage. Nevertheless, this pipelining technique has the same disadvantages as data-path pipelining does. First, each stage adds an extra t_n. Second, more importantly, it significantly increases the latency and data-path complexity. Generally, synchronization pipelining techniques have different impacts on three types of systems: Type 1 systems where synchronization-latency is non-important, for example, the proposed system with slow responses for synchronization results in Chapter 3; Type 2 Systems where synchronization-latency has some negative impact, for example, micro-processors with EDS need the corresponding extensions of the data-path; Type 3 Systems where synchronization-latency is critically important such as [16] where a local stall is used for error recovery and only one clock cycle is given for metastability resolution.

Thus, metastable-hardened synchronizers are beneficial for type-2 systems and critical for type-3 systems. We focus on improving the synchronizer circuit parameters (the FOM of synchronizers), especially τ which is determined by

$$\tau = \frac{C}{g_{m,sum}} \tag{2.14}$$

where $g_{m,sum}$ is the sum of the transconductance g_m of inverters in Jamb latch and C is lumped capacitance at nodes A and B of Jamb latch. Furthermore, $g_{m,sum}$ is expressed as

$$g_{m,sum} = g_{mP} + g_{mN} \tag{2.15}$$

where g_{mP} and g_{mN} are the transconductance of P-channel Metal-Oxide-Semiconductor (PMOS) and N-channel Metal-Oxide-Semiconductor (NMOS) transistors, respectively. The transconductance of a single Metal-Oxide-Semiconductor (MOS) transistor is calculated as

$$g_{m,MOS} = k'\frac{W}{L}(V_{GS} - V_{th}) = \sqrt{2I_D k'\frac{W}{L}} \tag{2.16}$$

where k' is a process parameter, W and L are width and length of the transistor, respectively, V_{GS} is the gate-source voltage, V_{th} is the threshold voltage, and I_D is the drain

current of the transistor. Thus, τ is greatly influenced by transistor topologies and sizing, supply voltages, and other factors.

So far, Eq. (2.14) implies a self constraint for synchronizer design between g_m and C. Besides, there needs to be a trade-off between t_n and τ. Nevertheless, both of these two trade-offs have been well studied. The bottom line is that the CCIs should be able to resolve the metastability at the critical nodes to stable states and the driving transistors should be able to drive the states at the critical nodes to the target stable states. New circuit topologies, transistor sizing, using low V_{th} and other techniques need a trade-off between τ and t_n to optimize t_d in Eq. (2.9). Transistor sizing is an inefficient way since g_m and C are both increased.

2.5.4 Synchronizer Validation

Since in Eq. (2.14) the capacitance and the transconductance cannot be directly validated, two methodologies for τ extraction in the measurable timing and voltage domains are developed: Method One in [30] and Method Two (The proofs of these two methodologies are provided in [27].). Method One is performed as follows: in Fig. 2.6a, an ideal switch K and a voltage source U with a small value are placed in between nodes A and B. K is first closed to force A and B to enter the true metastability state and then released to let them resolve. The small voltage source U (in our case this is smaller than 200 µV) initializes the metastability resolution, acting like the noise in real circuits (otherwise this resolution may not happen in simulations). The τ is calculated as

$$\tau = \frac{t_{r,1} - t_{r,2}}{ln\frac{V_{A-B,1}}{V_{A-B,2}}} \tag{2.17}$$

where $t_{r,1}$ and $t_{r,2}$ are two time points during the metastability resolution and $V_{A-B,1}$ and $V_{A-B,2}$ are the corresponding voltage differences of node A and B for time points $t_{r,1}$ and $t_{r,2}$, respectively. t_r for a specified N_r is further obtained by Eq. (2.6). Method One is simple and suitable for computer simulations.However, the extracted τ may be inaccurate technically because these course-grained simulations generate only a few data points for

the metastability resolution region.

Method Two is to change the input data time (δ_1 and δ_2) around the balance point and measure the synchronizer output time ($t_{d,1}$ and $t_{d,2}$). The τ is calculated as

$$\tau = \frac{t_{d,1} - t_{d,2}}{ln\frac{\delta_2}{\delta_1}}. \tag{2.18}$$

Method Two is complex and suitable for silicon measurement. The τ accuracy also depends on the selection of δ_1 and δ_2.

Furthermore, due to on-chip variability, τ measured from the dedicated circuits does not directly apply to the actual application scenarios. Nevertheless, synchronizer measurement in actual application circuits will induce large circuit penalty and high manual effort. Many silicon-measurements of τ are based on method Two. However, generally there exist a conflict of MTBF requirement between artifact design and reliability testing. For synchronizer measurement, special circuits namely digitally controllable delay lines are needed to generate aggressive input data with fine-tuned timing (usually precise in several *ps*) to set synchronizers in a deep metastable status. Even though, the observation of metastability in post-silicon measurements needs significant sampling time. On the contrary, the power, energy or speed measurement of other digital circuits requires less circuit penalty and manual effort. For example, power or energy can be obtained directly from measurements of supply rails.

Nevertheless, the above methodologies are only for one stage of synchronizer, i.e., one latch. Notice that a master-slave flip-flop is two consecutive synchronizers (pipelining) and showing τ of the master stage alone is insufficient to represent the overall τ of a flip-flop. However, even in simulations, it is usually impractical to measure τ of the slave stages by using method Two to input data from the master stage. Thus, mainstream research focuses on single latches. There are a few exceptions that people work on flip-flop synchronizers such as measuring metastability parameters of the readily-available circuits such as the standard cell or FPGA registers. Nevertheless, even these measurements only show τ of the master stages of tested circuits. Due to lack of related research literature, this dissertation believes that post-silicon measurements for multiple synchronizers need to be carried out

one synchronizer by one, even though these synchronizers form a synchronization pipeline like a master-slave flip-flop. On the contrary, though each component of a data-path may be designed and tested individually, the measurement of the data-path delay can be done by one overall test. In summary, much more design complexity and manual efforts are paid for post-silicon measurements for multiple synchronizers.

2.5.5 Existing Synchronizer Design

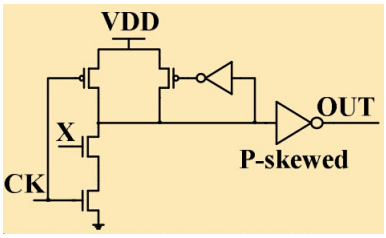
Many synchronizer techniques (summarized in [36]) have been developed, however, not all of them work sufficiently at low supply voltages. Even worse, several cases of deceptive synchronizers are described in [37]. Here this work provides some existing synchronizers.

Jamb latch (Fig 2.6a) in [30] is built using non-CLK-controlled CCI as the feedback loop instead of other gates because non-CLK-controlled inverters have a higher gain and less capacitance than other gates do [30] [27], thus achieving a low τ. A similar conclusion is drawn after extensive studies on the metastability of several high performance flip-flops ([38]). Thus, Jamb latch and its derivatives gain extensive research interest.

Assuming that flip-flops are becoming more susceptible to metastability due to the PVT variations ([39], [40], [38],[41]) and that the master stage of a flip-flop has the most impact on metastability resolution ([41] [39] [40]), [41] proposes two flip-flops with the master stages using differential CCI , namely pre-discharge flip-flop (first appeared in [42]) and sense-amplifier transmission-gate flip-flop.These two flip-flops both with significantly-sized CCI shows 30% and 24% τ reduction over a CLK-controlled flip-flop with minimum-sized CCI , respectively. Simulations results in [39] [40] show that a storage cell namely Quatro with two non-CLK-controlled coupled feedback paths has τ that are 11% smaller than that of a reference flip-flop with CLK-controlled CCI. In summary, these simulation results show that conventional methods such as non-CLK-controlled CCI and transistor sizing are still effective synchronizer techniques. Nevertheless, most innovatively and importantly, two useful FOM for flip-flops design, namely the metastability-delay-product ($\tau \times t_n$) and the metastability-power-delay-product ($\tau \times power \times t_n$), are proposed and used for analysis

throughout [41], [43],[39],[40]. Nevertheless, my dissertation still applies the FOM discussed in Section. 2.5.2 in this Chapter.

In [7], a new synchronizer is proposed as shown in Fig 2.7a and the measurement results in Fig 2.7b. However, τ is not mentioned in [7] and needs an estimation. Typically, τ of a Jamb latch for this situation is about 7 ps where t_{FO4} is 11 ps [7]. Assuming that all other factors (T_w, f_c, f_d) are the same, a simple math using Eq. (2.5) would lead to $\tau = 5.87$ ps for this dynamic synchronizer, around 16% τ improvement over the basic Jamb latch. Nevertheless, the power and area penalties are large.



(a)

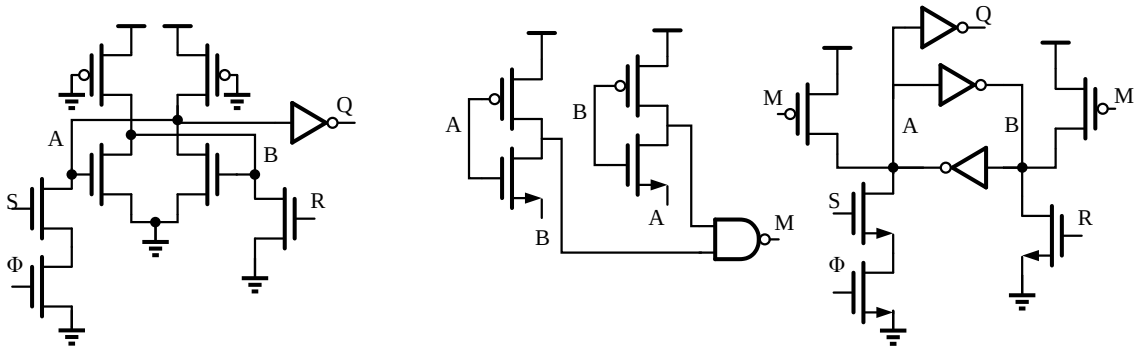
Synchronizer	Normalized MTBF at 2GHz	Power (1.2V, 25°C, simulated)	Area
3-stage, 7-inverter Dynamic Sync.	~5E7× (This work)	106 µW	52 µm²
2-FF	1×	60 µW	24 µm²
Jamb Latch	~50×	24 µW	8 µm²

(b)

Figure 2.7: The dynamic synchronizer (screen-shots from [7]) (a)schematic; (b) validation results.

Power-hungry (or current-mode) synchronizers based on Jamb latches have been developed. One power-hungry method is to ground the amplifying PMOS transistors (Fig. 2.8a) as biasing transistors to provide very large bias current [44], [29]. However, the transconductance of the PMOS transistor is removed as expressed in Eq. (2.15) and the direct-path or leakage current is large. To reduce the power consumption, extra biasing transistors can be added and controlled by a metastability detector [29] (Fig. 2.8b) or a signal pulse (symmetric boost synchronizers [45]). Nevertheless, for the first case of the metastability-driven synchronizers, delay improvement is significantly degraded due to the additional capacitance at the latching nodes and the metastability detector delay.

The other power-hungry method for low supply voltage is to reduce the V_{th} of transistors. FBB has been exploited to reduce the V_{th} of PMOS and NMOS of the driving transistors and the CCIs for sub-threshold regions [29] and thus both τ and t_n are sig-



(a) (b)

Figure 2.8: The power-hungry synchronizers: (a) GNDED; (b) GATED.

nificant reduced. Furthermore, FBB is controlled and disabled when operated in nominal voltages to reduce power consumption. However, as I demonstrate the delay bottleneck of the metastability-driven synchronizers lies in the metastability detector delay. The FBB implementation for NMOS is non-trivial and requires expensive process options. Similarly, [43] advocates using low-V_{th} transistors for only the CCI pair and standard-V_{th} transistors for the remaining circuits to improve τ in sub-threshold regions.

In [46], synchronizers are optimized by adapting them to the effect of on-chip variability at the post-silicon calibration stage. After the measurement of synchronizers on an FPGA, the best synchronizer among several redundant synchronizers is chosen to use, or, the clock frequency is adjusted accordingly for the measured synchronizer. For the latter method, the performance improvement is approximately 33% [46]. Nevertheless, the area penalty of additional circuits is very large.

Among all the synchronizers that this work has been researched on, this work deliberately chooses these techniques as our reference designs: the basic Jamb latch because of its excellent metastability performance and wide acceptance in mainstream research, the power-hungry synchronizers since they extremely improve τ at low supply voltages and the PVT-variation adaptation technique as it is a rarely-seen post-silicon optimization technique for synchronizers. Nevertheless, all these could place another tough task for our design.

2.6 Conclusion

EDS circuits can be used to detect PVT variations for further responses to save energy; however, they suffer the intrinsic and extrinsic reliability problems. The intrinsic EDS-reliability is caused by the metastability behavior of the synchronizers in EDS circuits. The MTBF due to metastability and the delay of synchronizers exclusively constrain each other and the τ is the key coefficient to describe this constraining relationship. In synchronizer designs, the MTBF due to metastability is prioritized the highest and specified first, while the synchronizer delay is usually weighted heavier than the energy consumption or area cost and targeted to be improved by methods such as reducing the τ of the synchronizer.The extrinsic EDS-reliability describes the ability of EDS circuits to avoid actual timing errors when the EDS circuits are deployed to detect the errors (slack deficit or timing errors). The extrinsic EDS-reliability depends on the EDS circuit deployment in the application systems.

Chapter 3

Design for an Energy-Efficient System with EDS Circuits

3.1 Introduction

To address the extrinsic EDS-reliability problem mentioned in Chapter 2, a new strategy for EDS circuit deployment is proposed in this chapter. Later, the proposed strategy is applied to an FPGA-based DCT unit together with EDS and DVS circuits as a proof of concept. In the next section, the experiment results are presented and finally the conclusions are drawn.

3.2 Proposed Circuit Design

In the proposed EDS deployment strategy, the non-critical paths with much higher data activities instead of the critical ones, are monitored by the EDS circuits. The speculative characteristic of EDS circuits is achieved by sampling at the clock falling edge and tuning clock duty cycle instead of sampling by an extra clock and tuning clock phase, precluding an extra clock. The effectiveness of the proposed design is demonstrated using FPGA-

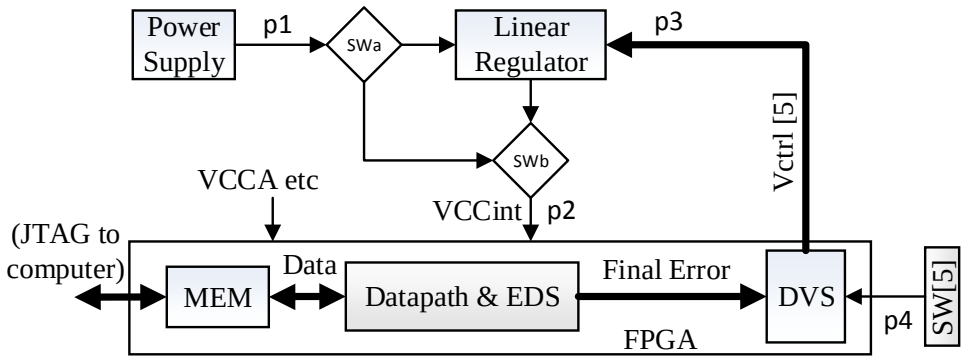


Figure 3.1: Block diagram of the proposed FPGA-based design.

based DCT with EDS and DVS circuits in a closed-loop system consists of an Altera FPGA board and a linear voltage regulator to dynamically adjust the FPGA's supply voltage (c.f. Fig. 3.1).The FPGA board (Terasic DE0 development board with an Altera Cyclone III FPGA) has two power configurations: powered either directly through the PS or via the LR (LT3070 [5]) by configuring the two switches SW_a and SW_b. In PS configuration, supply voltage can be finely tuned to nominal voltage ($V_{nominal}$) or a sub-critical voltage. The LR ([5]) is digitally programmable by DVS circuits inside FPGA. LR has two operating modes: the nominal mode where supply voltage is 1.20 V and automatic mode (the so-called "DVS" mode), controlled by the switches $SW[5]$ attached to FPGA.

3.2.1 Proposed EDS Design and Deployment

Fig. 3.2a shows that multiple EDS circuits are inserted at the endpoints of only non-critical paths such as Intermediate-Significant Bits (ISBs) in a DSP data-path. The midway nodes (such as $L1$ in Fig. 3.2b) of the critical paths can be alternative locations, however, this work does not consider this case. This is because this work does not intend to break down the arithmetic logics such as the "+" operator at the Register-Transfer Level (RTL). In FPGA, these built-in adders are implemented using special purpose logic and not easily modified without incurring a performance degradation. Thus, the latter case may be more suitable for balanced arithmetic circuits in an ASIC.

