

An Efficient Transition Detector Exploiting Charge Sharing

by

Yu Wang

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
December 13, 2014

Keywords: Better-than-worst-case design, error detection, transition detector,
Short circuit, charge sharing, PVT variations.

Copyright 2014 by Yu Wang

Approved by

Adit D.Singh, Chair, James B Davis Professor, Electrical and Computer Engineering
Vishwani Agrawal, James J Danaher Professor, Electrical and Computer Engineering
Victor P. Nelson, Professor and Assistant Chair, Electrical and Computer Engineering

Abstract

Transition detectors have been widely employed for online error and metastability detection, including in Better-Than-Worst-Case (BTWC) timing design of microprocessors that are designed to allow occasional timing errors. Early error detector designs such as Razor introduced a shadow latch with a delayed clock in parallel to the datapath flip-flop for timing-error detection through duplication and comparison. Intel and ARM have presented transition detectors in prototype implementations of BTWC designs that both detect timing errors, and also address the issue of flip-flop metastability. While these designs represent the state-of-art in transition detector design the high area overhead remains a major concern since the circuits may need to be incorporated in almost all the flip-flops in a large design. In this thesis, we propose a much more efficient transition detector exploiting charge sharing (TDCS) that displays ultra-low area overhead. Our design employs a novel combination of short circuit effects and charge sharing based discharge for operation, and reduces by more than half the complexity of conventional transition detector designs. Although the motivation for the TDCS design originates from BTWC design, it can be utilized in various other applications. A TDCS circuit is designed in 45nm technology for evaluation, and analyzed based on a high performance version of the PTM model. Simulation of our TDCS design shows that it can reliably achieve the same functionality as published designs with 60% fewer transistors. Furthermore, corner analysis shows that TDCS is also robust under extreme PVT variations.

Acknowledgments

I would like to express my gratitude to my supervisor Dr. Adit Singh for the continuous support of my Master's study and research. His guidance, motivation, enthusiasm, and immense knowledge helped me throughout my entire research. I am grateful for his patience with me since I am a beginner in VLSI design area and, sometimes, a stubborn student.

I would like to thank Dr. Vishwani Agrawal for his suggestions and help in my graduate study. No matter when I ask a question in or after class, I can always get a more detailed, insightful and easier to understand answer than I expected.

I would like thank Dr. Victor Nelson. He was the first professor who introduced me to digital circuit world. His class impressed me with vast and detailed information from which I learned the skills needed in my research.

I thank my friends: Jie Zou, Erlong Zhang and Chao Han for their encouragement and suggestions when I am depressed. Special thanks go out to Huirong Li and Bei zhang for helping me with writing the thesis.

Last but not least, I would like to thank my parents: Keyong Wang and Jing Kong, for giving me the chance to come Auburn for higher education. They are the motivation in my life to make myself a better person.

This research was supported in part by the National Science Foundation under Grants No. EECS-0903449 and CCF-1319529.

Table of Contents

Abstract	ii
Acknowledgments.....	iii
List of Tables.....	vi
List of Figures.....	vii
Chapter 1 Introduction	1
1.1 Background	2
1.2 The need for Better-Than-Worst-Case Design.....	3
1.3 Problem statement.....	6
1.4 Contribution of this thesis.....	7
1.5 Organization of the thesis	8
Chapter 2 Review of current transition detector circuit.....	9
2.1 Transition detector with timing-borrowing	9
2.2 ARM's transition detector	11
2.3 The need for high performance and low area overhead transition detector.....	13
Chapter 3 Proposed transition detector exploiting charge sharing	14
3.1 Dynamic CMOS gate	14
3.1.1 Dynamic CMOS gate principles	14
3.1.2 Issues with dynamic CMOS gates	16
3.2 Short circuit effect in CMOS	18
3.3 Design of transition detector exploiting charge sharing	20
3.3.1 Falling transition detection by short circuit discharge.....	20
3.3.2 Rising transition detection by charge sharing.....	21
3.4 Evaluation of TDCS.....	23
3.4.1 Transistor sizing in TDCS.....	23
3.4.2 SPICE simulation of TDCS	25

3.5 Evaluation of area overhead.....	28
Chapter 4 TDCS design robustness	30
4.1 Introduction.....	30
4.2 Process variation	31
4.3 Supply voltage.....	31
4.4 Temperature	32
4.5 PVT Corners.....	32
4.6 Corner analysis results	36
Chapter 5 Conclusion and future work	37
Bibliography.....	38

List of Tables

Table 3.1 Transistor sizing of TDCS	25
Table 3.2 TDCS performance.....	28
Table 3.3 Area overhead comparison	29
Table 4.1 Ambient temperature ranges.....	32
Table 4.2 Process corners	34
Table 4.3 Environmental Corners	34
Table 4.4 PVT corner list.....	34
Table 4.5 Corner analysis results	35

List of Figures

Figure 1.1 Electronic device toward smaller size	1
Figure 1.2 power density of some processors	2
Figure 1.3 Design complexity trends	4
Figure 1.4 Extreme Variations a) Heat Flux results in Vcc variation. b) Temperature variation result in hot spot. b) Random dopant fluctuations.....	5
Figure 1.5 Soft errors trends.....	6
Figure 2.1 The transition detector with time borrowing (TDTB) from Intel . (a) Schematic of TDTB (b) A possible implementation of the pulse generator in TDTB.....	11
Figure 2.2 The transition detector used by ARM in the prototype RAZOR design. (a) Schematic of the entire design. (b) Enlarged schematic of the transition detection circuit.....	11
Figure 3.1 a) CMOS gate structure b) dynamic CMOS gate structure.....	14
Figure 3.2 dynamic NAND gate.....	15
Figure 3.3 Charge sharing in domino COMS gate.....	17
Figure 3.4 domino AND gate with keeper.....	18
Figure 3.5 (a) Short circuit current in CMOS inverter (b) Approximate short circuit current.....	19
Figure 3.6 Transition detector exploiting charge sharing and short circuit discharge for detection rising and falling transitions respectively	20
Figure 3.7 Falling transition detection process through short circuit discharge.....	21
Figure 3.8 Rising transition detection through charge sharing (Note $C_e > C_n$).....	22
Figure 3.9 The architecture of TDCS with a latch in the datapath	24
Figure 3.10 Simulated timing-error detection demonstration	26
Figure 3.11 Detailed view of rising and falling transition detection. (a) Falling transition	
Figure 4.1 a) Uniform distribution b) Normal distribution	31

Figure 4.2 Five different types of process corners..... 33

Chapter 1

INTRODUCTION

Great success has been achieved by the semiconductor industry over the last 50 years, which has dramatically changed people's lifestyle. Electronic devices have become part of our life. They are everywhere from computers and cell phones to home appliances. When the first general purpose computer ENIAC (Electronic Numerical Integrator And Computer) [8, 9] was announced in 1946, nobody could imagine we could turn the giant machine into a hand-sized cell phone, while today's cell phone performs 10,000 times faster than the ENIAC. Fig. 1.1 shows the evolution of electronic devices.

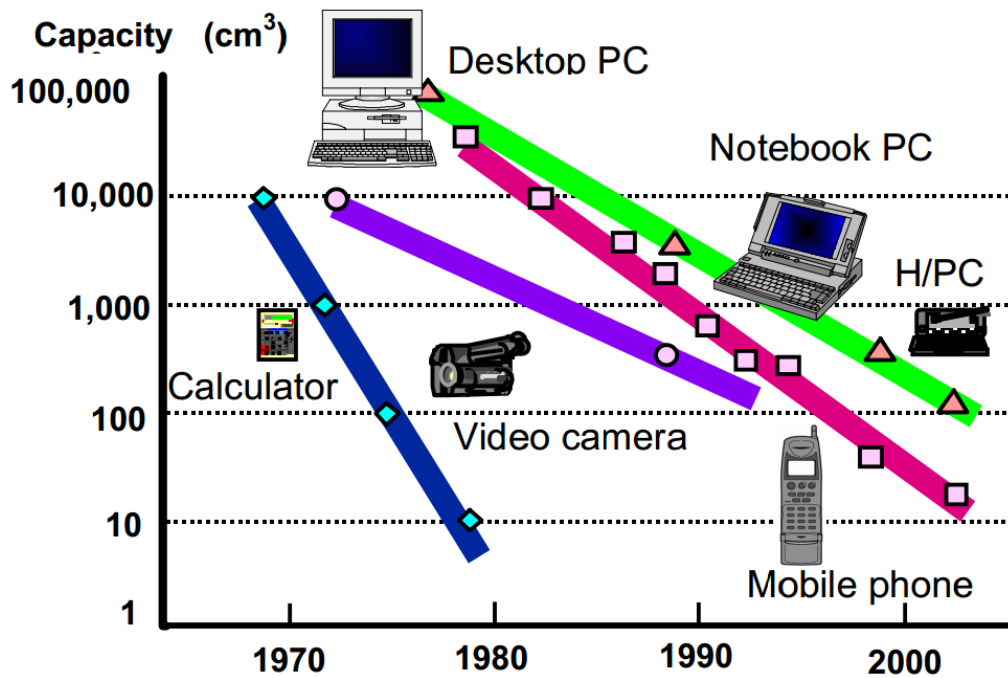


Figure 1.1 Electronic device toward smaller size [12]