Fig. 3.2c depicts a schematic of FPGA-based EDS circuit utilizing two flip-flops (DFF1 and DFF2). DFF1 samples signal D at the clock falling edge. The sampled and the late signals are XOR-ed as the ERROR signal that is held by DFF2 in the next clock cycle. In Fig. 3.2d, D should arrive before the clock falling edge in normal conditions and does not trigger the ERROR signal. On the other hand, an ERROR signal is generated if D arrives later than that edge due to PVT variations. Nevertheless, an ERROR signal is only an indication of slack-deficit (actually it is a timing error at the sampling DFF1 but not a system failure). The ERROR signals from EDS circuits are combined together as the final ERROR signal through an OR tree (Fig. 3.2a).

Here a knowledge about FPGA clocking is necessary for understanding the sampling at the clock falling edge and provided by [47, 48, 49]. FPGA clocks can be classified as base clocks that are externally supplied to FPGA and internally generated clocks. The internally generated clocks can be further categorized as gated clocks and derived clocks. Gated clocks are a clock gated by combinatorial circuits and thus no new timing nodes are generated. Usually this is not recommended in FPGA as it would introduce skew due to the added circuits or jitter due to the deviation from the dedicated clock network. Derived clocks are a new clock that are generated though registers or a PLL driven by a primary clock. Thus features such as frequency or phase can be different from the primary clock. Here an important special case is an inverted clock. It can be generated by an inverter or a PLL with a 180 degree phase shift. However, as [47, 48] clearly state, the best way is to sample the data on the opposite clock edge since most FPGA devices provide a programmable option to input a clock or an inverted clock to a register. In common sense, this is understandable since FPGA devices should provide a choice for system designers to use either the rising or falling clock edge. By this way, the clock signal traverses from the source such as a PLL to the destination registers through the dedicated clock-tree, regardless of whether its rising or falling edge will be used.

The non-monitored critical paths can also be protected from timing errors if the monitored paths (namely $p1$) maintain a *relative margin* over the most critical path (namely $p2$). To guide the Quartus II software to deploy the EDS circuits to achieve this, this work has developed a methodology that differentiates the duty cycles β and β' of the actual clock

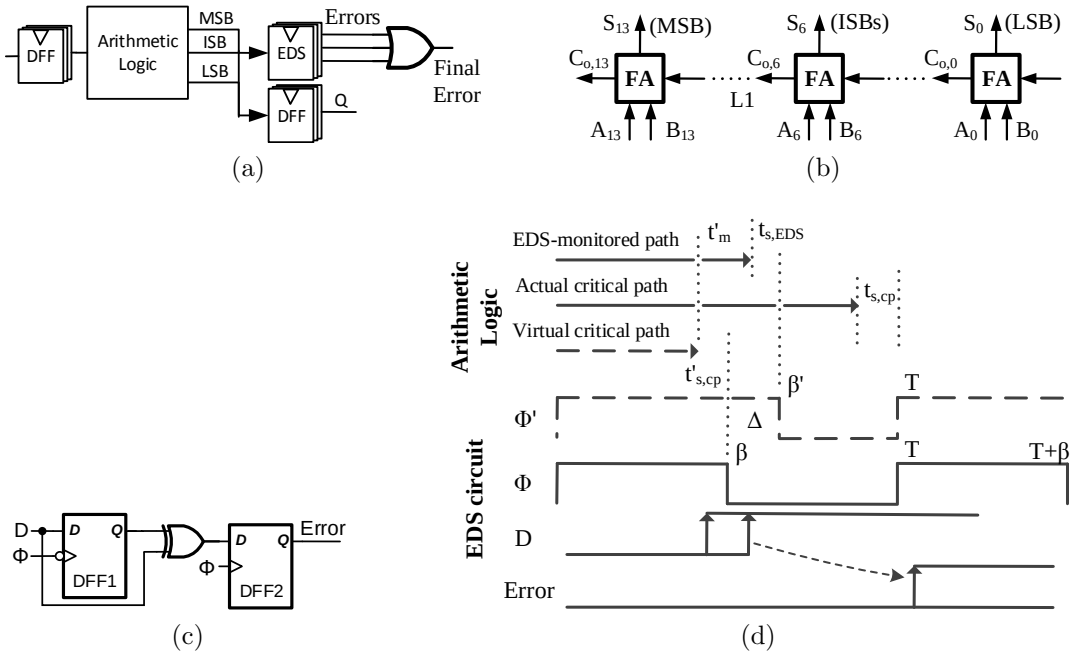


Figure 3.2: (a)DSP Datapaths deployed with EDS circuits, (b) an arithmetic logic example (ripple-carry adder); (c) EDS circuitry implementation; (d) timing analysis.

ϕ and the constraint clock ϕ' (only exists during the synthesis stage) by $\Delta = (\beta' - \beta) \cdot T$ where T is the clock period. As shown in Fig. 3.2d, the EDS-monitored and the critical paths are sampled at the clock falling and rising edges, respectively. Hence, a virtual critical path relative to the clock falling edge, which is scaled based on the actual critical path by β, is generated. Thus, the relative margin t'_m provided by the EDS-monitored paths over the virtual critical path is given by

$$t'_m = \Delta + (\beta \cdot t_{s,cp} - t_{s,EDS}) \tag{3.1}$$

where $t_{s,EDS}$, $t_{s,cp}$ and $t'_{s,cp} = \beta \cdot t_{s,cp}$ are the slacks of the EDS-monitored paths, the most critical path and the virtual critical path, respectively. Nevertheless, only those EDS-monitored paths that satisfy $t'_m > 0$ protect the most critical path and can be defined as speculative paths ("s-path") which are the so-called PVT-to-timing converters while EDS circuits are the timing-to-digital converters. That is

$$t_{s,EDS} < \Delta + \beta \cdot t_{s,cp} \tag{3.2}$$

3.2.2 DVS Algorithm Design and Implementation

This work applies a DVS algorithm driven by the comparison of the actual and reference error rates, $E_{rr}\%$ and $E_{ref}\%$ which are the ratios of the actual and maximum-allowable numbers of errors (E_{rr} and E_{ref}) over the sampling clock cycles, respectively. For the algorithm to match with t'_m, the operating voltage V_{DVS} should be scaled to where EDS circuits just start to detect errors. To achieve this, $E_{ref}\%$ should be configured extremely small and can be further simplified to whether an error occurs during ϕ_{cnt} clock cycles, where

$$E_{ref}\% = \frac{E_{ref}}{sampling\ clock\ cycles} = \frac{1}{\phi_{cnt}} \tag{3.3}$$

This configuration simplifies the error counter and the comparator to be only 1-bit in the DVS algorithm implemented in Fig. 3.3. The X-bit voltage counter divides the voltage range into 2^X levels. The final ERROR signal from Fig. 3.2a is recorded in the error

counter. After every ϕ_{cnt+} or ϕ_{cnt-} clock cycles counted by the clock counter, the comparator compares the error counter with $E_{ref} = 1$ and the error counter is reset. The voltage counter $VLevel$ raises a ΔV_+ level if the comparison result indicates an error after ϕ_{cnt+} clock cycles, or, lowers a ΔV_- level if no errors are detected after ϕ_{cnt-} clock cycles. The voltage counter $VLevel$ is decoded by the decoder as 5-bit control signals of the linear voltage regulator. A multiplexer controlled by the switches $SW[5]$ is inserted between the voltage counter and the decoder for switching the operating modes. Notice that the scaling up period $T_+ = T \cdot \phi_{cnt+}$ need to be greater than the LR response time $T_{regulator}$ to avoid redundant responses.

The probability P_+ (P_-) of scaling up (down) after every ϕ_{cnt+} (ϕ_{cnt-}) clock cycles depends on the actual error rate $E_{rr+}\%$ ($E_{rr-}\%$), respectively. For example, P_+ is

$$P_+ = p(E_{rr+}\% \geq E_{ref}\%) = p(E_{rr+} == 1) \tag{3.4}$$

A similar equation can be applied to $P_- = p(E_{rr-} == 0)$. Finally, the voltage scaling bandwidth $\frac{\Delta V}{\Delta t}$ is given as

$$\frac{\Delta V}{\Delta t} = \frac{\Delta V_+}{T \cdot \phi_{cnt+}} \cdot P_+ - \frac{\Delta V_-}{T \cdot \phi_{cnt-}} \cdot P_- \tag{3.5}$$

$\frac{\Delta V}{\Delta t}$ needs to be adaptive to the changing PVT variations so as to achieve energy efficiency or prevent VOS. A merit α can be used to evaluate the conservativeness of the DVS algorithm

$$\alpha = (\frac{\Delta V_+}{\Delta V_-}) \cdot (\frac{\phi_{cnt-}}{\phi_{cnt+}}) \tag{3.6}$$

A large α indicates a more reliable or conservative setting of the DVS algorithm, however, less energy-efficient.

3.2.3 Reliability Analysis

Other than α, P_+ and P_- also have a great impact on VOS. However, a primary limitation is that P_+ and P_- depends not only on the PVT variations but also on the activations

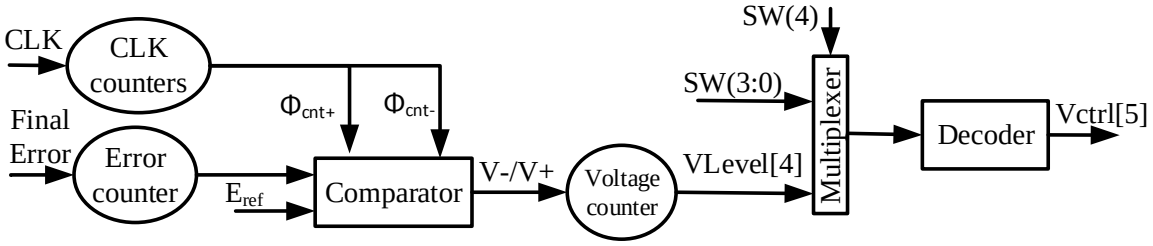


Figure 3.3: The proposed DVS circuit implementation.

of the s-paths. Thus it is very important to minimize the portion of s-path inactivations in P_+ and P_- for P_+ and P_- to mainly reflect the PVT variations. For input data with switching activity p_{in}, the probability p_N that an s-path with N logic stages is *inactivated* during M clock cycle is estimated as

$$p_{N,M} = \prod^{M}(1 - p_{in} \cdot \prod^{N} p) \tag{3.7}$$

where p is the propagation/activation probability of one logic stage and M is equivalent to ϕ_{cnt-} since P_- is vulnerable to path inactivations. For example, p for a 1-bit full adder (from carry-in to carry-out) and an inverter are 0.5 and 1, respectively. Eq. (3.7) indicates that the non-critical paths with smaller N usually have higher activations than the critical ones [2], thus the VOS probability due to $p_{N,M}$ is exponentially decreased in our strategy.

The probability that K s-path with N logic stages are inactivated during M clock cycles is estimated as

$$p_{N,M,K} = \prod^{K}\prod^{M}(1 - p_{in} \cdot \prod^{N} p) \tag{3.8}$$

Due to the light-weight characteristic of the proposed EDS deployment compared to TRC, more s-paths can be deployed to measure local PVT variations.

Nevertheless, one time of VOS might not trigger system failures due to the extra V steps of voltage margin of V_{DVS} relative to the minimum error-free operating voltage V_{safe} at a specific PVT condition. Thus the probability that the inactivation of K s-path with

N logic stages induces V times of consecutive VOS is

$$p_{N,M,K,V} = \prod^{V}\prod^{K}\prod^{M}(1 - p_{in} \cdot \prod^{N} p) = \prod^{K}\prod^{(M \cdot V)}(1 - p_{in} \cdot \prod^{N} p) \tag{3.9}$$

where $M \cdot V$ is the maximum allowable number of clock cycles for the s-paths being inactivated consecutively. Nevertheless, a larger X of the voltage counter can divide the DVS scope into finer steps and provide a larger V for the same amount of voltage margin. Besides countering VOS, sufficient voltage margin is also needed to counter sporadic transient variations or allow enough response time for the voltage regulator.

3.2.4 Design Methodology and Parameter Tuning

To realize the proposed system, two extra steps, denoted as "EDS deployment" and "DVS algorithm", are added to a conventional design methodology for an FPGA design, as shown in Fig. 3.4. For "EDS deployment", $t_{s,cp}$ and $t_{s,EDS}$ highly depend on Place and Route (PnR). In other words, if the EDS circuits are inserted into the "critical" paths relative to the clock falling edge or β' is modified slightly after one round of PnR, in the next round of PnR after the EDS circuit insertion, these paths may become "non-critical". This PnR problem is also observed by other works [1] [4]. Thus, an iterative cycle is a common case for the designs with EDS circuits [1] and performed as follows (An automatic placement tool [24] is valuable).

1. **Initialization/Modification:** For the initial stage, β and β' are differentiated and EDS circuits are assigned in various locations. For iterative stages, only β' and the locations of one or two EDS circuits are adjusted.

2. **Timing Evaluation:** After each successful compilation, K is checked. If expectations are not satisfied, go to 1);

3. **Validation:** Testing experiments using a specific file are carried out. V_{safe} and V_{DVS} are measured, at the manual mode of the PS configuration for that PVT condition

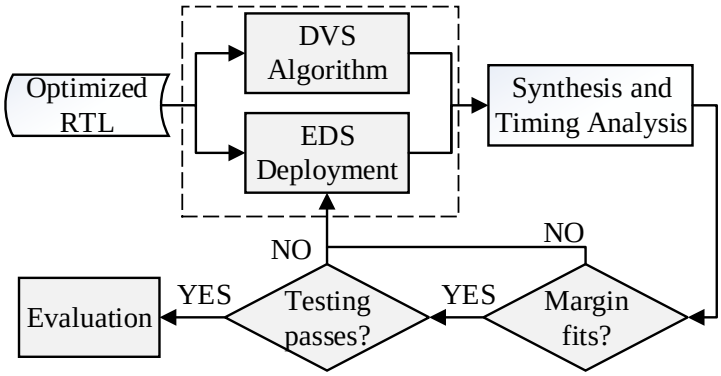


Figure 3.4: Design methodology for integrating EDS circuits and DVS systems into a standard FPGA design flow.

and at the DVS mode at the LR configuration, respectively. If expectations are not satisfied, go to 1);

4. **Evaluation:** The designs are tested with more input data.

The parameters of "DVS algorithm" are independent of PnR and require less tuning efforts.

3.3 Experimental Results

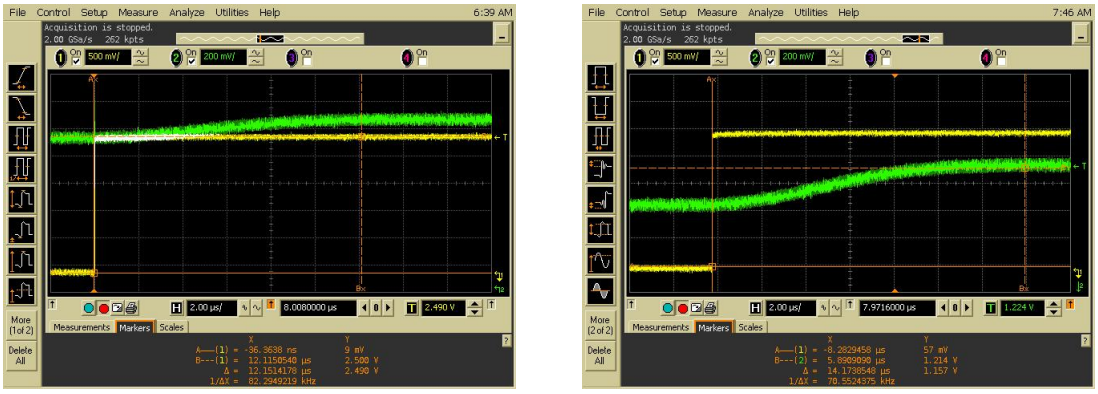
A DCT unit is chosen from [50], the core building block of JPEG as well as other video compressors, as a reference design to demonstrate the effectiveness of the proposed design. This DCT unit is fully pipelined and utilizes a parallel distributed arithmetic architecture with an 8-bit input and a 12-bit output bus width. 76% of the FPGA on-chip memory is allocated to input and output data storage. The original RTL code for the DCT is further optimized ("Baseline"). DVS and EDS circuits are added into "Baseline" as the "Proposed". For the optimized EDS circuit deployments, 8 EDS circuits are assigned to the ISBs of the 14-bit or 12-bit "+" adders outputs since these adders are the critical paths. The maximum number of N in Eq. (3.7) is 7. Frequency constraints and the actual clocks

are set as 200 MHz ($T = 5$ ns) that meets the requirements of real-time full high definition video processing applications. β and β' are set to 50% and 61% (obtained by using the above iterative tuning cycle), respectively.

Table 3.1: Voltage Level Division for the voltage regulator. "VLevel" stands for the 4-bit voltage counter. "Vctrl" stands for the 5 tri-state control signals (Vo_2, Vo_1,Vo_0, $MARGSEL$, $MARGTOL$) of the voltage regulator. "Voltage" is the output voltage calculated according to [5].

VLevel	VCtrl	Voltage
0	0Z0ZZ	0.950
1	0ZZ0Z	0.970
2	0ZZ00	0.990
3	0ZZZZ	1.000
4	0ZZ10	1.010
5	0Z10Z	1.019
6	0ZZ1Z	1.030
7	0Z100	1.040
8	0Z1ZZ	1.050
9	0Z110	1.061
A	0Z11Z	1.082
B	01000	1.089
C	010ZZ	1.100
D	0101Z	1.133
E	01ZZZ	1.150
F	011ZZ	1.200

The 4-bit ($X = 4$) voltage counter divides the voltage range 0.95 V to 1.20 V by 16 levels as shown in Tab. 3.1 according to [5]. $\Delta V_+ = \Delta V_-$ are both set as 1 level-step. $T_{regulator}$ is measured around 12 µs to 14 µs as follows. A simple voltage control circuit (not the to-be-used DVS circuit) controlled by the switches shifts the LR output from 1.10 V or 0.95 V to 1.20 V. $T_{regulator}$ is the signal delay δ from the location $p4$ (channel 1, yellow signals in Fig. 3.5) to the location $p2$ (channel 2, green signals in Fig. 3.5). ϕ_{cnt+} is set to 4096, which translates to $T_{cnt+} = 20.48$ µs. The $\alpha = \phi_{cnt-}/\phi_{cnt+}$ is set as 3 and thus $M = \phi_{cnt-}$ in Eq. (3.7) is 12288 ($T_{cnt-} = 61.44$ µs). So far, for the input data with



(a) (b)

Figure 3.5: $T_{regulator}$ measurement: (a) from $1.10V$ to $1.20V$ ($\delta = 12\,\mu s$) ; (b) from $0.95V$ to $1.20V$ ($\delta = 14\,\mu s$)

$p_{in} = 100\%$, $p_{N,M}$ for a single s-path in ϕ_{cnt-} clock cycles is 1.4e−42 which is 10^{12} times smaller than the metastability-failure probability in a microprocessor. If an EDS circuit was placed at the 14th bit of a ripple-carry adder, $p_{N,M}$ would be as large as 0.47.

Both designs are compiled by Quartus II 13.1 with the optimization for speed and the highest Fitter effort setting. As demonstrated in Table. 3.2, the proposed design only incurs 0.8% Logic Element (LE) penalties compared to the baseline. Nevertheless, here two factors need to be considered for understanding the LE penalties. One is that it is the choice of the FPGA PnR tool to implement the functionality using either combinational functions or logic registers for optimizations. The other is that even though the highest fitting effort is set, the difference of LE usage is pretty small and probably falls below the stopping threshold of the PnR algorithm.

At the "slow 1200 mV 85 °C" corner, the number K of s-paths is obtained in TimeQuest Timing Analyzer as shown in Fig. 3.6. In Fig. 3.6a, the second path is the most critical path relative to the clock rising edge and thus $t_{s,cp} = 0.317\,ns$, which also translates into its f_{max} in Tab. 3.2. Judged by Eq. (3.2), $t_{s,EDS}$ must be less than 0.7085 ns and among 183 EDS-monitored paths in Fig. 3.6b K is 30.

Lena, Ruler and Gray (Fig.3.7) from [51] with 512×512 8-bit gray-scale pixels are used

	Slack	From Node	To Node
1	0.244	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][0]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS3|data_falling_edge
2	0.317	MDCT:The_MDCT|RAM:U1_RAM|altsyncram:...ated|ram_block1a0~portb_address_reg0	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][3]
3	0.367	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][2]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS3|data_falling_edge
4	0.379	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][1]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS3|data_falling_edge
5	0.417	MDCT:The_MDCT|DCT1D:U_DCT1D|even_not_odd	MDCT:The_MDCT|DCT1D:U_DCT1D|dcto_1A[17]
6	0.418	MDCT:The_MDCT|DCT2D:U_DCT2D|even_not_odd	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS1|data_falling_edge
7	0.422	MDCT:The_MDCT|DCT1D:U_DCT1D|even_not_odd	MDCT:The_MDCT|DCT1D:U_DCT1D|dcto_1A[15]
8	0.427	MDCT:The_MDCT|DCT2D:U_DCT2D|even_not_odd~_Duplicate_14	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS1|data_falling_edge
9	0.445	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][4]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS3|data_falling_edge
10	0.459	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][3]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS3|data_falling_edge
11	0.469	MDCT:The_MDCT|RAM:U1_RAM|altsyncram:...ated|ram_block1a0~portb_address_reg0	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][2]
12	0.486	MDCT:The_MDCT|RAM:U1_RAM|altsyncram:...ated|ram_block1a0~portb_address_reg0	MDCT:The_MDCT|DCT2D:U_DCT2D|latchbuf_reg[7][8]

(a)

	Slack	From Node	To Node
18	0.626	MDCT:The_MDCT|ROMO:\G_ROM_ST2:5:U2_ROMO|datao[3]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS1|data_falling_edge
19	0.633	MDCT:The_MDCT|ROMO:\G_ROM_ST2:4:U2_ROMO|datao[2]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS1|data_falling_edge
20	0.634	MDCT:The_MDCT|DCT1D:U_DCT1D|even_not_odd	MDCT:The_MDCT|DCT1D:U_DCT1D|EDS:EDS7|data_falling_edge
21	0.646	MDCT:The_MDCT|DCT2D:U_DCT2D|even_not_odd~_Duplicate_2	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS2|data_falling_edge
22	0.652	MDCT:The_MDCT|DCT2D:U_DCT2D|dcto_1D[9]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS4|data_falling_edge
23	0.655	MDCT:The_MDCT|DCT2D:U_DCT2D|even_not_odd	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS2|data_falling_edge
24	0.657	MDCT:The_MDCT|DCT2D:U_DCT2D|dcto_1C[10]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS4|data_falling_edge
25	0.660	MDCT:The_MDCT|ROMO:\G_ROM_ST1:2:U1_ROMO|datao[2]	MDCT:The_MDCT|DCT1D:U_DCT1D|EDS:EDS6|data_falling_edge
26	0.670	MDCT:The_MDCT|ROME:\G_ROM_ST2:8:U2_ROME|datao[2]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS2|data_falling_edge
27	0.671	MDCT:The_MDCT|DCT1D:U_DCT1D|even_not_odd	MDCT:The_MDCT|DCT1D:U_DCT1D|EDS:EDS7|data_falling_edge
28	0.694	MDCT:The_MDCT|ROMO:\G_ROM_ST1:2:U1_ROMO|datao[3]	MDCT:The_MDCT|DCT1D:U_DCT1D|EDS:EDS6|data_falling_edge
29	0.695	MDCT:The_MDCT|ROME:\G_ROM_ST1:2:U1_ROME|datao[4]	MDCT:The_MDCT|DCT1D:U_DCT1D|EDS:EDS6|data_falling_edge
30	0.696	MDCT:The_MDCT|ROMO:\G_ROM_ST2:9:U2_ROMO|datao[1]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS2|data_falling_edge
31	0.720	MDCT:The_MDCT|DCT2D:U_DCT2D|dcto_1C[8]	MDCT:The_MDCT|DCT2D:U_DCT2D|EDS:EDS4|data_falling_edge

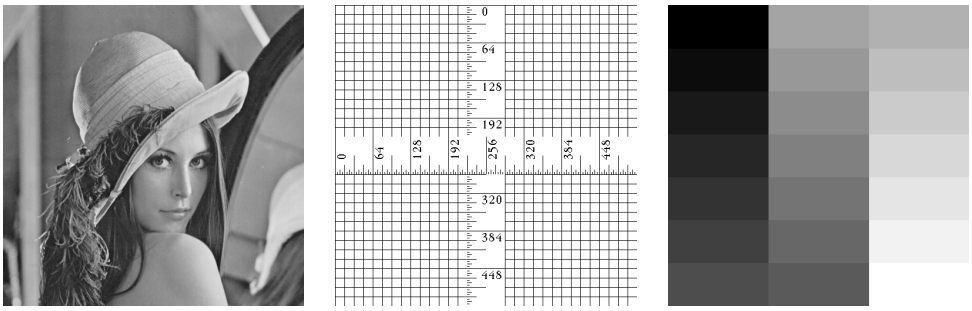
(b)

Figure 3.6: Timing Reports for : (a) Critical paths; (b) EDS-monitored paths.

Table 3.2: Compilation Results (Combinational functions and logic registers are inside Logic Elements (LE)).

Design	LE	Combinational functions	logic registers	f_{max}(MHz)
Baseline	4161	3860	1883	222
Proposed	4195	3899	1897	214
Difference	0.8%	0.4%	1.0%	-3.5%

for the evaluations. Each image is divided into 16 blocks each containing 512×32 pixels, downloaded to the FPGA memory, and finally processed repeatedly by the DCT circuits (The specific file for validations is Lena). An inverted DCT is then performed and the PSNR is computed in Matlab.



(a) (b) (c)

Figure 3.7: Images: (a) Lena, (b) Ruler and (c) Gray represent the input data with normal, fairly-low and extremely-low entropy, respectively.

V_{safe} for both the "Baseline" and "Proposed" is measured using Lena as 1.02 V and 1.04 V, respectively. V_{safe} of "Baseline" suggests 15% maximum headroom similar to that of 18% from [2]. V_{safe} of "Proposed" is increased by 2% as a result of f_{max} degradation. However, the remaining 7% residual margin of f_{max} (over the constraint clock) is explorable by DVS circuits and becomes one of our design advantages. Furthermore, the PSNRs around V_{safe} are measured and the processed images are presented in Fig. 3.8. Notice that PSNR is a widely-used metric for evaluating the quality of image compression and for

8-bit image compression. A PSNR greater than 40 dB is generally considered acceptable. This is demonstrated in Fig. 3.8 where Lena images with PSNR smaller than 40 dB show significant noise. The current and power of "Baseline" are measured at the nominal voltage of the PS configuration. Another useful finding from Table. 3.3 is that both "Baseline" and "Proposed" might be well balanced (i.e. optimized) since a sudden degradation of PSNR is observed around V_{safe}.

Table 3.3: PSNR vs. Operating Voltage around V_{safe} at the PS configuration.

Baseline		Proposed	
Voltage(V)	PSNR(dB)	Voltage(V)	PSNR(dB)
1.03	48.9	1.04	48.9
1.02	48.9	1.03	48.0
1.01	48.1	1.02	40.7
1.00	21.1	1.01	6.6

(a) (b) (c)

Figure 3.8: Lena at PSNR : (a) 40.7 dB; (b) 21.1 dB; (c) 6.6 dB.

At the LR configuration, "Proposed" is tested and the $SW(4)$ shifts the nominal mode to the automatic DVS mode. The signal waveform of V_{DVS} at location $p2$ shows that the DVS circuits can be properly initialized from 0.94 V to 1.10 V (Fig. 3.9a). Due to the negligibly-small $p_{N,M}$ as calculated, V_{DVS} of "Proposed" in evaluations are reliably scaled to around 1.10 V. This is illustrated by the stable waveform of the LR output voltage

in Fig. 3.9b. No VOS was observed for Lena and Ruler. The processed images of Lena and Ruler from both "Baseline" at the PS configuration and "Proposed" are equivalent and also show negligible differences from the original ones. This is also evident from the equally large PSNRs in Table. 3.4. Nevertheless, judging from Fig. 3.9b, it is unclear how the internal DVS circuit is working. To better demonstrate DVS , the control signals of LR for Lena and Ruler at room temperature are shown in Fig. 3.9c (corresponding to Fig. 3.9b) and Fig. 3.9d. The yellow, red, green and blue waveforms represents the lower 4 bits of the $VCtrl$ signals (the highest bit is always grounded as seen in Table. 3.1 and not measured). In Fig. 3.9c, $VCtrl$ toggles in $\{Z11Z, 1000, 10ZZ, 101Z, 10ZZ, 1000, Z11Z\}$ steadily and repeatedly. The V_{DVS} range is from 1.082 V to 1.133 V. Fig. 3.9d shows a different toggling style $\{Z11Z, 1000, Z11Z\}$. The V_{DVS} range is from 1.082 V to 1.089 V. Different toggling styles indicate that s-paths are capable of handling various types of data (i.e., being activated). It implies that besides monitoring the paths with less logic stages, more s-paths could help improving the reliability as analyzed in the Section. 3.2.3.

Most importantly, when the FPGA chip is heated by the hair dryer, V_{DVS} is automatically raised from 1.10 V to 1.13 V. When the chip is cooled down back using an ice bag, V_{DVS} is reduced. This procedure is video-recorded. Due to lack of this unique feature of "Proposed", "Baseline" is supposed to be only operated at the nominal voltage 1.20 V calculated by STA, neither V_{safe} nor V_{DVS}.

Assuming the circuits usually operates the data with normal entropy when calculating the energy consumption, "Proposed" achieves 8.3% V_{DVS} reduction and 16.5% energy (i.e., power) saving for both images compared to "Baseline" at the same clock-frequency setting, as shown in Table. 3.4.

For the image Gray that tests the unfavorable input data, however, one of its 16 blocks causes the V_{DVS} drop to 1.03 V (a VOS) (Fig. 3.10b), inducing a degradation of PSNR (Fig. 3.10a). Nevertheless, this test is carried out repeatedly on that file block. If the whole image can be continually fed to the "Proposed" design, the VOS will probably be eliminated. The LR output voltages when processing two other blocks of Gray are shown in Fig. 3.10c and Fig. 3.10d, showing different styles. This indicates that "proposed" is able to handle some of the extremely-low entropy data. Another finding is that the minimum

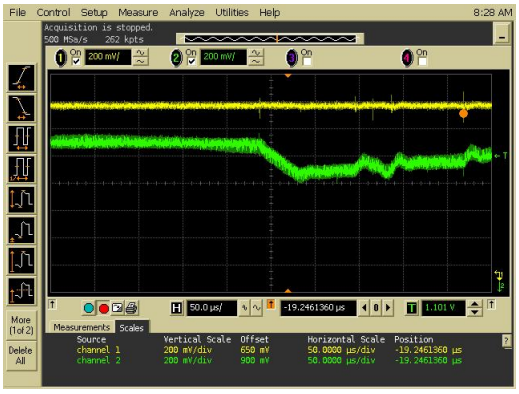
Table 3.4: Evaluations Results for Lena and Ruler.

Image	Design	PSNR (dB)	Voltage (V)	Current (mA)	Power (mW)
Lena	Baseline	48.9	1.20	0.188	0.226
	Proposed		1.10	0.171	0.188
	Savings	-	8.3%	9.0%	16.6%
Ruler	Baseline	53.6	1.20	0.148	0.178
	Proposed		1.10	0.135	0.149
	Savings	-	8.3%	8.8%	16.4%
Average Savings		-	8.3%	8.9%	16.5%

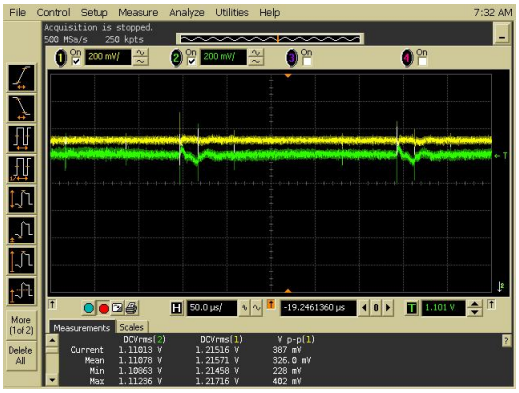
LR output voltage 1.049 57 V in Fig. 3.10c explains why there still needs to be a voltage margin for "proposed".