It is well known that the integration capacity has doubled every two years for many decades. This trend was first discovered by Gordon Moore in 1965 [10, 11], which results in the famous Moore's Law. The major driving force behind the technology in the last few decades was scaling. The foundation of scaling was laid out with the invention of the self-aligned silicon gate process in the late 1960s. Thanks to scaling, the performance of a chip

has been increasing exponentially. As we scaled down into submicron regime, new issues such as power inefficiencies and reliability arose in VLSI design

1.1 Background

Recently, Better-Than-Worst-Case (BTWC) timing designs have been proposed as a way to overcome the performance and power inefficiencies arising from the need to guardband highly scaled designs against uncertainty incurred by process, voltage and temperature (PVT) variations. [1, 18]

As we stepped into the submicron regime, technology allowed billions of transistors integrated on a single chip, running at gigahertz frequencies. However, the growing die size and high frequency led to huge power consumption. Researchers had once predicted that “If scaling continues at present pace, by 2005, high speed processor would have power density of nuclear reactor, by 2010, a rocket nozzle, and by 2015, surface of sun” [13]. Fig. 1.2 shows the trend of power density. Although these things did not happen, nevertheless the high power density causes serious reliability issues in a processor.

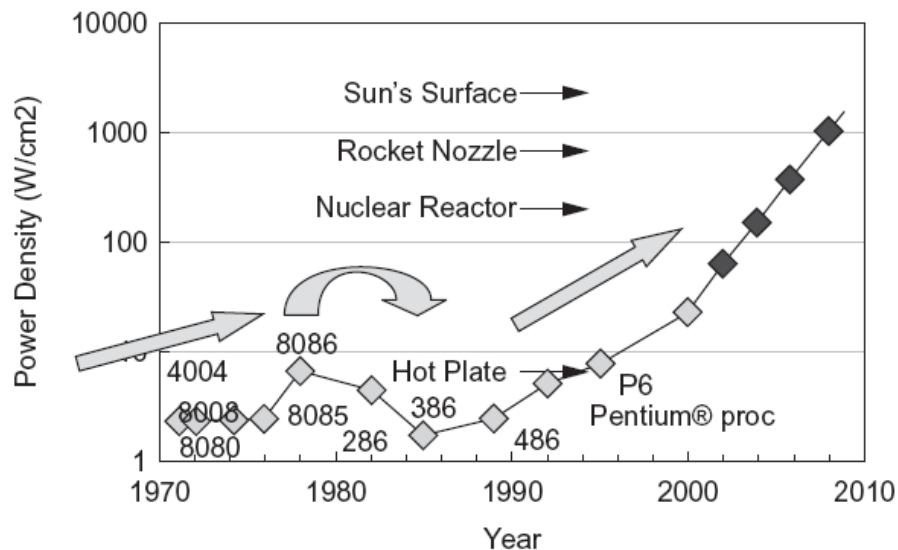


Figure 1.2 power density of some processors

In order to limit the power density of high speed digital systems, such as microprocessors, digital signal processor (DSPs) and other applications, researchers have come up with several strategies to reduce the power consumption. These strategies are collectively called low power design.

The other major driving force behind low power design are portable application requiring low power dissipation and high throughput, such as notebook computers and portable communication devices. In these applications, the power consumption must be minimized without having a major impact on the performance. Therefore, low power design strategies have become an important research field of CMOS design.

Strained silicon, hi-k metal gate and FinFETs are process related innovations that combine with low power design to ensure the possibility of further scaling. But none of them has fully re-enabled the scaling we once enjoyed. People are still looking for new design approaches to help us pursue higher performance while keeping acceptable power consumption.

1.2 The need for Better-Than-Worst-Case Design

For more than half of a century, Moore's Law has been an accurate prediction for development of the semiconductor industry. Thanks to the continuous scaling down of technology, the numbers of transistors on a single chip doubled every 18 months, which is the major driving force of today's technology. With the exponential growing transistor budget, computers have been made smaller and smaller, mobile computing device such as smart phones are created, and also the growing trend of wearable smart devices has been made possible.

As the fabrication technologies step into the nanometer regime, a series of design challenges have arisen for computer architecture: design complexity, uncertainty in environmental and fabrication conditions, and soft errors caused by charged high energy

particles [1]. In addition, these challenges make it even more difficult to meet the power and reliability budgets as we try to scale the system performance.

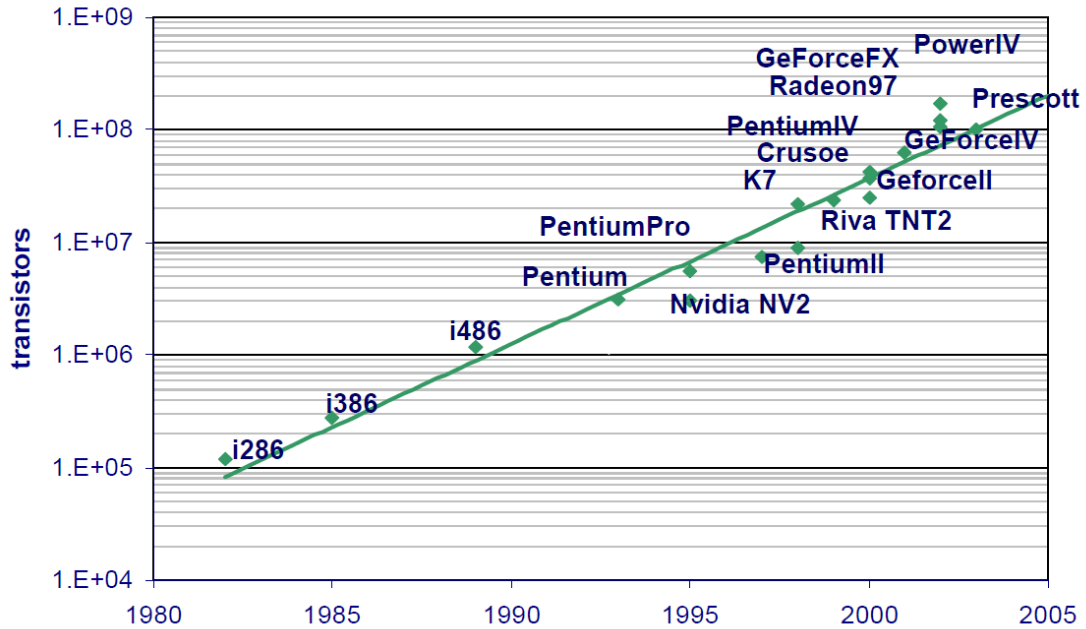
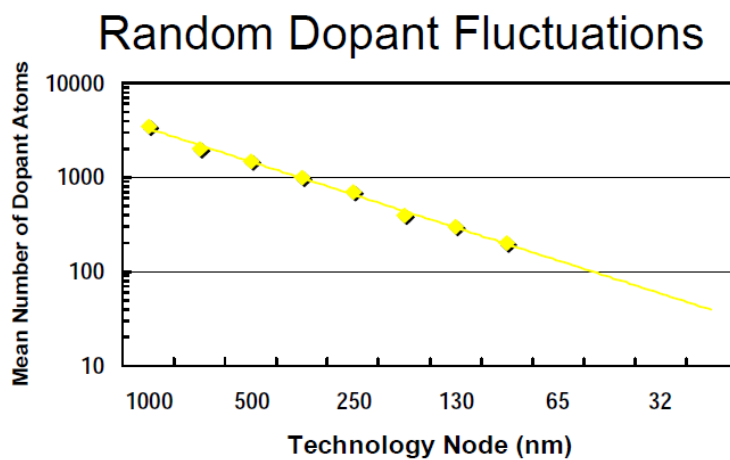
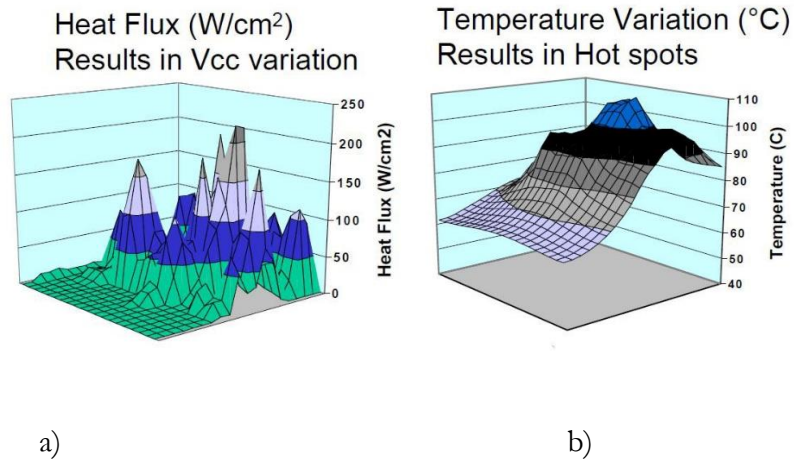