3.4 Conclusion

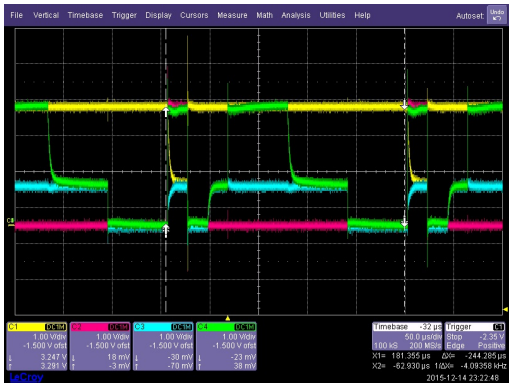
To solve the extrinsic EDS-reliability problem, this chapter proposes an EDS deployment strategy that requires neither buffer insertions nor extra clocks and speculatively and accurately detects slack-deficits. As a proof of concept, an FPGA-based DCT unit with EDS circuits and a DVS system is realized and evaluated with realistic data. On average, compared to just a DCT design, the proposed work produces equivalent outputs and achieves 16.5% energy saving at the same clock-frequency setting, with a 0.8% logic element and 3.6% maximum-frequency penalties. The proposed non-application-specific strategy can be generalized for DSP circuits that process normal-entropy data.



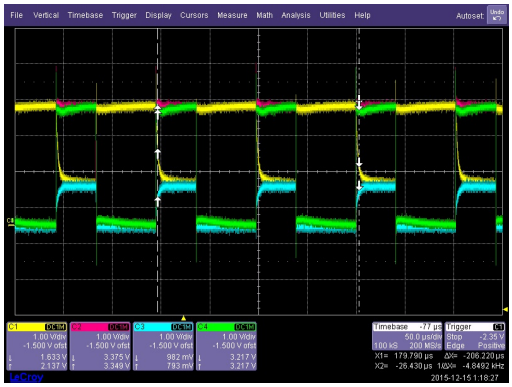
(a)



(b)

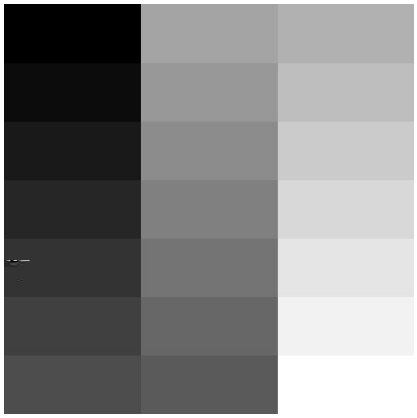


(c)

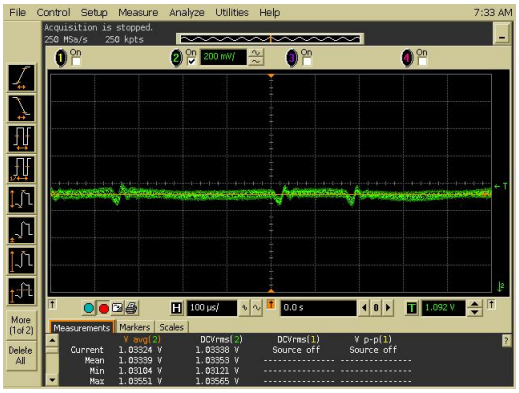


(d)

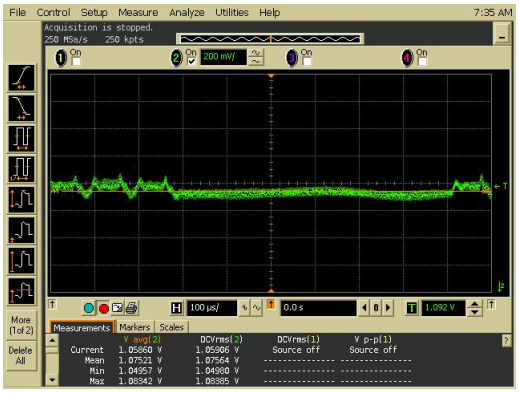
Figure 3.9: Oscilloscope Waveforms: the LR output supply voltage for (a) the initialization of DVS for Lena and (b) the normal status of DVS for Lena; the LR control signals for (c) Lena and (d) Ruler.



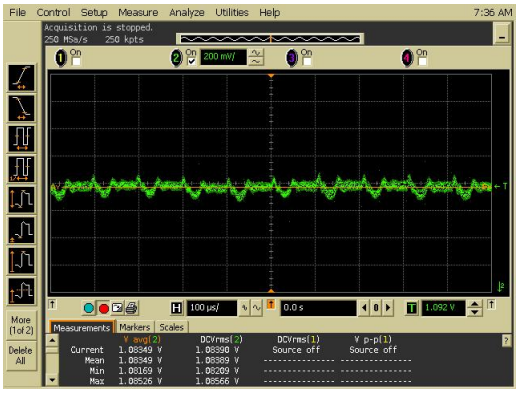
(a)



(b)



(c)



(d)

Figure 3.10: Evaluations for Gray: (a)the processed Gray with some scratches with the PSNR of 43.0 dB and three patterns of supply voltage when processing Gray : (b) $V_{DVS} = 1.033\,\mathrm{V}$ which is a VOS and causes the scratches; (c) $V_{DVS} = 1.075\,\mathrm{V}$; (d) $V_{DVS} = 1.083\,\mathrm{V}$.

Chapter 4

Design for Metastable-Hardened Voltage-Boosted Synchronizers

VBSs are proposed in this chapter to solve the intrinsic reliability problem of EDS circuits ("synchronization metastability") mentioned in Chapter 2. Section 4.1 introduces the proposed synchronizers and describes a new methodology of metastability parameter extraction. Section 4.1.4 presents and compares the simulation results for the baseline and proposed synchronizers. Final section draws the conclusions.

4.1 Voltage-Boosted Synchronizer Design

This work proposes VBSs (Fig. 4.1) to improve the FOM τ of synchronizers in low-voltage supply environments. A VBS integrates a Jamb latch with a charge pump implemented by using switched capacitors [52] [53]. The proposed synchronizers provide a temporary voltage boost to the latching element so as to improve its metastability resolution time. The charge pump works in two phases, *Precharging* and *Powering*. This characteristic is naturally matched with the transparent and latching modes of the Jamb latch, which is further controlled by the clock high and low phases. During the precharging (transparent) phase, Jamb latch receives new data and the charge pump is precharged by PMOS $P1$.

During the powering (latching) phase, the control signal b turns on PMOS $P2$ and turns off $P1$. The voltage at the $P2$ side of the capacitor C_b is boosted up and the voltage V_{bst} at the $P1$ side of C_b rises beyond V_{dd}, thus speeding up the metastability resolution of the CCIs. $P1$ is drain-body connected to force the current unidirectionally to flow from the supply rail to the charge pump. The proposed designs have two major advantages:

1. The characteristics of *double transconductors* and *self switch-off* persist since the CCIs of the Jamb latch remains.

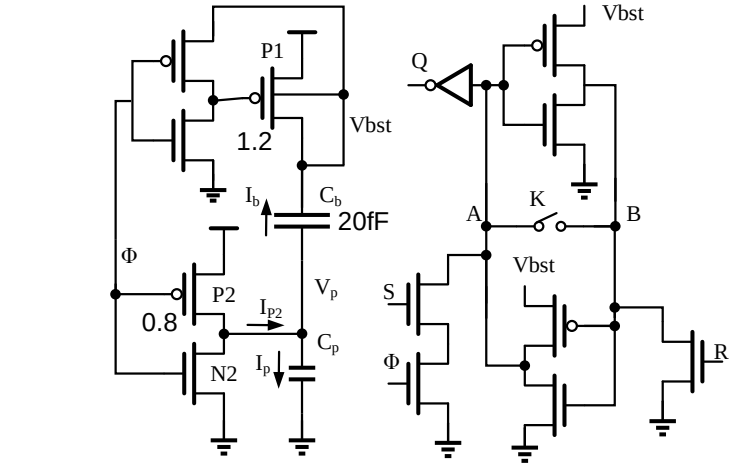
2. The charge pump can be sized arbitrarily by designers to provide a large current and sufficient electrical charge for powering the CCIs so as to meet design specifications without a direct impacting on the critical nodes of Jamb latch. Meanwhile, Jamb latch can use small-sized transistors to minimize the power consumption. This characteristic avoids the constraint of the criterion 2 in Chapter 2.

4.1.1 Consistency between Jamb Latch and Charge Pump

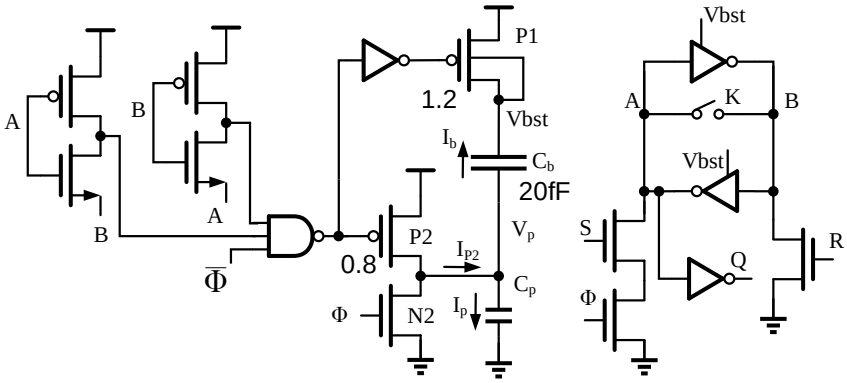
The Jamb latch and the charge pump are consistent with each other in both timing and voltage domains.

Timing Consistency The charge pump works in two phases, *Precharging* and *Powering*. This characteristic is naturally matched with the transparent and latching phases of the Jamb latch, which is further easily controlled by the clock high and low phases. During the precharging (transparent) phase, the Jamb latch receives new data and the charge pump is precharged by PMOS $P1$. During the powering (latching) phase, the control signal b turns on PMOS $P2$ and turns off $P1$.

Voltage Consistency The voltage at the $P2$ side of the capacitor C_b is boosted up and the voltage V_{bst} at the $P1$ side of C_b rises beyond V_{dd}, thus speeding up the metastability resolution of the CCIs. Unlike other logic circuits where low voltage circuits cannot drive



(a)



(b)

Figure 4.1: VBSs: (a) CVBS, (b) MVBS.

high voltage ones, the cross-coupled inverters can be powered by a supply voltage higher than the voltage level of the input signal due to the cross-coupled feedback effect. An example is a level-converting flip-flop.

4.1.2 The Working Mechanism of Charge Pump

Powering Strategies of Charge Pump

Two Powering strategies of the charge pump on the powering phase have been developed, including CVBS (c.f. Fig. 4.1a) and MVBS (c.f. Fig. 4.1b). CVBS powers the CCIs every powering phase and MVBS powers the CCIs only when the CCIs are metastable. MVBS utilizes the metastability detector to detect the metastable status and generate the control signal M for the input b of the charge pump. The supply voltage of the inverter driving PMOS $P1$ is connected with V_{bst} to perform the same function as the $P1$ drain-body connection does. During the powering phase, the synchronizer is targeting a specified N_r for the metastable resolution and possibly enters the following situations:

Situation 1: There is no metastability. This occupies the majority of all the possible situations. The charge pump of MVBS will not be triggered, thus no extra energy is wasted. However, CVBS will raise V_{bst} and the transistors including $P1$, $P2$ and CCIs will dissipate dynamic energy due to the raised V_{bst}. Nevertheless, the energy loss in these small inverters is much smaller than that of the fully switched-on the GNDED .

Situation 2: Metastability is unresolved before N_r is reached. To guarantee the overall metastability resolution performance, τ should be maintained smaller than those of the baseline synchronizers by boosting V_{bst}.

Situation 3: Metastability is resolved before N_r is reached. The Jamb latch switches off but still consumes the leakage current. This current is provided by both the charge pump and the small current of $P1$ when $V_{bst} < V_{dd}$ in CVBS, or, by the switched-on $P1$ in MVBS. The stability for this situation is determined by other failure factors such as SEU [12] (p346). In this sense, MVBS is more flexible to frequency scaling than CVBS.

Situation 4: Metastability is unresolved after N_r is reached. Though electrical charge on C_b may be consumed further and possible run out due to the *switched-on* CCIs, the N_r specification for this stage is already met. The subsequent synchronizing stages (if they exist) will provide more metastability resolution time.

In summary, MVBS targets low power and high flexibility for frequency scaling while CVBS targets high performance.

The Boosting Mechanism of Capacitor

According to KCL, the current flowing through $P2$ is

$$i_{P2}(t_r) = i_b(t_r) + i_p(t_r) \tag{4.1}$$

where $i_b(t_r)$ and $i_p(t_r)$ are the currents flowing through C_b and C_p, respectively. The voltage $V_b(t_r)$ across C_b is determined by

$$V_b(t_r) = V_{b0} - \frac{1}{C_b}\int_0^{t_r} i_b(t)\,\mathrm{d}t \tag{4.2}$$

where V_{b0} is the initial voltage across C_b (in this work V_{b0} is V_{dd}). The voltage $V_p(t_r)$ across C_p is determined by

$$V_p(t_r) = V_{p0} + \frac{1}{C_p}\int_0^{t_r} i_p(t)\,\mathrm{d}t \tag{4.3}$$

where V_{p0} is the initial value of $V_p(t_r)$ (in this work V_{p0} is near 0 V). Normally, during the powering phase, the condition holds

$$V_p(t_r) \leq V_{dd}. \tag{4.4}$$

The output voltage $V_{bst}(t_r)$ of the charge pump is given by

$$V_{bst}(t_r) = V_b(t_r) + V_p(t_r) \tag{4.5}$$

Situation 2 requires

$$V_{bst}(t_r) > V_{dd}. \tag{4.6}$$

This requirement is combined together with condition (Eq. (4.4)) to produce

$$V_b(t_r) > 0. \tag{4.7}$$

That is

$$\int_0^{t_r} i_b(t)\,\mathrm{d}t \leq C_b V_{b0} \tag{4.8}$$

where $Q_{b0} = C_b V_{b0}$ is the total electrical charge stored during the precharging phase. In other words, Q_b provides the entire electrical charge needed for the metastability resolution and C_b should be sized according to the desired τ and N_r.

Eq. (4.5) can be differentiated to obtain $V_{bst}(t_r)$ derivative over t_r

$$\frac{dV_{bst}}{dt_r} = \frac{i_p(t_r)}{C_p} - \frac{i_b(t_r)}{C_b} \tag{4.9}$$

$$= \frac{i_{P2}(t_r)}{C_p} - i_b(t_r)(\frac{1}{C_b} + \frac{1}{C_p}) \tag{4.10}$$

$$\simeq \frac{1}{C_p}[i_{P2}(t_r) - i_b(t_r)] \tag{4.11}$$

assuming $C_b \gg C_p$. This indicates that when $i_{P2}(t_r)$ is greater than $i_b(t_r)$, $V_{bst}(t_r)$ is raised to increase the load current or else $V_{bst}(t_r)$ is reduced to reduce the load current. In other words, *the charge pump maintains the* $i_{P2}(t_r) \simeq i_b(t_r)$ *relationship by adapting* V_{bst} *to adjust the load current* $i_b(t_r)$. Giving that $C_b \gg C_p$, $P2$ negatively charges C_b, forcing $V_b(t_r)$ to drop from V_{dd} towards 0 during the powering phase.

So far a detailed analysis of the voltage-boosting mechanism is demonstrated. More importantly, a qualitative guideline for iterative transistor-sizing during simulations is obtained based on the analysis. The capacitor C_b should be sized large to provide sufficient electrical charges. $P2$ transistor should sized large enough to provide a large current. $P1$ transistor should be sized larger so as to precharge the capacitor in time. $N2$ transistors can be sized properly for discharging the electrical charge stored in C_p. Nevertheless, a

quantitative guideline for transistor sizing would require more comprehensive analysis, such as the relationship between $i_b(t_r)$ and τ. This leaves for future work.

4.1.3 Metastability Parameters Calculation

The conventional methodology for τ extraction assumes a constant bias current and a constant τ during the metastability resolution for the ORI and the GNDED, however, the assumption is not the case for GATED and VBSs with temporally changing V_{bst} and τ. Thus here this work has developed a new methodology for calculating t_r and the average $\bar{\tau}$ for a given N_r.

(1) **Metastability Simulations**. τ is extracted using the mentioned methodology parametrically over V_{bst}. In each simulation, the charge pump is replaced with an extra voltage source that provides a constant V_{bst} while the metastability detector remains in MVBS. The extracted simulation data builds a mapping function between V_{bst} and τ

$$\tau = \tau(V_{bst}) \tag{4.12}$$

Similarly, for the baseline GATED, the control signal of two PMOS current sources is disconnected from the metastability detector and connected to ground or V_{dd} to extract τ_{on} and τ_{off}, respectively.

(2) **Voltage-Boosting Simulation** . The temporally changing V_{bst} is extracted. Opposite from metastability simulations, the switch K is first opened to initialize VBSs. Then K is closed to force CCIs into the persisting metastable status. Thus the V_{bst} response (including the delay t_{MD} of metastability detector in MVBS) is evaluated as

$$V_{bst} = V(t_r) \tag{4.13}$$

Similarly, the t_{MD} of the GATED is extracted.

(3) N_r **Integration**. N_r is obtained by integrating the inversed τ over t_r as a function

$N(t_r)$

$$N_r = \int_0^{t_r} \frac{1}{\tau(V_{bst}(t))}\,\mathrm{d}t = N(t_r) \tag{4.14}$$

For constant V_{bst} and τ, Eq. (4.14) degenerates to Eq. (2.6).

(4) t_r **and** $\bar{\tau}$ **Calculation**. t_r for a given N_r is calculated using the inverse function of Eq. (4.14)

$$t_r = N^{-1}(N_r) \tag{4.15}$$

The t_r of the GATED is calculated as

$$t_r = t_{MD} + (N_r - \frac{t_{MD}}{\tau_{off}})\tau_{on}. \tag{4.16}$$

Assuming fixed N_r and τ_{off}, t_r is bottlenecked by t_{MD} and τ_{on}. $\bar{\tau}$ is calculated using Eq. (2.6). Notice that Eq. (4.14) and Eq. (4.15) already implicitly includes t_{MD}.

4.1.4 Simulation Results

To intuitively illustrate τ (FOM of synchronizers), our simulations will first set a 35τ N_r specification for all synchronizers and then simulate t_r and t_d of synchronizers. The synchronizer circuits were simulated with the Cadence environment in Taiwan Semiconductor Manufacturing Company (TSMC) 65 nm technology. The ideal capacitor C_b is 20 fF (notice that C_p is the parasitic capacitance in the transistors); the widths of $P1$ and $P2$ are 0.8 µm and 1.2 µm, respectively, for both CVBS and MVBS. All other transistors are minimum-sized.

Simulations were first carried out at $V_{dd} = 0.7$ V. In Fig. 4.2, the trend of the curves indicates that as V_{bst} becomes larger, τ becomes smaller due to a larger overdrive voltage. The gap between CVBS and MVBS curves shows that the metastability detector deteriorates τ. Fig. 4.3a shows that in the powering phase the charge pump raises V_{bst} to power the CCIs and is precharged in the precharging phase. Fig. 4.3b demonstrates the bias currents flowing through the CCIs of the five synchronizers. MVBS and CVBS need

smaller bias currents than the GATED and GNDED do. The feedback response time of MVBS is larger than that of GATED due to the charge pump. However, due to the PMOS transconductors, MVBS and CVBS have better τ than GATED and GNDED, as shown in Fig. 4.4. In Fig. 4.5, CVBS and MVBS reach the specified N_r (such as 35) earlier than the GNDED and GATED do, respectively.

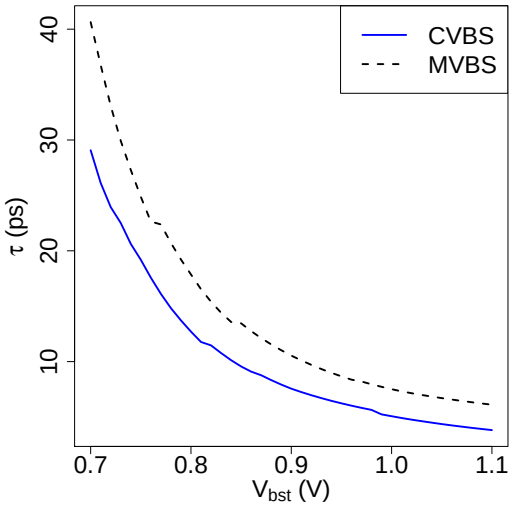


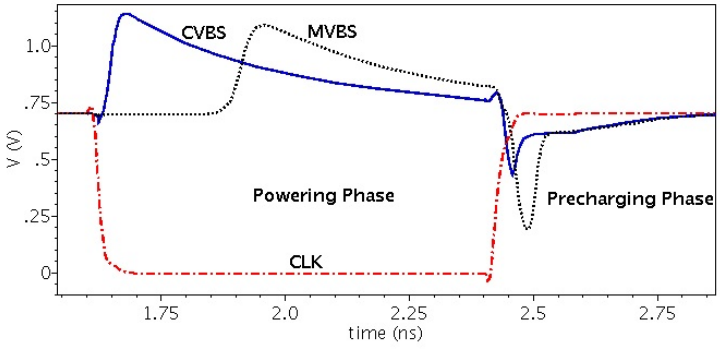
Figure 4.2: τ for VBSs at $V_{dd} = 0.7\,\mathrm{V}$.

Table 4.1: The Performance of the GATED

Vdd (V)	0.7	0.6	0.5	0.4
τ_{off} (ps)	46	125	448	1086
τ_{on} (ps)	18	24	42	143
t_{MD} (ps)	193	477	1430	4360

Simulations are carried out parametrically for $V_{dd} = \{0.4\,\mathrm{V}, 0.5\,\mathrm{V}, 0.6\,\mathrm{V}, 0.7\,\mathrm{V}\}$. The energy of each synchronizer is calculated by integrating the power (including that of the charge pump) with respect to the corresponding clock period $2t_d$. The results are shown in Tab. 4.1 and Tab. 4.2. Some conclusions can be drawn as follows.

(1) Performance. The r_0 column represents the performance (frequency) ratio of the corresponding synchronizer over the basic Jamb latch. The r_0 for the GATED and MVBS



(a)

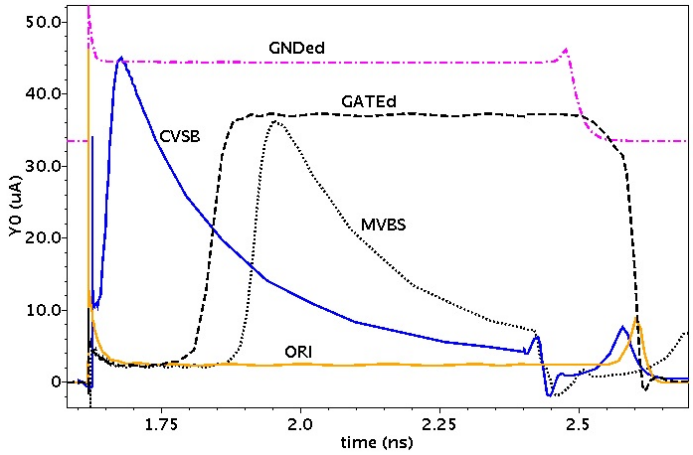


Figure 4.3: Voltage-boosting simulations at $V_{dd} = 0.7\,\mathrm{V}$: (a) V_{bst}; (b) biasing currents.

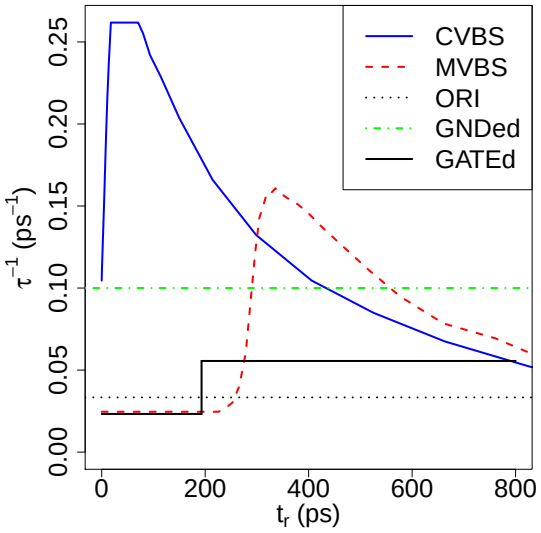


Figure 4.4: τ^{-1} vs t_r at $V_{dd} = 0.7\,\mathrm{V}$.

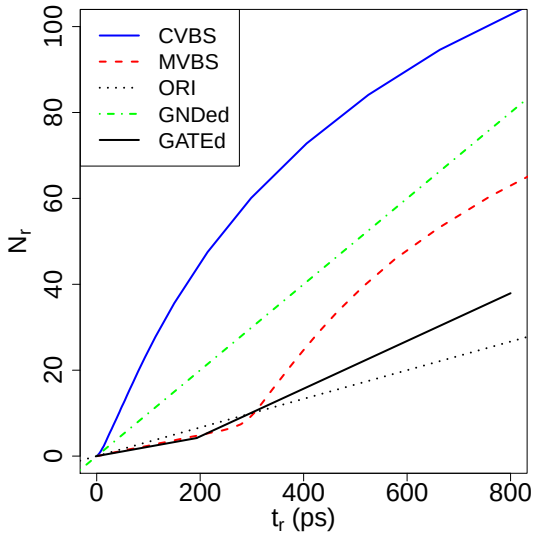


Figure 4.5: N_r vs. t_r at $V_{dd} = 0.7\,\mathrm{V}$.

are 1.3 to 2.2 and 2.0 to 2.7, respectively. The r_0 for the GNDED and CVBS shows 3.0 to 7.9 and 5.1 to 9.8, respectively. The r_1 column represents the performance (frequency) ratio of MVBS over the GATED or CVBS over the GNDED. The r_1 values show that MVBS and CVBS are 1.12 to 1.49 and 1.19 to 1.73 faster than the GATED and the GNDED, respectively.

(2) Energy. The energy of each synchronizer is calculated by integrating the power (including that of the charge pump) with respect to the corresponding clock period $2 \cdot t_d$. The other synchronizers consume a similar amount of E_{norm} except the GNDED consumes 4 to 5 times more. E_{idle} of the ORI and GATED are the least primarily because of low clock energy. However, 71% to 95% of E_{idle} of MVBS comes from the clock inverter (in Fig. 4.1b) that is usually also required by practical master-slave flip-flop designs. Thus this work believes that ORI, GATED and MVBS consume the same level of idle energy. E_{idle} of CVBS and GNDED are 5 to 8 and 30 to 50 times of that of MVBS, respectively.

Table 4.2: The Energy Consumption and Performance of the Baseline and Proposed Synchronizers with a given $N_r = 35$

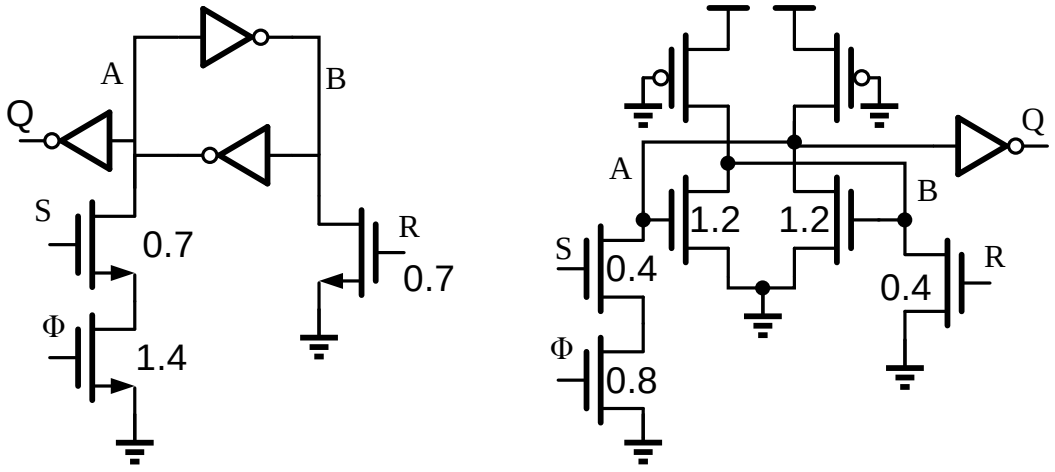
Vdd (V)	Energy (fJ)		Delay(ps)					
ORI								
Vdd	E_{idle}	E_{norm}	τ	t_r	t_n	t_d	r_0	r_1
0.7	0.02	1.91	31	1085	98	1183	1.00	
0.6	0.03	1.64	86	3010	180	3190	1.00	
0.5	0.05	1.36	206	7210	445	7655	1.00	
0.4	0.07	0.89	566	19810	1578	21388	1.00	
GNDED								
Vdd	E_{idle}	E_{norm}	τ	t_r	t_n	t_d	r_0	r_1
0.7	19.49	20.82	10	350	50	400	2.96	
0.6	13.19	14.04	13	455	78	533	5.99	
0.5	8.94	9.42	23	805	161	966	7.93	
0.4	6.51	6.76	80	2800	560	3360	6.37	
GATED								
Vdd	E_{idle}	E_{norm}	$\bar{\tau}$	t_r	t_n	t_d	r_0	r_1
0.7	0.02	3.89	21	747	154	901	1.31	
0.6	0.02	3.18	35	1225	287	1512	2.11	
0.5	0.03	2.54	79	2766	719	3485	2.20	
0.4	0.06	1.64	251	8791	2522	11313	1.89	
MVBS								
Vdd	E_{idle}	E_{norm}	$\bar{\tau}$	t_r	t_n	t_d	r_0	r_1
0.7	0.42	3.30	14	477	126	603	1.96	1.49
0.6	0.32	2.64	27	961	225	1186	2.69	1.27
0.5	0.24	2.02	71	2498	534	3032	2.52	1.15
0.4	0.19	1.33	235	8229	1882	10111	2.12	1.12
CVBS								
Vdd	E_{idle}	E_{norm}	$\bar{\tau}$	t_r	t_n	t_d	r_0	r_1
0.7	2.93	4.12	4	147	85	232	5.11	1.73
0.6	2.15	3.41	6	218	151	369	8.65	1.44
0.5	1.49	2.44	12	423	357	780	9.81	1.24
0.4	0.96	1.65	48	1667	1163	2830	7.56	1.19

4.2 Improved Voltage-Boosted Synchronizers

The VBSs are improved in several perspectives: synchronizer optimizations, simulation methodology improvement, and post-layout simulations.

4.2.1 Synchronizer Improvements

Schematic Optimizations The schematic Optimizations are made in four aspects.



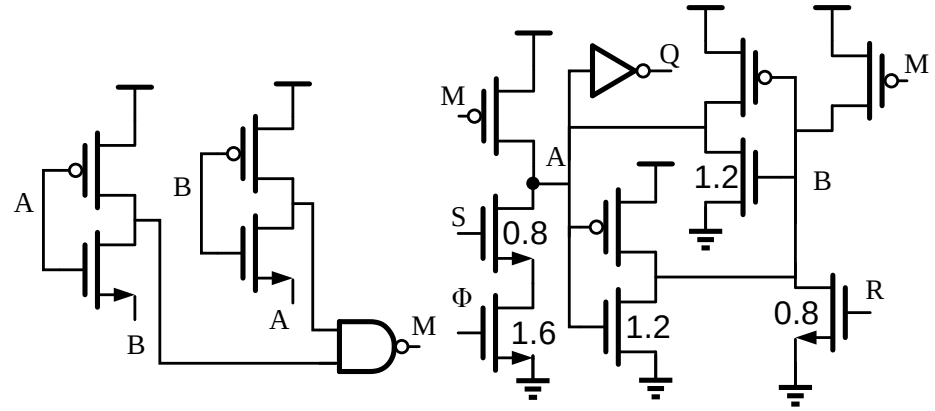
(a) (b)

Figure 4.6: Improved Synchronizers (sizing in µm): (a)ORI; (b) GNDED.