Figure 1.3 Design complexity trends

The first design challenge is design complexity. As the technology is continuously shrinking down, designers have increasingly large transistor budget. According to Moore's law, the designers' transistor budget doubles every 18 months. The increasingly large number of transistors increases the burden of verification every year. In a paper that discusses the design and verification of the Pentium IV processor, researchers indicated that the Pentium IV required 250 person-year of effort to be verified. That is two times more effort compared to the earlier Pentium Pro processor [2]. Moreover, according to ITRS (the International Technology Roadmap for Semiconductors), despite all the efforts on verification, processors still reach the market with hundreds of bugs [3].



c)

Figure 1.4 Extreme Variations. a) Heat Flux results in Vcc variation. b) Temperature variation results in hot spot. c) Random dopant fluctuations.

The second design challenge is design uncertainty. Uncertainty consists of environmental and process variation. Environmental variation is caused by changes of temperature and supply voltage, while process variation arises from device dimension and doping variation that occur during silicon fabrication. An increasing concern is process variation because its impact is becoming more significant as the feature size shrinks down. Because of process variation, designers have to design for the worst case scenario which can be overly conservative. In practice the worst case scenario is not frequently encountered which explains the possibility of processor overclocking. Today's high-end

processor, for example, Intel I7 4770k can often be easily overclocked from 3.9 Ghz to 4.4 Ghz.

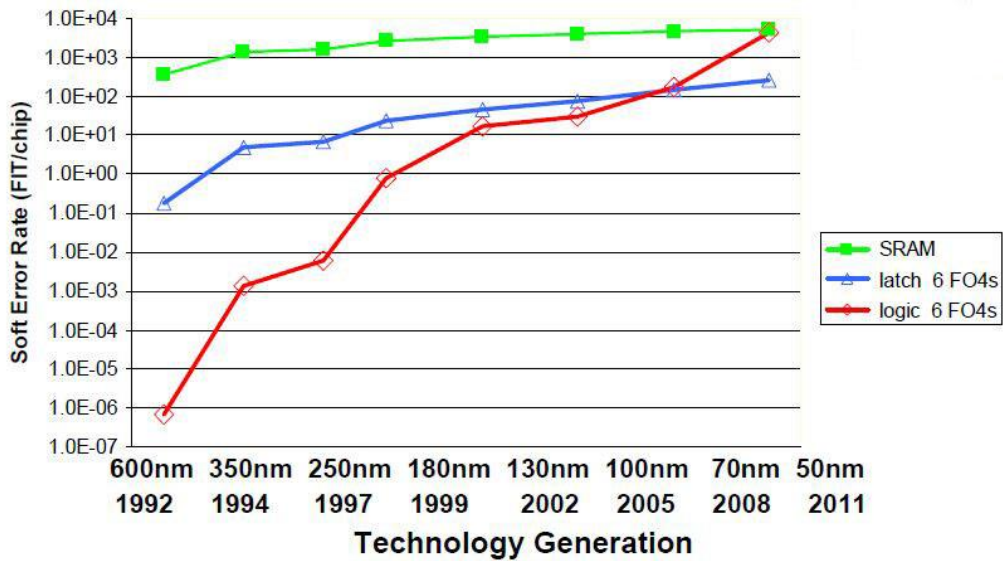


Figure 1.5 Soft errors trends

The third design concern is providing protection from soft errors that are caused by charged particles (such as neutrons and alpha particles) that strike the bulk of silicon portion of a die [4]. This may result in a logic glitch and potentially corrupted combinational logic computation. Although this is a low probability event [5], it is still a concern because of the reduced power supply and increasingly growing transistor budgets.

The combination of the three challenges is forcing designers to spend more and more effort on design and verification. On one hand they have to pursue the performance and meet power budget, on the other hand they have to achieve reliability and robustness goals.

1.3 Problem statement

Recently, a whole new design methodology has been developed to deal with the challenges that we discussed in section 1.2. This is called better-than-worst-case (BTWC) design. BTWC design releases the verification pressure by separating the performance and power concerns from correctness and reliability concerns. This is accomplished by

allowing circuits to be clocked faster than the worst-case delay even though timing-error may arise. Once the timing error occurs, it can be detected and recovered from. In this way, the designer can focus on power and performance optimization without worrying about occurrence of errors.

Although at first BTWC design was developed for low power application, recently, researchers have found that it can be adopted in high performance applications. If error correction is infrequent, the advantage of BTWC design is high. For example, suppose BTWC design is clocked 50% faster than WC design. In WC design, all instructions can be executed in one cycle. In BTWC design, 80% of the instructions can be executed in one cycle and 20% of the instructions take two cycles to be executed. The performance ratio can be calculated as follows.

$$\frac{T}{\frac{1}{2}T(0.8+0.2 \times 2)} = \frac{5}{3} \approx 1.67 \quad (1)$$

Where T is the clock period of the WC design. The result of equation (1) shows that the throughput of BTWC is 67% higher than that of WC case design. Recently, researchers have found methods to apply BTWC to high performance circuits. A high-throughput multiplier based on BTWC is developed with 2.36-times-better performance compared to the traditional multiplier. [14].

Based on the discussion above, BTWC design is a promising design methodology for various applications. In this thesis we explore the design of a key piece of BTWC support circuitry to achieve a reliable design with low area overhead.

1.4 Contribution of this thesis

A BTWC design consists of two core parts. One is the error detection circuit, and the other is the recovery mechanism. The error detection circuit is responsible for detecting timing errors that can lead to system failure. If an error is detected, the recovery circuit will

stall the pipeline and utilize extra clock cycles to correct the error. When the recovery is complete, the pipeline resumes normal operation.

The main work of this thesis is focused on the transition detector, which is an important type of error detection circuit. Compared to conventional designs, a transition detector can provide both fast error detection and a potential metastability detection capability with less hardware. Recently, researchers have proposed different designs for transition detectors. Two recent designs of transition detectors are discussed in this thesis, as background towards a new lower cost transition detector: the transition detector with time borrowing (TDTB) designed by Intel[18], and a 32nm prototype transition detector design by ARM[7]. These represent the state-of-art in the design of transition detectors.

In this thesis, we design and evaluate an efficient new transition detector that exploits charge sharing to achieve ultra-low area overhead. A combination of short circuit current and charge sharing based node discharge is used to detect both rising and falling transition, respectively, in a single sensing circuit. The proposed transition detector is designed in 45nm technology. Simulation studies are presented to show its high performance and robustness to manufacturing and environmental variations.