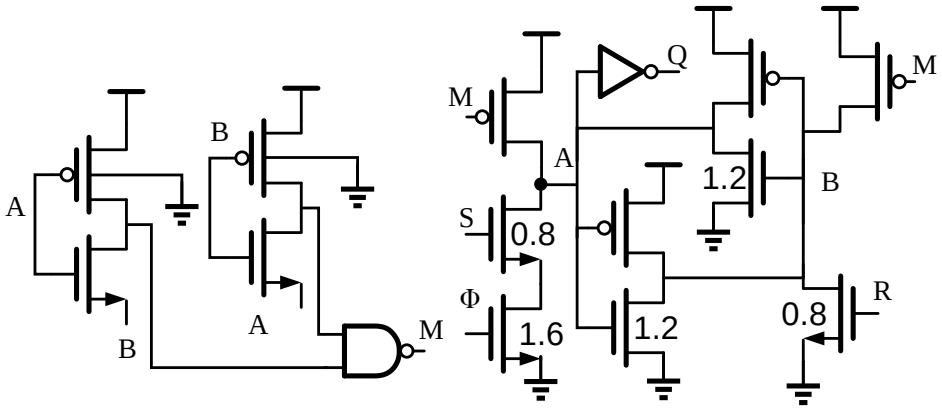
(1)The capacitor C_b is replaced with a PMOS and an NMOS transistor as seen in Fig. 4.8 and Fig. 4.9. This enables the feasibility of the layout and fabrication of VBSs.

(2)For CVBS, the signal V_{bst} is disconnected from the body of $P1$ transistor and the source of a driving PMOS transistor. By doing so, the improved CVBS is more compatible with standard cell design.

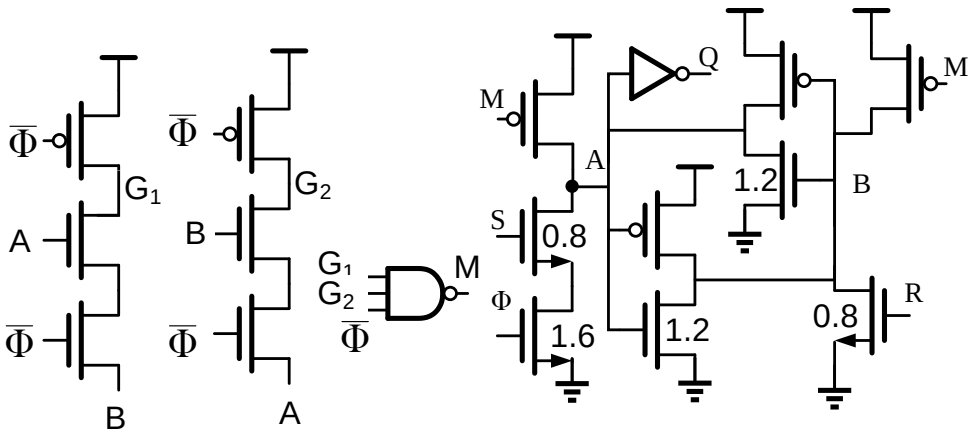
(3) The propagation chain of the metastability detection signal is slightly modified so as to first turn off the $P1$ transistor before the $P2$ transistor boosting the capacitors, as seen in Fig. 4.8 and Fig. 4.9. Though schematic simulations shows less differences, this



(a)



(b)



(c)

Figure 4.7: Improved Synchronizers (sizing in µm): (a)GATED_sta; (b) GATED_fbb; (c) GATED_dyn.

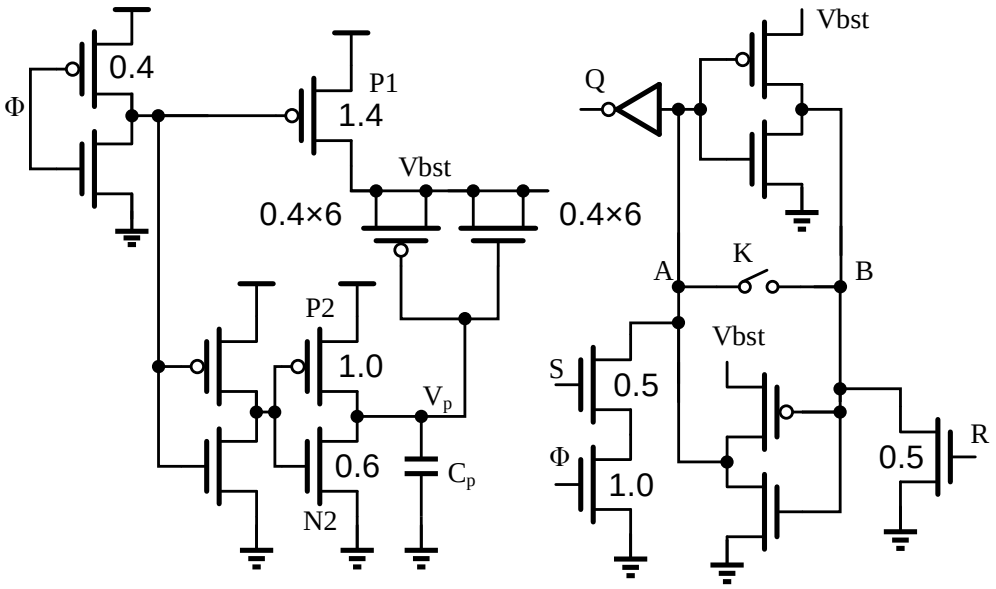
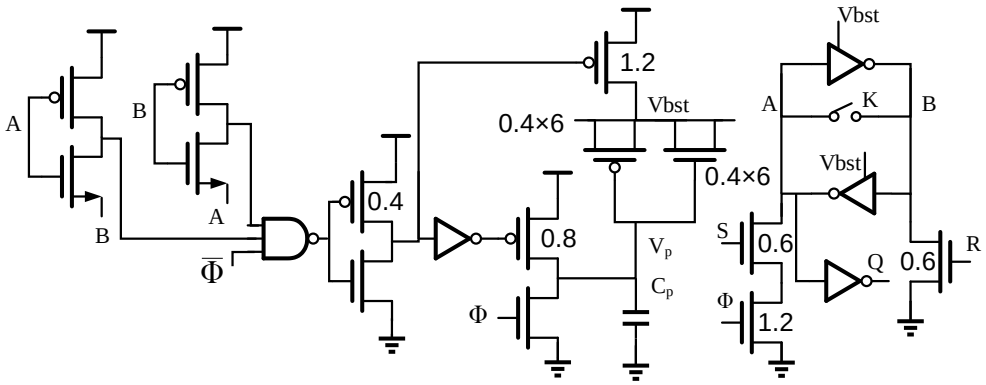


Figure 4.8: Improved Synchronizer CVBS (sizing in µm).

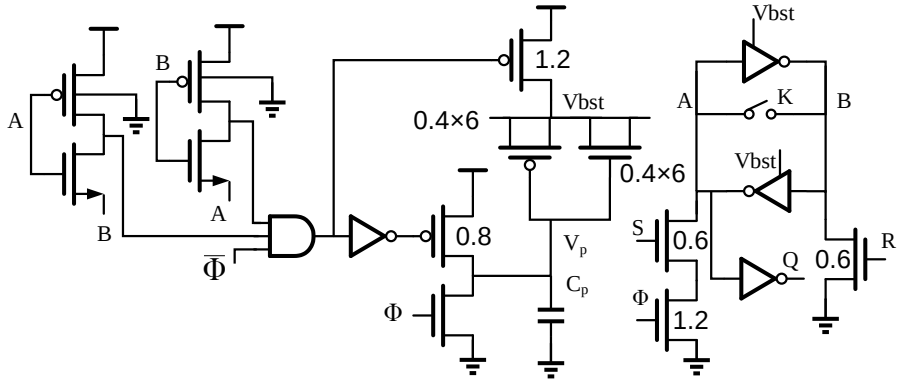
new arrangement does avoid significant charge loss in post-layout simulations. This is similar to generating non-overlapping clocks for switched-capacitor circuits. Nevertheless, this arrangement induces one buffer delay to t_{MD}, which is acceptable.

(4) All the synchronizer transistors are properly sized to meet at least two requirements, corner analysis (for all synchronizers) and charge pump fullness (for VBSs). In corner analysis, the three minimum-sized driving NMOS transistors corresponding to the signals S, Φ and R may not be able to drive CCIs into the expected states at slow NMOS corners. In other words, t_n is infinity in these situations. Hence, theses driving NMOS transistors needs to be sized large enough as shown in each schematic figures. For GNDED and GATED, since PMOS transistors are added as the current sources, nodes A and B of CCIs with minimum-sized NMOS transistors may both simply stuck at the supply voltage level at the corner of slow NMOS and fast PMOS. Hence, these NMOS transistors needs to be sized large for nodes A and B to be able to resolve to the ground voltage level.

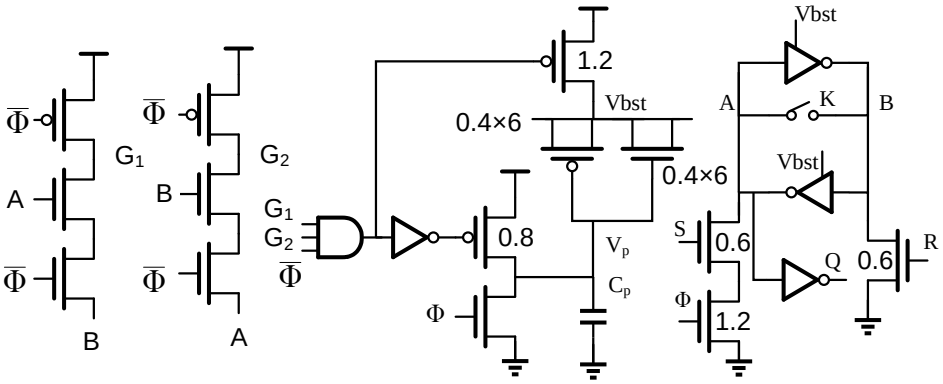
VBSs should (at least nearly) fully pre-charge the capacitor and maintain sufficient



(a)



(b)



(c)

Figure 4.9: Improved Synchronizers (sizing in µm): (a) MVBS_sta; (b) MVBS_fbb; (c) MVBS_dyn.

charge for powering. This fullness $F_{1,2}$ is evaluated as

$$F_{1,2} = \frac{V_{bst}(t_d)}{V_{dd}} \tag{4.17}$$

where $V_{bst}(t_d)$ is V_{bst} at the end of the pre-charging (F_1) or powering (F_2) phases, i.e., at t_d. For VBSs, the PMOS and NMOS transistors used as capacitors are sized as ($6\,\mu m \times 0.4\,\mu m$). Notice that a pair of these transistor is similar to a capacitor of 60 fF. $P1$, $P2$ and $N2$ sizings are modified. *Transistor sizings (widths in* µm*) are illustrated in the corresponding schematics. Transistors without specified sizing are minimum-sized.*

(5)Most importantly, the meta-detectors are optimized by applying FBB to the two PMOS transistors of the meta-detector or dynamic logic implementation of the meta-detector. Previous work applied FBB to the driving transistors and CCIs of Jamb latch. FBB in this work adds less leakage power and precludes the triple-well technology for NMOS FBB. More importantly, as our simulations results demonstrate, t_{MD} significantly bottlenecks t_r. Applying Amdahl's law to Eq. (4.16), more optimization effort should be put on the meta-detector. The meta-detector is essentially an analog XOR gate that compares the inputs A and B. The speed of the meta-detector is related to the common-mode voltages V_A and V_B of nodes A and B and its PMOS speed. Notice that transistor sizing for optimizing CCIs adds a large capacitance to the critical nodes. Thus this method is not considered in this work. MVBS with static, FBB and dynamic meta-detectors are added suffixes _sta, _FBB and _dyn, respectively. In other words, there are three MVBSs: MVBS_sta, MVBS_FBB and MVBS_dyn. Similarly, there are three GATEDs: GATED_sta, GATED_FBB and GATED_dyn.

Methodological Accuracy As mentioned previously, τ may be inaccurate due to the extraction method. Thus a fine-grained simulation for the metastability resolution region is proposed here. It is performed as a two-round simulation. The first round is a coarse-grained simulation. The metastability resolution time t_r is extracted using a clock of a predefined large period. The second round is fine-grained simulations. The simulator is forced to execute multiple predefined steps during t_r to extract enough data points (in

my case it is to set "strobeperiod" parameter in the Spectre simulator). Without this fine-grained simulation, the curve of V_{A-B} in Fig. 4.10 will not be smooth. This method is similar to the "timing step control" in [27]. More importantly, τ extracted using Eq. (2.17)

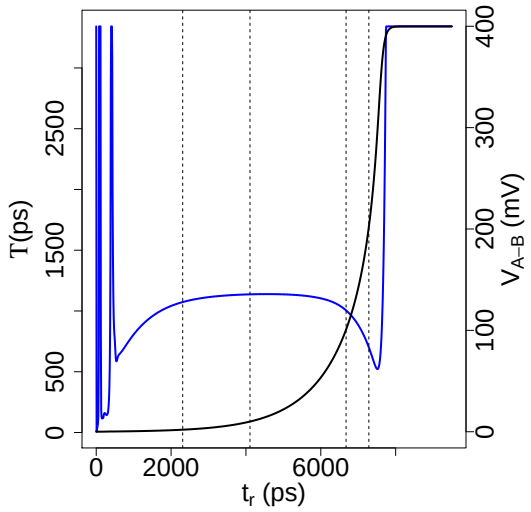


Figure 4.10: The curves of V_{A-B} (the black one) and $T(t_r)$ (the blue one).

should clearly specify the voltage region where τ is extracted, i.e., the values $V_{A-B,1}$ and $V_{A-B,2}$ need to be carefully chosen. Besides theoretical analysis in [30], this work provides a simple simulation method to determine $V_{A-B,1}$ and $V_{A-B,2}$ by calculating the derivative of V_{A-B}.

$$T(t_r) = \frac{dt_r}{dln(V_{A-B}(t_r))}. \tag{4.18}$$

$T(t_r)$ is plotted in Fig. 4.10. Based on the observations, this work deliberately chooses $V_{A-B,1} = 10\,\text{mV}$ and $V_{A-B,2} = 100\,\text{mV}$ where $T(t_r)$ is flat for later simulations. A brief explanation for this is that under 10 mV the noise (thermal or numerical) is dominant and beyond 100 mV the constant factor is dominant in the precisely-expanded function of $V_{A-B}(t_r)$.

By doing these two improvements, τ values in this work can be much more accurate.

Table 4.3: Improvement Ratio of t_d over the basic Jamb latch.

Voltage (V)	ORI	GNDED	GATED			CVBS		MVBS					
			sta	fbb	dyn			sta		fbb		dyn	
	r0					r0	r1	r0	r1	r0	r1	r0	r1
0.4	1.00	6.48	3.70	4.11	4.46	8.13	1.26	3.98	1.08	5.51	1.34	7.57	1.70
0.5	1.00	7.45	3.96	4.64	5.14	11.66	1.57	3.75	0.95	5.72	1.23	7.96	1.55
0.6	1.00	5.47	3.11	3.63	3.80	7.46	1.36	3.05	0.98	4.40	1.21	5.27	1.39
0.7	1.00	3.38	2.20	2.41	2.40	4.14	1.23	2.27	1.03	2.81	1.16	2.97	1.24

4.2.2 Schematic Simulation Results

Schematic simulations are carried out parametrically for $V_{dd} = 0.4\,\text{V}, 0.5\,\text{V}, 0.6\,\text{V}, 0.7\,\text{V}$ at the typical process corner and temperature 27 °C. Fig. 4.11a shows the t_w values of these synchronizers at 0.4 V vary little among these synchronizers and have little variance impact on MTBF. Fig 4.11 shows the trends of timing parameters of each simulated synchronizer. Table. 4.3 shows the performance (frequency) ratios $r0$ of GATED, MVBS, GNDED and CVBS over the ORI and the performance (frequency) ratio $r1$ of MVBS over GATED and CVBS over GNDED.

Fig. 4.11 illustrates the comparison of the timing parameters of each synchronizer where $t_d = t_n + 35 \cdot \tau$. These figures can be viewed together with Table. 4.3. The data shows that the optimizations on metastability detectors effectively improve the synchronizer delays, especially for VBSs. For 35τ N_r specification, in terms of speed among VBS, CVBS is the fastest, followed by MVBS_dyn, MVBS_fbb and MVBS_sta.

The Fullness of charge pumps in VBSs at the clock rising or falling edges are shown in Table. 4.4. Several observations can be made: (1) For three MVBSs, the faster the synchronizer is, the smaller its $F1$ value is. This is due to the less time for precharging. However, CVBS shows good values of $F1$ because its precharging transistors $P1$ and $N2$ are sized larger. (2) For all VBSs, the faster the synchronizer is, the smaller its $F2$ value is. This is because the faster synchronizer consumes more charge from its charge pump. (3)The higher the supply voltage is, the smaller $F1$ values are. This is due to the less time for precharging in higher supply voltages where synchronizers are faster. (4) Nevertheless,

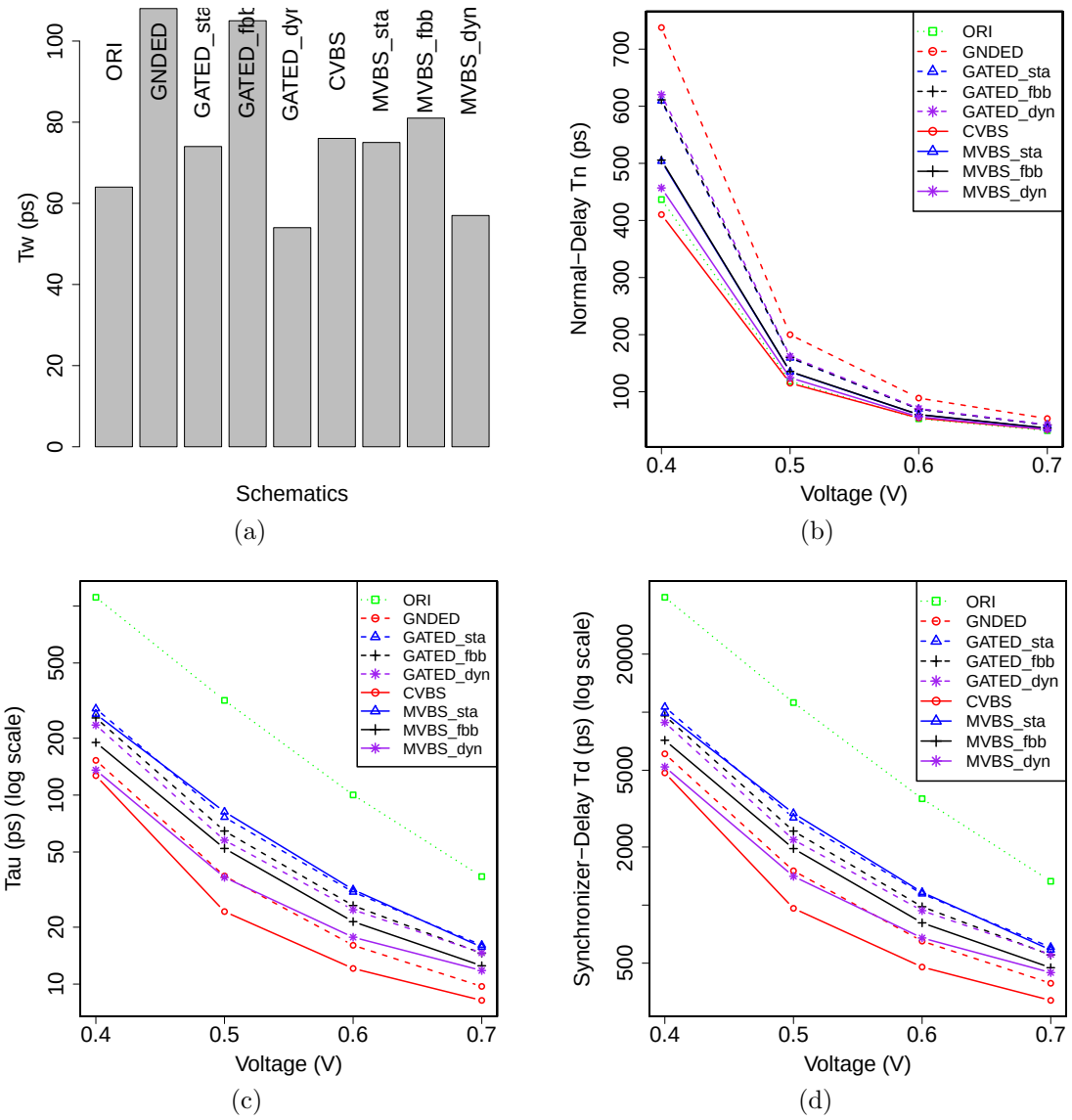


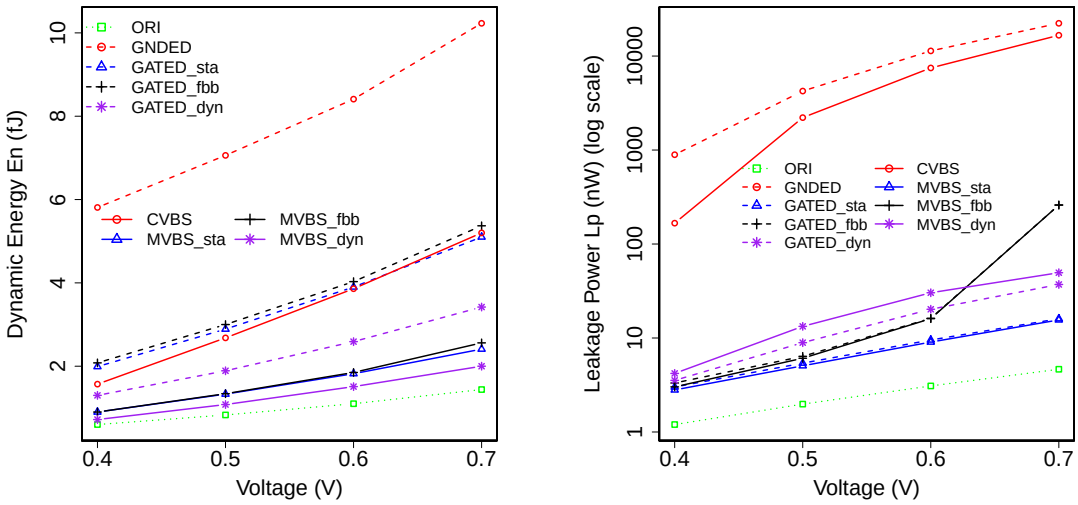
Figure 4.11: Synchronizer timing parameters: (a)t_w at 0.4V; (b)t_n; (c) τ ($\bar{\tau}$);(d) t_d.

it is more difficult to discover the correlation between supply voltages and $F2$ values. A simple explanation can be that the higher supply voltage is, the more charge the capacitor has, however, the more charge the synchronizer consumes.

Table 4.4: The Fullness of Charge Pump (in %)

Voltage (V)	CVBS		MVBS sta		MVBS fbb		MVBS dyn	
	F1	F2	F1	F2	F1	F2	F1	F2
0.4	99.9	116.9	100	151.2	99.7	150.4	98.4	145.8
0.5	99.1	139	99.9	154.8	99	153.6	96.6	148.4
0.6	98.6	136.2	99.3	151.2	97.1	148.1	95.3	144.3
0.7	98.5	131.4	97.3	144.8	95.3	140.9	94.5	139.1

The energy and leakage power are shown in Fig. 4.12.



(a) (b)

Figure 4.12: (a)Dynamic energy; (b)Leakage power.

4.2.3 Layout Implementation and Simulations

The layouts of synchronizers are seen in Fig. 4.13 and Fig. 4.14. In terms of area, VBSs are much smaller than the additional flip-flops used in [46] for post-silicon calibrations of synchronizers, though being larger than the basic Jamb latch. Nevertheless, it does not introduce a large overall area expense since synchronizers are usually much less than the storage flip-flops in digital systems.

The post-layout simulation results are presented in Table. 4.5. The $r2$ is the ratio of τ_{PL} in post-layout simulations over τ_{SCH} in schematic simulations. Nevertheless, a large $r2$ suggests the layout needs more optimizations. Values of τ and t_d are also illustrated in Fig. 4.15. A conclusion is drawn that MVBS and CVBS are 2.51 to 3.26 and 4.21 to 8.18 times faster than the basic Jamb latch, respectively.

4.3 Conclusions

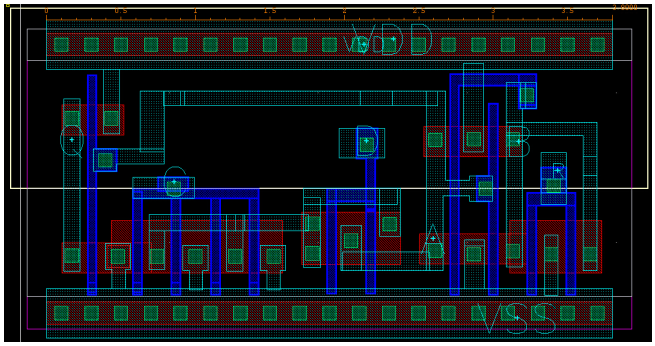
To solve the delay bottleneck due to synchronizers at low supply voltages with a specified target of MTBF due to metastability, this work proposed two VBSs consisting of a basic Jamb latch and a charge pump, MVBS and CVBS. The capacitor of the charge pump is sized large enough to achieve a high powering capability to speed up the metastability resolution in the Jamb latch. For a equivalent 1-year MTBF specification, MVBS and CVBS show 2.0 to 2.7 and 5.1 to 9.8 times delay improvements over the basic Jamb latch, respectively, without incurring large power consumption.

The VBSs are further improved in several aspects. The transistors and the capacitors are well sized to meet the extra constraints and the metastability detectors are implemented with full body forward biasing or dynamic logic to reduce the detection delay. An accurate methodology is proposed for extracting metastability circuits parameters for synchronizers under changing biasing currents. The VBSs can be precharged to a fullness of at least 94.5% within the given minimum synchronizer delay. For the 35τ MTBF specification at four levels of supply voltages, MVBS and CVBS show maximally 2.97 to 7.57 and

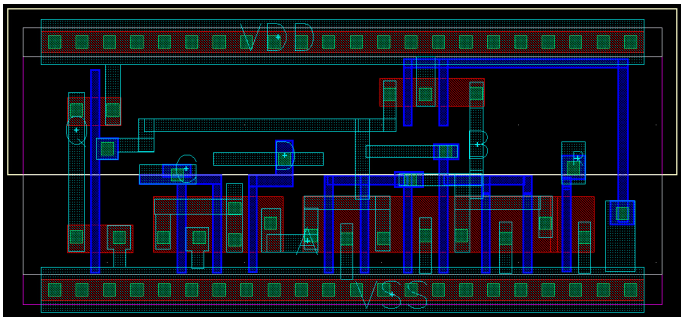
Table 4.5: Post-layout simulations results for the 35τ specification.

	Voltage (V)	τ_{SCH} (ps)	τ_{PL} (ps)	r2	t_n (ps)	t_d (ps)	r0
ORI	0.7	37.1	69.7	1.88	47.8	2488.5	1
	0.6	100.2	206.8	2.06	80.8	7318.3	1
	0.5	317.1	706.8	2.23	185.6	24924.5	1
	0.4	1113.2	2678.2	2.41	695.3	94432.9	1
GNDED	0.7	9.7	20.0	2.06	53.1	752.9	3.31
	0.6	16.0	36.3	2.26	88.9	1358.3	5.39
	0.5	37.3	93.4	2.50	200.0	3469.5	7.18
	0.4	152.7	423.3	2.77	737.8	15553.0	6.07
GATED_sta	0.7	16.1	29.9	1.86	61.9	1108.2	2.25
	0.6	30.7	61.7	2.01	103.6	2264.5	3.23
	0.5	76.4	168.6	2.21	236.9	6137.1	4.06
	0.4	286.6	678.9	2.37	892.9	24655.7	3.83
CVBS	0.7	8.2	15.4	1.88	51.7	590.7	4.21
	0.6	12.1	26.4	2.18	85.2	1009.2	7.25
	0.5	24.2	81.7	3.38	188.8	3048.3	8.18
	0.4	126.7	481.6	3.80	685.3	17541.3	5.38
MVBS_sta	0.7	15.7	26.8	1.71	52.6	990.6	2.51
	0.6	31.6	73.4	2.32	88.3	2657.3	2.75
	0.5	81.6	212.7	2.61	202.0	7646.5	3.26
	0.4	268.2	988.2	3.68	763.2	35350.2	2.67

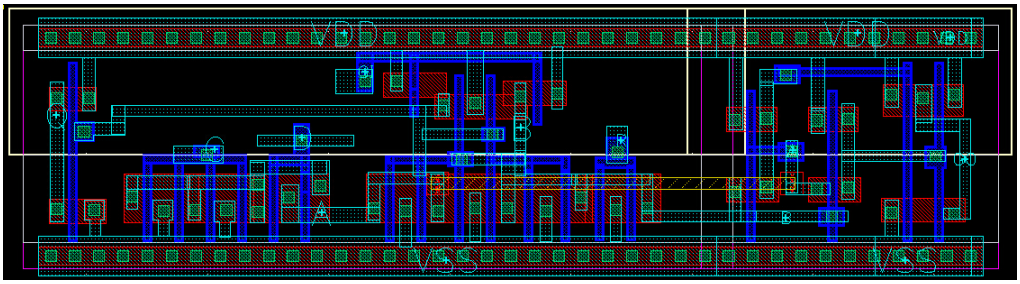
4.14 to 8.13 times delay improvements over the basic Jamb latch, respectively, without incurring large power consumption. For the same condition, post-Layout simulations show MVBS and CVBS are 2.51 to 3.26 and 4.21 to 8.18 times faster than the basic Jamb latch, respectively.



(a)

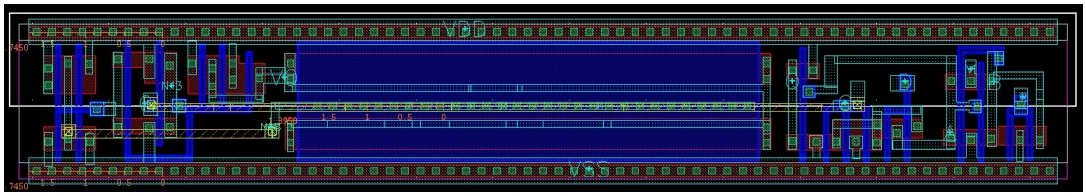


(b)

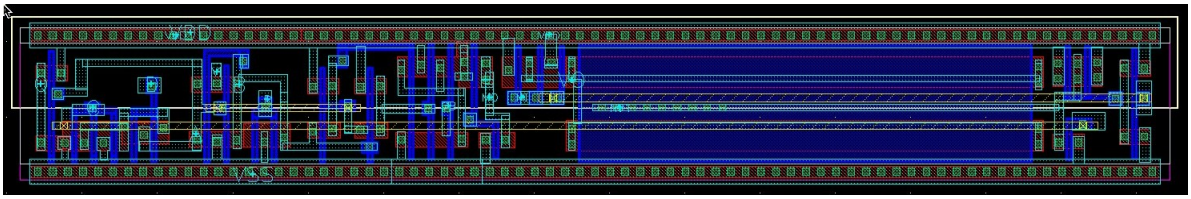


(c)

Figure 4.13: Layouts and Area of synchronizers (all widths $W = 1.8\,\mu m$ in standard cells): (a)ORI(Length $L = 3.8\,\mu m$ and area $A = 6.84\,\mu m^2$); (b) GNDED $L = 4.4\,\mu m, A = 7.92\,\mu m^2$; (c) GATED ($L = 7.8\,\mu m, A = 14.04\,\mu m^2$).

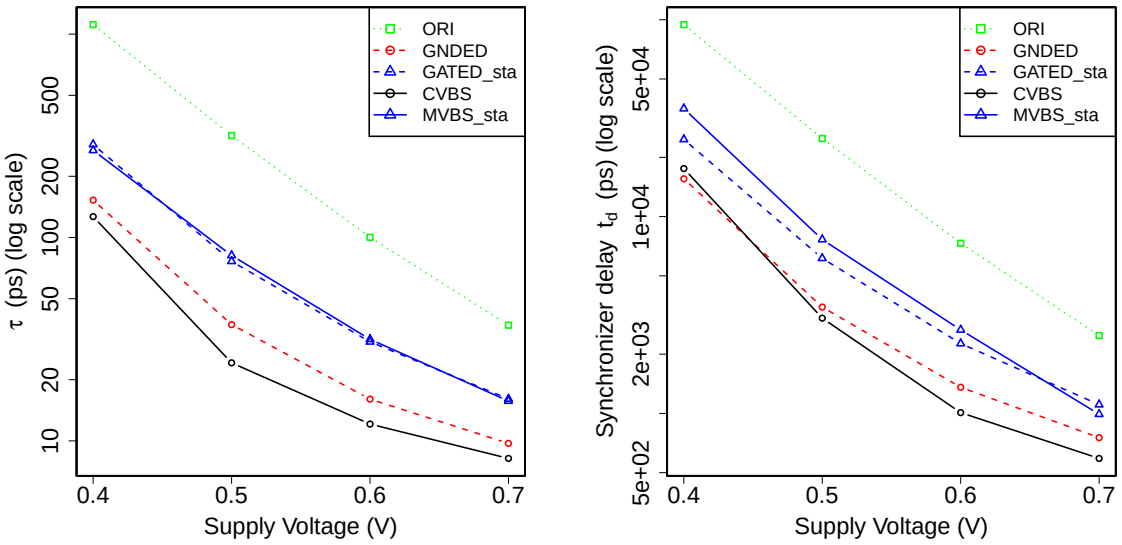


(a)



(b)

Figure 4.14: Synchronizer Layouts (all widths $W = 1.8\,\mu m$ in standard cells): (a) CVBS ($L = 13.4\,\mu m, A = 24.12\,\mu m^2$); (b)MVBS ($L = 15\,\mu m, A = 27\,\mu m^2$).