1.5 Organization of the thesis

This thesis is organized as follows. In Chapter 2, Intel's TDTB and ARM's transition detectors, which represent the state-of-art in transition detector design, are introduced. Chapter 3 presents our proposed design for an efficient new transition detector circuit. The core concepts behind the operation of the design, are explained in the Chapter, along with SPICE simulation to illustrate the operation and evaluate the performance of the proposed design. In Chapter 4, corner analysis is performed to test the robustness of TDCS. Finally, conclusions are drawn in Chapter 5 based on the all the analysis and simulation results in the thesis.

Chapter 2

REVIEW OF CURRENT TRANSITION DETECTOR CIRCUITS

Transition detectors were originally introduced to detect potential metastability in circuit flip-flops, but now are also commonly used for timing-error detection. The circuit is designed to detect any switching transition on a signal line within a target timing window relative to each clock edge. In BTWC designs, the goal of the transition detector is to detect input signal transitions arriving close to the clock edge to avoid metastability or during some window after the clock edge to detect timing errors. If a transition is detected in this erroneous timing window, an error signal is flagged and the recovery mechanism is activated. In this chapter, we will introduce two state-of-the-art transition detectors which are designed by Intel and ARM.

2.1 Transition detector with time-borrowing.

Fig.2.1 shows a schematic of Intel's TDTB. Notice in Fig. 2.1(a) that the conventional flip-flop used to capture the circuit state is replaced by a latch to address metastability; however, this substitution does not impact the working of the rest of the circuitry shown, which detects any transition during the high clock phase.

First, during the clock (precharge) phase, the output of the dynamic inverter is charged high. This high signal at the input of the output inverter subsequently sets the error signal low. During the high clock (evaluation) phase, this dynamic node is cut off from the power supply, while the pull down evaluation transistor is turned on. If the other N-transistor, driven by the EXOR gate, turns on, even for a short time, during this high clock phase, the dynamic node will be discharged leading to the error signal going high and indicating an error. Any transitions at the input D will generate a pulse at the EXOR output because of the deliberately introduced delay in the transition at the upper EXOR input, which will briefly cause the two EXOR inputs to differ. Thus, any transition at the input of the latch will generate a pulse at the EXOR output during the evaluation (high clock) phase and flag an error.

Fig. 2.1(b) shows one possible transistor level implementation of the pulse generating circuit. We can see that it takes 18 transistors to implement a pulse generator, which explains for the most part the hardware consumption of TDTB.

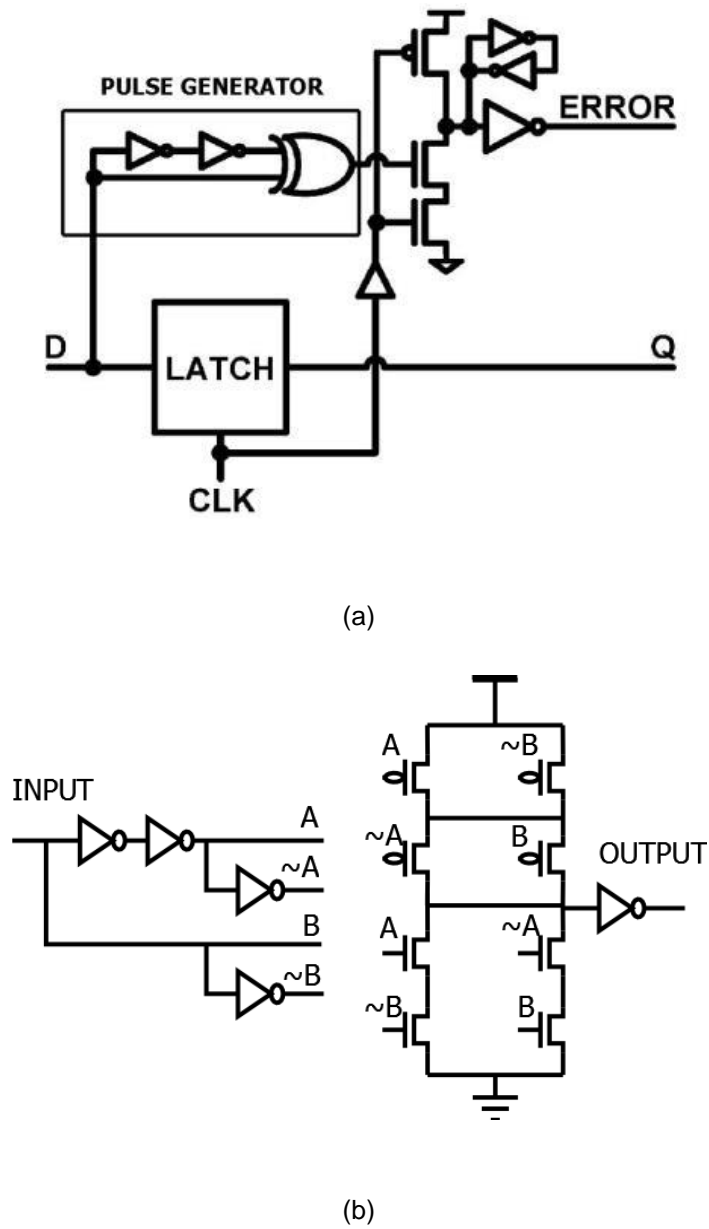


Figure 2.1 The transition detector with time borrowing (TDTB) from Intel [18]. (a) Schematic of TDTB (b) A possible implementation of the pulse generator in TDTB.

In terms of metastability issue, any transitions close to the rising clock edge will be flagged as an error due to the delay of the pulse generator. Additionally, the datapath latch can avoid datapath metastability due to its transparency during high clock phase. Although

it is possible that the transition detector itself can be in a metastable state when a transition occurs close to the edge of the error detection window, the possibility is negligible. Even if metastability occurred in transition detector, it is easier to be managed and controlled in the error detection path as compared to the functional path.

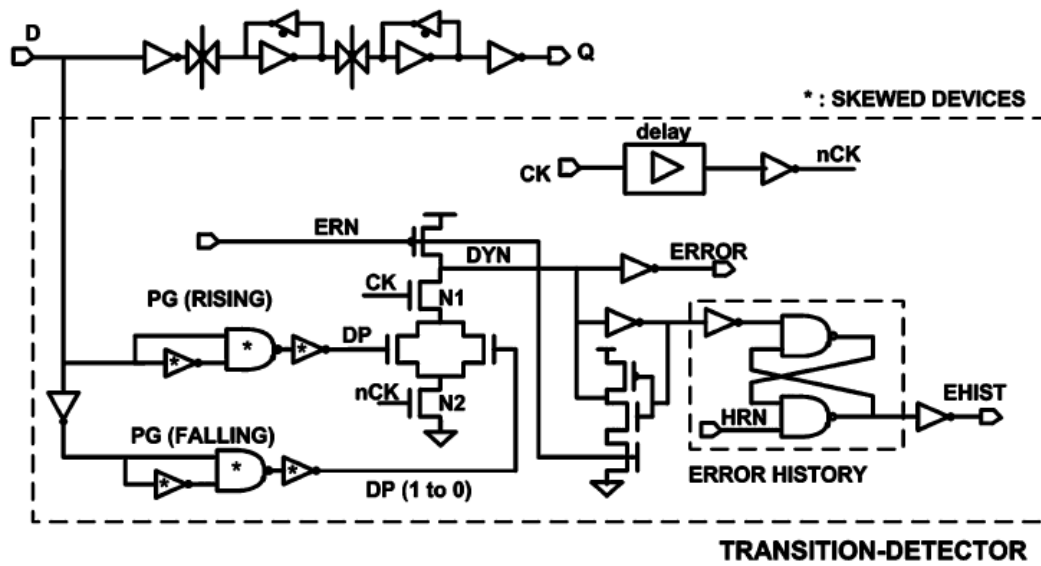
2.2 ARM's transition detector

Fig. 2.2 shows the architecture of a transition detector used in ARM's experimental RAZOR [6, 7] design. The complete circuit schematic in Fig. 2.2(a) shows an additional RS-latch structure to generate the error history bit, but here we focus on the transition detector which is extracted in Fig.2.2(b). In principle, this transition detector design shares the same basic idea for operation as Intel's TDTB. They both employ circuits to generate positive pulses based on input transitions that discharge the output of a precharged dynamic node during the evaluation phase of the clock. The difference between the two lies in the fact that the ARM's transition detector employs two explicit circuit paths to generate pulses for the rising and falling transitions respectively, whereas the TDTB combines the functions at the (relatively high) cost of an extra EXOR gate.

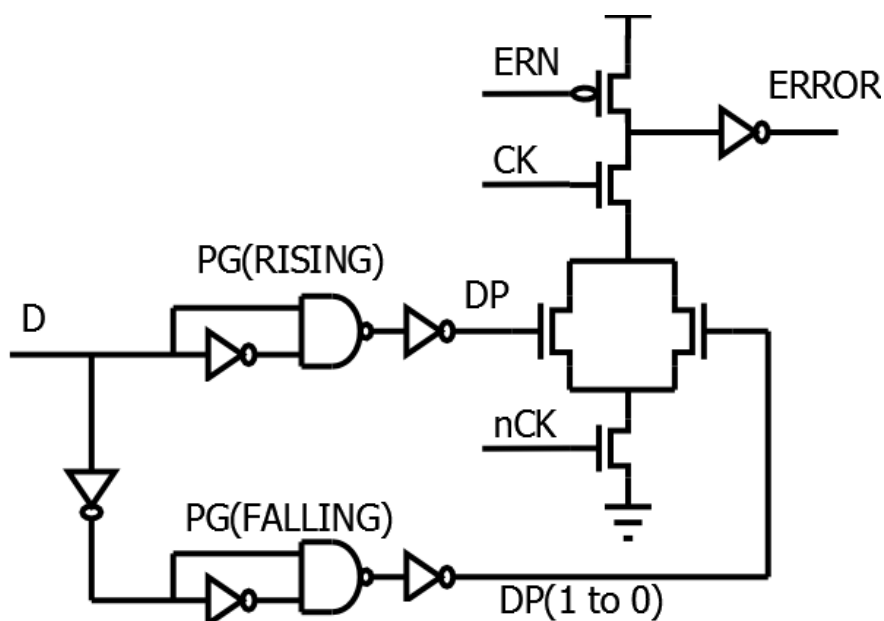
It can be observed that the upper path in Fig. 2.2(b) will only generate a pulse for rising (0 to 1) transitions at D. The initial 0 value at D initializes the lower input of the upper NAND gate to a 1 through the inverter. Now when D switches to 1, the inverter delay briefly keeps this 1 at the lower NAND input while the upper input transits to the new 1 value, causing the output of the NAND to briefly pulse low. This pulse discharges the dynamic NOR output gate, signaling a transition (error). Falling transitions are similarly detected via first converting them into rising transitions by inverting the input signal as shown along the lower path in Fig. 2.2(b). Notice that the precharged NOR gate here is controlled by two separate clocks, CK and nCK (the delayed and inverted CK). This transition detector is therefore active over the window when both CK and nCK are high, while Intel's TDTB designs detects transitions during the entire high clock window. Thus in the ARM design, the width of error detection phase can be controlled by the inversion delay of the inverter on nCK (delayed clock) network. However, if needed, such an

adjustment to the detection window can be made via the precharge circuit in any design, and therefore is not of primary importance in our discussion.

The same can be said for Intel's TDTB, ARM's transition detector, which can detect datapath metastability due to the delay of pulse generating circuit. In terms of hardware consumption, the pulse generator takes 18 transistors to be implement, which is the same with TDTB.



(a)



(b)

Figure 2.2 The transition detector used by ARM in the prototype RAZOR design [7]. (a) Schematic of the entire design [7]. (b) Enlarged schematic of the transition detection circuit

2.3 The need for a high performance and low area overhead transition detector.

Although the two transition detectors from Intel and ARM have provided a approach to metastability control and error detection in BTWC designs, they impose a significant area overhead. This is particularly so because virtually all flip-flops in a BTWC design must be protected by such circuits. Thus there is great motivation to reduce the area overhead of transition detectors. Observe that both the transition detectors employ duplicated circuits in their pulse generator designs. Since the transition detectors are incorporated in all the flip-flops, this duplication imposes large area overhead in a processor chip. Our effort is focused on eliminating the duplicated circuits to reduce area overhead while retaining high performance timing-error detection features.

In the next chapter, the proposed design of a transition detector will be introduced.

Chapter 3

PROPOSED TRANSITION DETECTOR EXPLOITING CHARGE SHARING

In this chapter, we will present the proposed design of a transition detector called “transition detector exploiting charge sharing” (TDCS). Our proposed design is able to detect both rising and falling transition without circuit duplication by exploiting charge sharing in the output gate in a novel way. The design complexity is greatly reduced in comparison to the current design. The core concept and operation principle will be illustrated in detail.

Some key structures and concepts in TDCS include dynamic CMOS gate structure, short circuit effect and charge sharing. In the subsequent section, we will briefly review these structures and concepts.

3.1 Dynamic CMOS gate

3.1.1 Dynamic CMOS gate principles

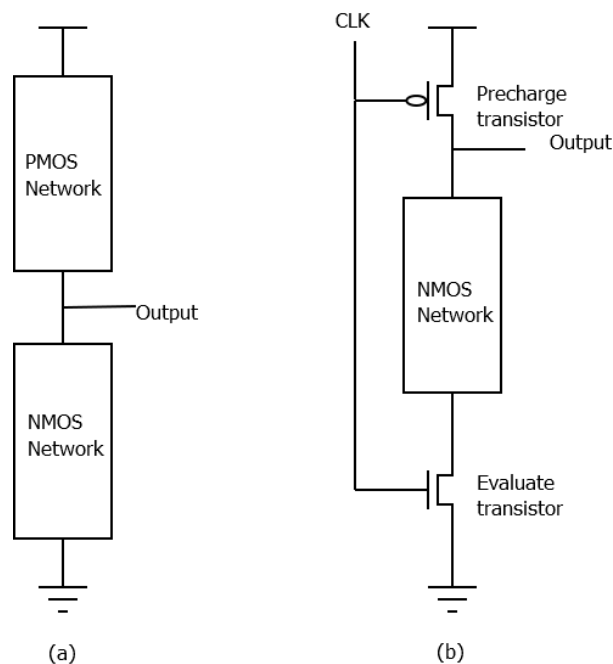


Figure 3.1 a) CMOS gate structure b) dynamic CMOS gate structure

By the mid 1980's, complementary metal oxide semiconductor (CMOS) started to become the primary choice for digital semiconductor designs due to the advantage that CMOS logic gates dissipate almost no power when the inputs to the gate do not change. However, CMOS logic requires a large circuit area due to logic redundancy and CMOS logic can get slow in series connections.

One alternative to static CMOS gates is dynamic CMOS gates. Dynamic gates as shown in Fig. 31.b, employ a clocked logic. There are two phases in one clock cycle: precharge (low clock) phase and evaluation (high clock) phase. When the clock is low, the dynamic gate is in precharge phase where the precharge transistor at the top is on and the evaluation transistor at the bottom is off. The load capacitances at the output node consists of the next logic level gate capacitances, diffusion capacitances and interconnect capacitances are exposed to the voltage supply. Thereby the output node is precharged to logic high. As the clock transitions to logic high, the dynamic gate goes to evaluation phase where the NMOS network is cut off from voltage source and evaluation transistor is turned on. At this time, the logic value at output node depends on the state of NMOS network. If a path is activated in the NMOS network, the load capacitance at output node is discharged to logic low. Otherwise, the output node stays charged high.

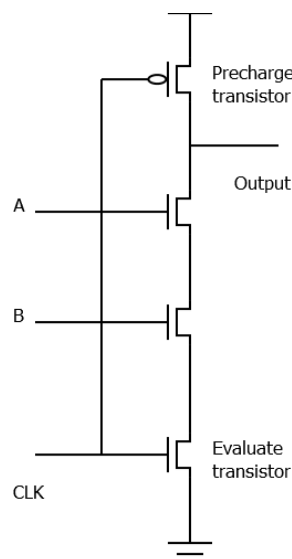


Figure 3.2 dynamic NAND gate

A simple example will now be provided which shows how a dynamic gate works. Fig. 3.2 is a dynamic NAND gate. In precharge phase, the output node capacitance is charged to logic high. As the clock transitions to logic low, the states of node A and B determine the value of output. If both A and B are logic high, both of the transistors are turned on. Therefore, there is a pull-down path to discharge the output node load capacitance and output node is logic low. In any of the other cases, there is no path for discharging. Thus, the output remains logic high.