(a) (b)

Figure 4.15: Post-Layout Simulation Results: (a) τ; (b)t_d.

Chapter 5

Conclusion and Future Work

5.1 Conclusions

This dissertation made an attempt to enhance the reliability of EDS circuits by designing and deploying EDS circuits in energy-efficient digital systems. The dissertation made contributions in following areas:

Voltage-Boosted Synchronizer Design for EDS Circuits The synchronizers in EDS circuits undergo the metastability problem. Any types of metastable signals possibly yield system status inconsistencies that can initiate system failures. These observations necessitate metastable-hardened synchronizers for building EDS circuits. To solve the delay bottleneck due to synchronizers at low supply voltages, this work proposed two VBSs consisting of a basic Jamb latch and a charge pump, MVBS and CVBS. The capacitor of the charge pump is sized large enough to achieve a high powering capability to speed up the metastability resolution in the Jamb latch. For a equivalent 1-year MTBFs specification, MVBS and CVBS show 2.0 to 2.7 and 5.1 to 9.8 times delay improvements over the basic Jamb latch, respectively, without incurring large power consumption.

VBSs are further modified and optimized to improve synchronizer delay. Transistor-level optimization techniques including transistor sizing, forward body biasing and dy-

namic implementations were applied to the baseline and proposed synchronizers. For a 35τ MTBFs specification in typical PVT conditions, MVBSs and CVBSs show 2.97 to 7.57 and 4.14 to 8.13 times delay improvement over the basic Jamb latch, respectively, without incurring large power consumption. For the same conditions, post-Layout simulations show MVBSs and CVBSs are 2.51 to 3.26 and 4.21 to 8.18 times faster than a basic Jamb latch, respectively.

EDS Circuit Deployment To enhance the extrinsic EDS-reliability, a new EDS deployment methodology have been developed. The EDS circuits are augmented to the non-critical paths with high activations to assure the sampling accuracy and the duty cycle of the clock signal is tuned to achieve the speculative requirement. This methodology requires neither buffer insertion nor dual clocks and is applicable for FPGA implementations. An FPGA-based Discrete Cosine Transform with EDS and DVS circuits deployed in this fashion and demonstrates up to 16.5% energy savings over a conventional design at equivalent frequency setting and image quality, with a 0.8% logic element and 3.5% maximum frequency penalties.

In summary, this dissertation significantly expand the application scope of energy-efficient digital systems with EDS circuits to the low supply-voltage and/or FPGA-based DSP applications.

5.2 Future Work

Voltage-Boosted Synchronizers Advanced versions of VBSs targeting better metastability resolution will be proposed and simulated. An important topic for the VBSs can be the reliability issue under noisy circumstances. A metastability test chip will be designed and taped out to measure the metastability behaviour of the simulated VBSs in a realistic environment. A key challenge for the test chip is to measure the average τ under changing biasing currents.

Energy-Efficient FPGA-based Microprocessor with EDS FPGA-based microprocessors are pervasively used. Nevertheless, the timing error correction that is applicable for FPGA implementation is still needed and "Multiple-issue" error recovery strategy proposed by [1] can be one solution.

Publications

1. Yaoqiang Li, Pierce I-Jen Chuang, Andrew Kennings, and Manoj Sachdev. "An FPGA Implementation of a Timing-Error Tolerant Discrete Cosine Transform" (Abstract Only). ACM/SIGDA International Symposium on FPGA, Monterey, CA, 2015: 266-266.

2. Yaoqiang Li, Pierce I-Jen Chuang, Andrew Kennings and Manoj Sachdev. "Voltage-Boosted Synchronizers". Great Lakes Symposium on VLSI (GLSVLSI), Pittsburgh, PA, 2015: 307-312.

3. Yaoqiang Li, Pierce I-Jen Chuang, Andrew Kennings, Manoj Sachdev. "Runtime Slack-Deficit Detection for a Low-Voltage DCT Circuit". International Midwest Symposium on Circuits and Systems, August 2-5, Fort Collins, Colorado, 2015.

4. Yaoqiang Li, Pierce I-Jen Chuang, Andrew Kennings and Manoj Sachdev. "Advanced Voltage-Boosted Synchronizers". Microelectron. Journal. (to be submitted)

References

[1] K. Bowman, J. Tschanz, S. Lu, P. Aseron, M. Khellah, A. Raychowdhury, B. Geuskens, C. Tokunaga, C. Wilkerson, T. Karnik, and V. De, "A 45 nm Resilient Microprocessor Core for Dynamic Variation Tolerance," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 194–208, Jan 2011.

[2] P. Whatmough, S. Das, D. Bull, and I. Darwazeh, "Circuit-Level Timing Error Tolerance for Low-Power DSP Filters and Transforms," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, no. 6, pp. 989–999, June 2013.

[3] J. Levine, E. Stott, G. Constantinides, and P. Cheung, "Online Measurement of Timing in Circuits: For Health Monitoring and Dynamic Voltage & Frequency Scaling," in *Int. Symp. on Field-Programmable Custom Computing Machines (FCCM)*, April 2012, pp. 109–116.

[4] J. M. Levine, E. Stott, and P. Y. Cheung, "Dynamic Voltage & Frequency Scaling with Online Slack Measurement," in *Int. Symp. on Field-Programmable Gate Arrays (FPGA)*. New York, NY, USA: ACM, Feb 2014, pp. 65–74.

[5] L. Technology, "LT3070 datasheet," http://cds.linear.com/docs/en/datasheet/3070fc.pdf, 2015, [Online; accessed 12-12-2015].

[6] S. Ghosh and K. Roy, "Parameter Variation Tolerance and Error Resiliency: New Design Paradigm for the Nanoscale Era," *Proceedings of the IEEE*, vol. 98, no. 10, pp. 1718–1751, Oct 2010.

[7] B. Giridhar, M. Fojtik, D. Fick, D. Sylvester, and D. Blaauw, "Pulse amplification based dynamic synchronizers with metastability measurement using capacitance derating," in *IEEE Custom Integrated Circuits Conference*, Sept 2013, pp. 1–4.

[8] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low-power CMOS digital design," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 4, pp. 473–484, Apr 1992.

[9] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. USA: Addison-Wesley Publishing Company, 2010.

[10] M. Alioto, "Ultra-Low Power VLSI Circuit Design Demystified and Explained: A Tutorial," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 1, pp. 3–29, Jan 2012.

[11] A. Putnam, A. Caulfield, E. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmaeilzadeh, J. Fowers, G. P. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J.-Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Y. Xiao, and D. Burger, "A Reconfigurable Fabric for Accelerating Large-Scale Datacenter Services," in *41st Annual International Symposium on Computer Architecture (ISCA)*, June 2014, pp. 13–24.

[12] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital integrated circuits : a design perspective*, 2nd ed., ser. Prentice Hall electronics and VLSI series. Pearson Education, Jan 2003.

[13] K. Kang, K. Kim, and K. Roy, "Variation Resilient Low-Power Circuit Design Methodology using On-Chip Phase Locked Loop," in *44th ACM/IEEE Design Automation Conference*, June 2007, pp. 934–939.

[14] D. Ernst, N. S. Kim, S. Das, S. Pant, R. Rao, T. Pham, C. Ziesler, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "Razor: A Low-Power Pipeline Based on Circuit-Level Timing Speculation," in *36th annual IEEE/ACM International Symposium on Microarchitecture*. Washington, DC, USA: IEEE Computer Society, Dec 2003, pp. 7–18.

[15] S. Das, C. Tokunaga, S. Pant, W.-H. Ma, S. Kalaiselvan, K. Lai, D. Bull, and D. Blaauw, "RazorII: In Situ Error Detection and Correction for PVT and SER Tolerance," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, Jan 2009.

[16] M. Fojtik, D. Fick, Y. Kim, N. Pinckney, D. Harris, D. Blaauw, and D. Sylvester, "Bubble Razor: Eliminating Timing Margins in an ARM Cortex-M3 Processor in 45 nm CMOS Using Architecturally Independent Error Detection and Correction," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 66–81, Jan 2013.

[17] D. Bull, S. Das, K. Shivashankar, G. Dasika, K. Flautner, and D. Blaauw, "A Power-Efficient 32 bit ARM Processor Using Timing-Error Detection and Correction for Transient-Error Tolerance and Adaptation to PVT Variation," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 18–31, Jan 2011.

[18] J. Crop, R. Pawlowski, and P. Chiang, "Regaining throughput using completion detection for error-resilient, near-threshold logic," in *49th Design Automation Conference*. ACM, June 2012, pp. 974–979.

[19] R. Pawlowski, E. Krimer, J. Crop, J. Postman, N. Moezzi-Madani, M. Erez, and P. Chiang, "A 530mV 10-lane SIMD processor with variation resiliency in 45nm SOI," in *IEEE International Solid-State Circuits Conference, Digest of Technical Papers*, Feb 2012, pp. 492–494.

[20] E. Krimer, P. Chiang, and M. Erez, "Lane decoupling for improving the timing-error resiliency of wide-SIMD architectures," *SIGARCH Comput. Archit. News*, vol. 40, no. 3, pp. 237–248, Jun 2012.

[21] M. Alba, A. Chua, W. Lofamia, R. Maestro, J. Hizon, J. Madamba, H. Aquino, and L. Alarcon, "An aggressive power optimization of the ARM9-based core using RAZOR," in *TENCON 2012 - 2012 IEEE Region 10 Conference*, Nov 2012, pp. 1–5.

[22] S. Das, D. Roberts, S. Lee, S. Pant, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "A self-tuning DVS processor using delay-error detection and correction," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 4, pp. 792–804, April 2006.

[23] S. Beer, M. Cannizzaro, J. Cortadella, R. Ginosar, and L. Lavagno, "Metastability in Better-Than-Worst-Case Designs," in *20th IEEE International Symposium on Asynchronous Circuits and Systems*, May 2014, pp. 101–102.

[24] J. Levine, E. Stott, G. Constantinides, and P. Cheung, "SMI: Slack Measurement Insertion for online timing monitoring in FPGAs," in *Int. Conf. on Field Programmable Logic and Applications (FPL)*, Sept 2013, pp. 1–4.

[25] C. Chow, L. Tsui, P. Leong, W. Luk, and S. Wilton, "Dynamic voltage scaling for commercial FPGAs," in *Proc. IEEE International Conference on Field-Programmable Technology*, Dec 2005, pp. 173–180.

[26] R. Ginosar, "Metastability and Synchronizers: A Tutorial," *IEEE Design & Test of Computers*, vol. 28, no. 5, pp. 23–35, Sept 2011.

[27] D. J. Kinniment, *Synchronization and Arbitration in Digital Systems*. Wiley Publishing, 2008.

[28] C. Portmann and H. Meng, "Metastability in CMOS library elements in reduced supply and technology scaled applications," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 1, pp. 39–46, Jan 1995.

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

[30] C. Dike and E. Burton, "Miller and noise effects in a synchronizing flip-flop," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 6, pp. 849–855, Jun 1999.

[31] T. Sakurai, "Optimization of CMOS arbiter and synchronizer circuits with submicrometer MOSFETs," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 4, pp. 901–906, Aug 1988.

[32] D. Kinniment, C. Dike, K. Heron, G. Russell, and A. Yakovlev, "Measuring Deep Metastability and Its Effect on Synchronizer Performance," *IEEE Transactions on*

Very Large Scale Integration (VLSI) Systems, vol. 15, no. 9, pp. 1028–1039, Sept 2007.

[33] L.-S. Kim and R. Dutton, "Metastability of CMOS latch/flip-flop," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 4, pp. 942–951, Aug 1990.

[34] K. Bowman, J. Tschanz, N. S. Kim, J. Lee, C. Wilkerson, S.-L. Lu, T. Karnik, and V. De, "Energy-Efficient and Metastability-Immune Resilient Circuits for Dynamic Variation Tolerance," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 1, pp. 49–63, Jan 2009.

[35] D. Sengupta and R. Saleh, "Power-Delay Metrics Revisited for 90Nm CMOS Technology," in *6th International Symposium on Quality of Electronic Design*. IEEE Computer Society, March 2005, pp. 291–296.

[36] S. Beer and R. Ginosar, "Eleven Ways to Boost Your Synchronizer," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 6, pp. 1040–1049, June 2015.

[37] R. Ginosar, "Fourteen ways to fool your synchronizer," in *9th International Symposium on Asynchronous Circuits and Systems*, May 2003, pp. 89–96.

[38] D. Li, P. Chuang, and M. Sachdev, "Comparative analysis and study of metastability on high-performance flip-flops," in *11th International Symposium on Quality Electronic Design*, March 2010, pp. 853–860.

[39] D. Rennie, D. Li, M. Sachdev, B. Bhuva, S. Jagannathan, S. Wen, and R. Wong, "Performance, metastability and soft-error robustness tradeoffs for flip-flops in 40nm CMOS," in *Custom Integrated Circuits Conference (CICC)*, Sept 2011, pp. 1–4.

[40] ——, "Performance, Metastability, and Soft-Error Robustness Trade-offs for Flip-Flops in 40 nm CMOS," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 8, pp. 1626–1634, Aug 2012.

[41] D. Li, D. Rennie, P. Chuang, D. Nairn, and M. Sachdev, "Design and analysis of metastable-hardened and soft-error tolerant high-performance, low-power flip-flops," in *12th International Symposium on Quality Electronic Design*, March 2011, pp. 1–8.

[42] D. Li, P. Chuang, and M. Sachdev, "Design of a novel high-performance pre-discharge flip-flop," in *8th IEEE International NEWCAS Conference*, June 2010, pp. 233–236.

[43] D. Li, P.-J. Chuang, D. Nairn, and M. Sachdev, "Design and analysis of metastable-hardened flip-flops in sub-threshold region," in *International Symposium on Low Power Electronics and Design*, Aug 2011, pp. 157–162.

[44] J. Zhou, D. Kinniment, G. Russell, and A. Yakovlev, "A robust synchronizer," in *IEEE Computer Society Annual Symposium on Emerging VLSI Technologies and Architectures*, March 2006, pp. 2 pp.–.

[45] M. Kayam, R. Ginosar, and C. Dike, "Symmetric Boost Synchronizer for Robust Low Voltage, Low Temperature Operation," EE Tech. Rep., 2007, technion.

[46] J. Zhou, D. Kinniment, G. Russell, and A. Yakovlev, "Adapting Synchronizers to the Effects of on Chip Variability," in *14th IEEE International Symposium on Asynchronous Circuits and Systems*, April 2008, pp. 39–47.

[47] K. Ayob, *Fundamentals of Timing in FPGAs*. CreateSpace Independent Publishing Platform, Feb 2015.

[48] Altera, "Timing Analysis of Internally Generated Clocks in TimeQuest 2.0," https://www.alteraforum.com, 2009, [Online; accessed 01-12-2015].

[49] ——, "Inverted Clocks," https://www.altera.com/support/support-resources/knowledge-base/solutions/rd05242007_585.highResolutionDisplay.html, 2010, [Online; accessed 01-12-2015].

[50] M. Krepa, "Discrete Cosine Transform core," http://opencores.org/project,mdct, 2009, [Online; accessed 15-08-2014].

[51] W. Allan, "The USC-SIPI Image Database," http://sipi.usc.edu/database/database.php, 2014, [Online; accessed 15-08-2014].

[52] P. Favrat, P. Deval, and M. Declercq, "A high-efficiency CMOS voltage doubler," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 3, pp. 410–416, Mar 1998.

[53] B. Razavi, *Design of Analog CMOS Integrated Circuits*, 1st ed. New York, NY, USA: McGraw-Hill, Inc., 2001.