Typically, dynamic logic circuits are faster conventional CMOS gates, while the requirement of dynamic gates for noises and glitches compression is more stringent than CMOS gates.

3.1.2 Issues with dynamic CMOS gates

Dynamic CMOS gates are more sophisticated than conventional CMOS gates, so it is possible that issues such as charge sharing and charge leakage may arise when designing a CMOS gates. It is interesting to note that some of the issues are used to help achieve the design goal of the proposed transition detector design.

Charge sharing issue.

In Fig. 3.3, we suppose nodes A, B, C and D are 0, 1, 1 and 1 during the precharge phase. The dynamic node capacitance is charged while all the capacitances at the intermediate nodes remain uncharged. If input D drops to 0, then input A rises to 1 during the evaluation phase, the charges stored in the capacitance of the dynamic node will redistribute to the intermediate nodes, which leads to the undesired voltage drop at the dynamic node. The final voltage of the dynamic node can be calculated as follows;

$$V_{dyn} = \frac{2C}{2C+3C}VDD = 0.4VDD \quad (2)$$

The result is low enough to switch the output inverter. Typically, charge sharing should be avoided in VLSI circuit design. However, in this particular work, charge sharing is

utilized to reduce the voltage of the dynamic node (to be explained in detail in later part of this chapter).

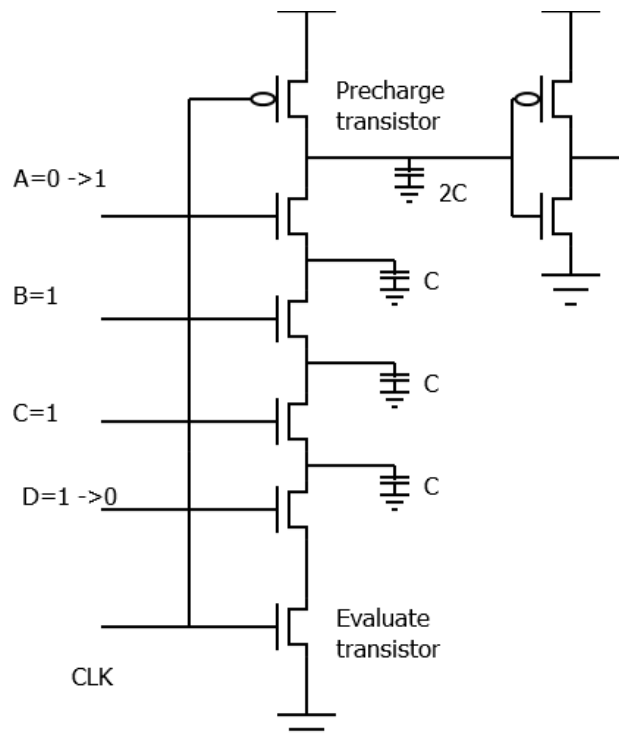


Figure 3.3 Charge sharing in domino CMOS gate.

Charge leakage

Another issue with dynamic gates is charge leakage, which happens in every dynamic gate. During the evaluation phase, there still will be a small but finite current discharging the dynamic node even though all the transistors are turned off. This is the leakage current. In CMOS gates, leakage power is a part of the static power which should be minimized. However, it does not change the logic state since the output voltage level depends on the resistance ratio of the pullup network and pulldown network. In domino logic, leakage can be a crucial problem because the stored charge will not be compensated until the next precharge phase.

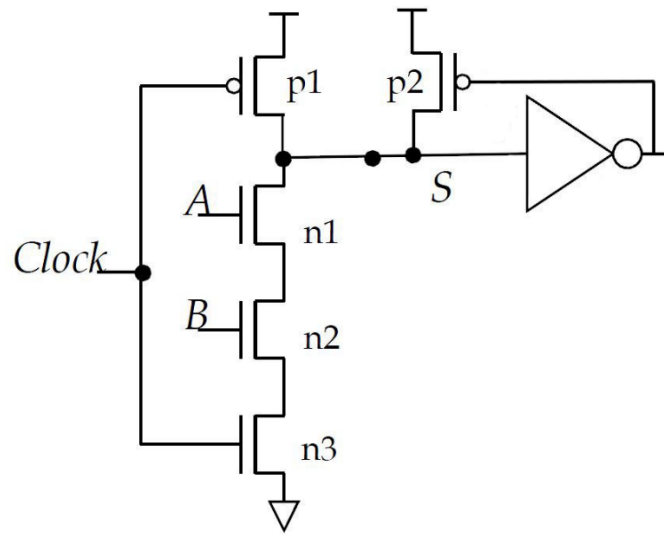
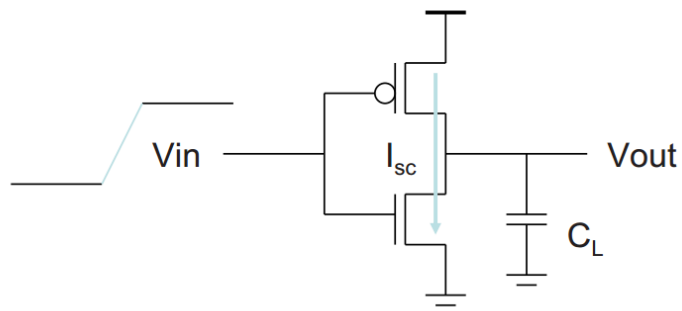


Figure 3.4 Domino AND gate with keeper transistor

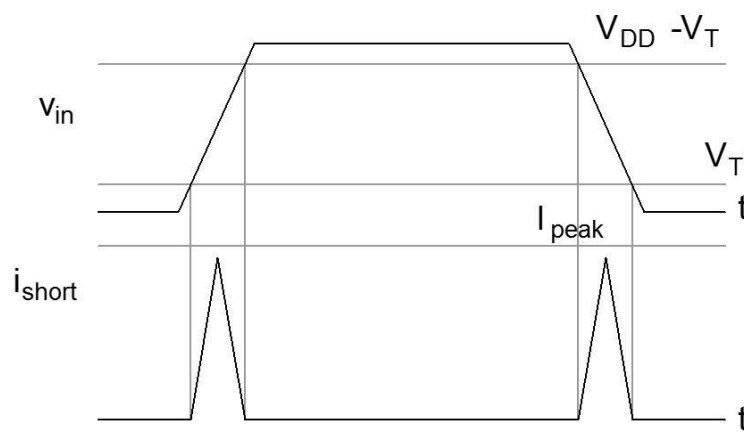
The solution to leakage is illustrated in Fig. 3.4. A keeper transistor p2 is added to the dynamic node. The keeper is simply a weak PMOS transistor with a gate that is controlled by the output. Therefore, the voltage drop caused by charge leakage can be recovered by the keeper transistor.

3.2 Short circuit effect in CMOS

Another core concept in the proposed transition detector design is short circuit effect. Taking a CMOS inverter as an example (Fig. 3.5), pullup and pulldown transistors can be both conducting for a short period of time as the input transitions from low to high or high to low since the slope of the transition of input signal is not infinite. During this time period, a pulldown path is created which leads to short circuit current from VDD to GND.



(a)



(b)

Figure 3.5 (a) Short circuit current in CMOS inverter (b) Approximate short circuit current

Notice that the short circuit current is extremely sensitive to the relation between $V_{thn} + V_{thp}$ and V_{DD} . In the limit that $V_{thn} + V_{thp} > V_{DD}$, short circuit current can be entirely eliminated because the pullup and pulldown network are never simultaneously conducting. Therefore it is possible to increase the short circuit current by raising the supply voltage.

In conventional CMOS circuit design, short circuit effect must be minimized to reduce undesired power consumption. However, in our proposed design of transition detectors, a short circuit is utilized to detect signal transition.

3.3 Design of transition detector exploiting charge sharing

Fig. 3.6 is a schematic representation of our new transition detector (TDCS) which exploits charge sharing to detect rising transitions while using a single falling transition detection circuits. Thus, TDCS employs a combination of two concepts: short circuit discharge and charge sharing, to detect both falling and rising transitions respectively, using a single circuit structure.

3.3.1 Falling transition detection by short circuit discharge

The concept that is utilized to detect a falling transition during the evaluation phase is short circuit discharge in the ARM transition detector. The key idea is to discharge the dynamic node N (during the evaluate high clock phase) using a short circuit temporarily created when input A (M6) briefly stays high because of the inverter (Inv2) delay after a 1 to 0 transition at D, while the B input (M7) switches to a high.

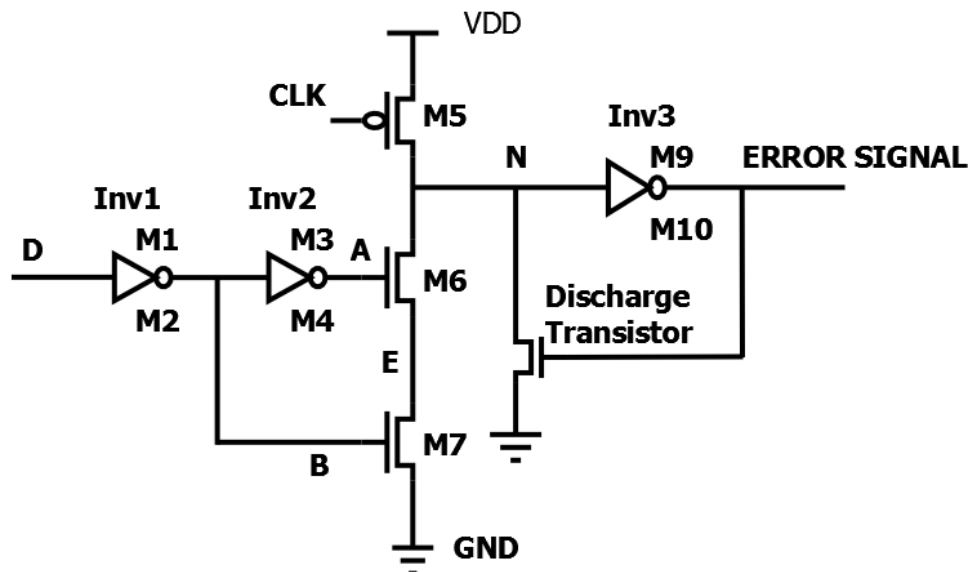


Figure 3.6 Transition detector exploiting charge sharing and short circuit discharge for detection rising and falling transitions respectively

Fig. 3.7 illustrates the falling transition detection process. Initially, D is high at the beginning of the evaluation phase, meaning that M6 is ON and M7 is OFF. The load capacitance C_n at dynamic node N is charged by the voltage supply. As D transitions to low, M7 is turned on while M6 is still ON due to the delay of Inv2. Then, C_n is discharged through the pulldown path and the error signal is activated. After the delay of Inv2, M6 is turned off and therefore, the pulldown path is closed.

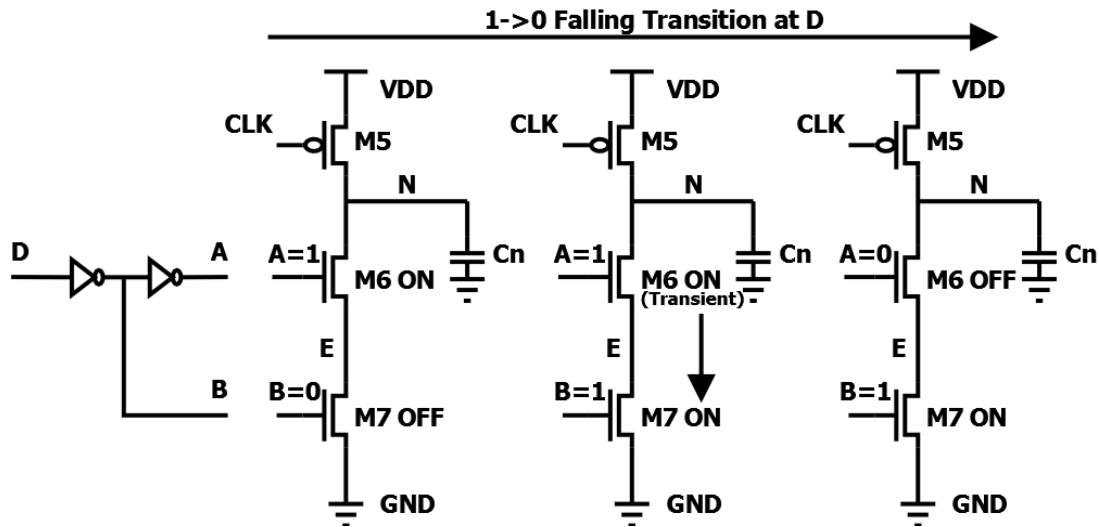


Figure 3.7 Falling transition detection process through short circuit discharge

3.3.2 Rising transition detection by charge sharing

For a falling transition, a reliable short circuit is created using the delay of Inv2. However, when it comes to a rising transition, the delay of Inv2 prevents the creation of a short circuit discharge path since M6 switches prior to M7. A new mechanism to discharge the dynamic node is needed for this transition.

Here, we employ charge sharing for rising transition detection. Charge sharing, which can cause an undesirable voltage drop at dynamic nodes, is a common phenomenon in dynamic CMOS gates. Commonly, this can degrade signals and must be minimized. However, here we exploit charge sharing in a novel way by using it to discharge the output node for the complementary transition. Fig. 3.8 illustrates the process of rising transition detection.

At the beginning of the evaluation phase, C_n is charged high and C_e is discharged low. As the D input transitions to high, M7 is turned off first. After the inversion delay of Inv2, M6 is turned on. At this time, a path from node N to node E is created. The charge at node N redistributes and is shared with node E due to the voltage difference between the two nodes. This phenomenon is known as charge sharing. Charge sharing continues until the voltages of the two nodes equalize. Thus, the voltage at node N decreases in proportion to the relative capacitances C_n and C_e . If the capacitances are equal, the voltage at node N will drop to half of its initial value. Making C_e two or more times larger than C_n can ensure that the voltage at N drops sufficiently for it to be reliably seen as a low by the output inverter, resulting in a high error signal indication from a transition.

In order to reduce the voltage to be as low as possible, M7 must have a relatively large size compared to M6, M5, M9 and M10 to ensure that C_e is larger than C_n . However, charge sharing obviously cannot reduce the voltage of the dynamic node N to a perfect “0.” Furthermore, it requires unrealistic sizing of M7 to drop the voltage of the dynamic node close to 0.1V or below.

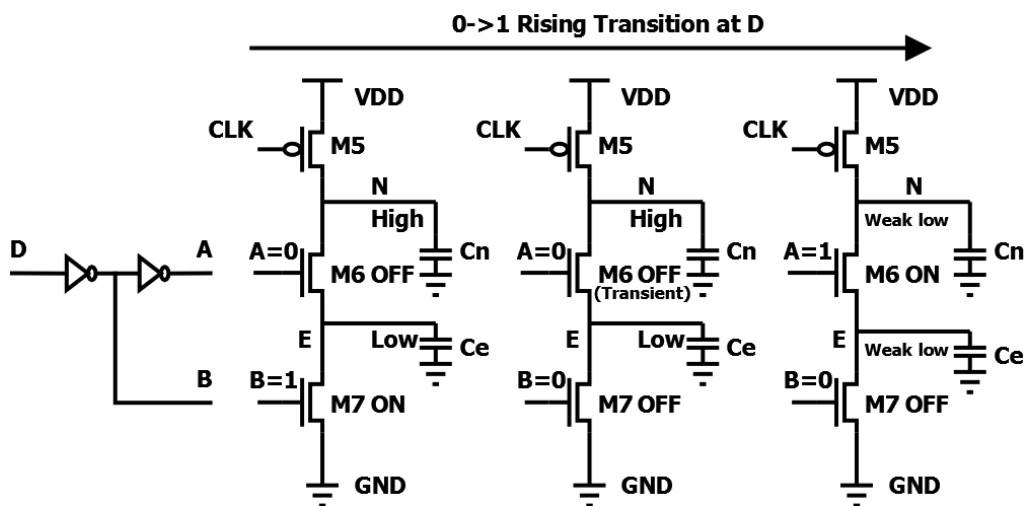


Figure 3.8 Rising transition detection through charge sharing (Note $C_e > C_n$)

To fully discharge the capacitance at the dynamic node in this situation, a new discharge mechanism is introduced. A weak N-transistor is connected to dynamic node as shown in Fig. 3.6. During the precharge phase, this discharge transistor is “OFF” since the ERROR

SIGNAL node is low. If a transition occurred in evaluation phase, the voltage drop at dynamic node would lead to a voltage rise at the ERROR SIGNAL output node. This causes the discharge transistor to turn on and speed up the discharge process until the capacitance is fully discharged.

This discharge transistor not only helps achieve a full discharge during charge sharing based transition detection, but also relaxes the sizing requirements for falling transition detection to fully discharge the dynamic node, depending solely on short circuit. However, the discharge transistor must have a small size compared to M5 for obvious reasons. Small sizing allows quick voltage recovery of the dynamic node at the beginning of the precharge phase. Also, it adds minimal diffusion capacitance to dynamic node.

Notice that although we employ a dynamic gate structure in the proposed design of TDCS, the evaluation transistor in the pulldown network is removed because it may hinder the charge sharing based detection at the beginning of evaluation phase. Additionally, the performance of TDCS is improved by removing the evaluation transistor since the total resistance on the pulldown path is reduced. While removing the evaluation transistor can result in increased dynamic power consumption during input transitions in the precharge phase, this is not expected to add significantly to the very large number of spurious switching transients (glitches) commonly encountered in CMOS circuits since the number of flip-flops is typically small compared to the total gate count in microprocessors.

3.4 Evaluation of TDCS

In this work, TDCS is implemented in a 45nm high performance version of predictive technology model (PTM). TDCS is a precise circuit that needs to be carefully sized and optimized to get the best performance. In the following section, the sizing of TDCS and SPICE simulation will be introduced.

3.4.1 Transistor sizing in TDCS

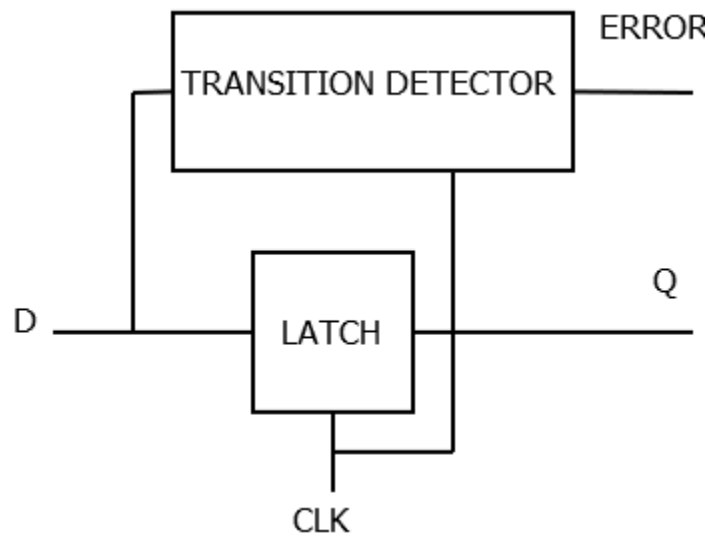


Figure 3.9 The architecture of TDCS with a latch in the datapath

Fig.3.9 illustrates the architecture of the proposed error detection circuit. Note that as in the Intel design, the TDCS is shown with a latch in the datapath. A flip flop can also be used. This latch is a standard latch from NanGate FreePDK45 Open Cell Library. In order to obtain best performance, the sizing of TDCS needs to be carefully optimized.

There are several aspects of note in the sizing process. The size of M7 in Fig. 3.6, must be relatively large compared to M6 to provide a large diffusion capacitance for charge sharing. The discharge transistor must be relatively small compared to M5 to get fast voltage recovery and good noise compression in each precharge phase. Inv2 is a skewed inverter in this design to provide a better performance.

Table 3.1 shows the optimal sizing for each transistor in TDCS, obtained through simulation. Note that nominal transistor sizing in this 45nm cell library is a relative width $N=10$, $P=15$ units. So except for M7, the remaining transistors fall within traditional sizing range.

Sizing	M1	M2	M3	M4	M5
TDCS	10	5	25	5	15
Sizing	M6	M7	M9	M10	Discharge transistor
TDCS	20	75	8	4	5

Table.3.1 Transistor sizing of TDCS

3.4.2 SPICE simulation of TDCS

Fig. 3.10 shows the simulation result of TDCS. Note that there are quick charge and slow discharge patterns in the last two evaluation phases. The quick rise of voltage is due to capacitive coupling and the slow drop can be attributed to leakage. These patterns will not affect the function and performance of TDCS.

Fig. 3.11 shows simulation details of rising and falling transition detection. In the detection of a falling transition, Fig. 3.11(a), a transient short circuit, with both pulldown transistors momentarily on, is employed to discharge node N. The voltage of the intersection point of node A and B's curves is above V_{th} . (0.46V for NMOS), which shows that a short circuit is indeed created. The voltage of node N starts to drop when the voltage of node B exceeds the threshold voltage. The voltage drops quickly at the beginning until the voltage of node A decreases below the threshold voltage, turning off the discharge. Then the voltage of node N is low enough to turn on the discharge transistor. The voltage continues to drop at a slow rate since the discharge transistor is initially not fully turned on. As the voltage of node N drops and the discharge transistor turns on more strongly, the discharge current grows larger which leads to a faster voltage drop rate. The discharge process does not stop until node N is fully discharged.

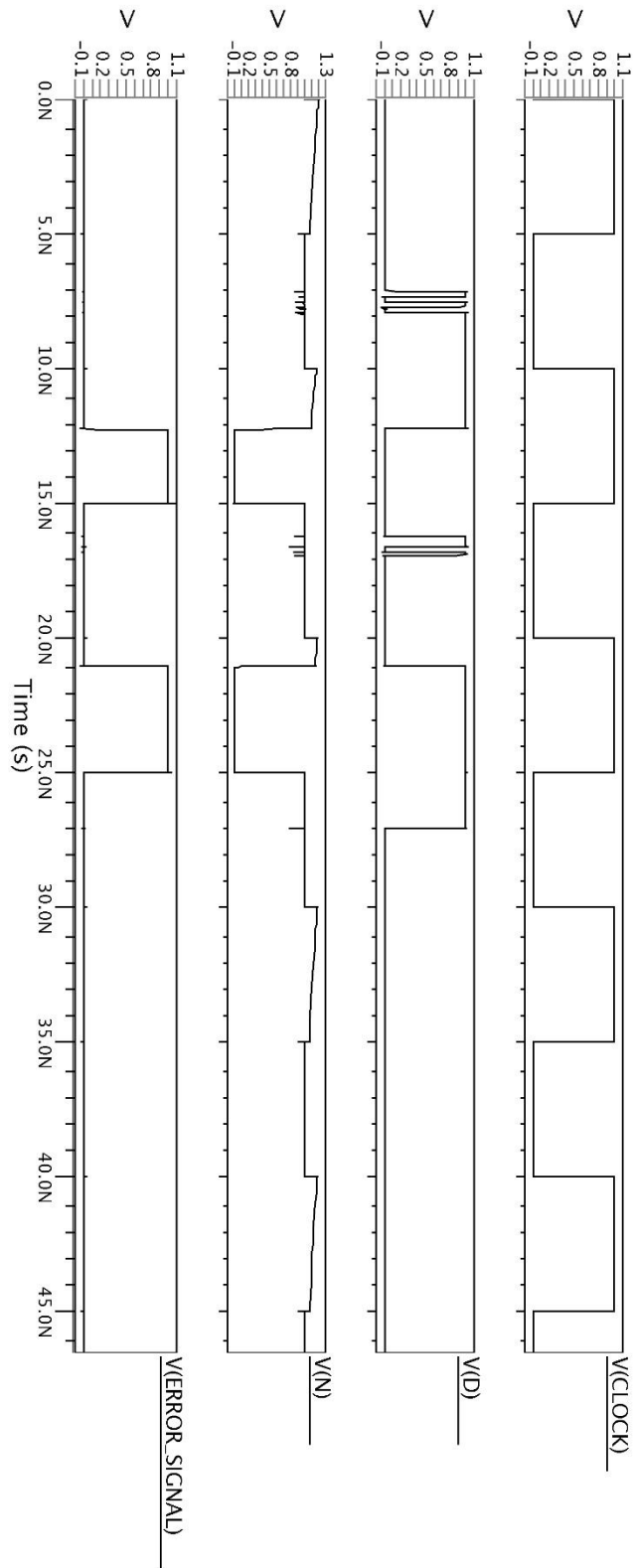
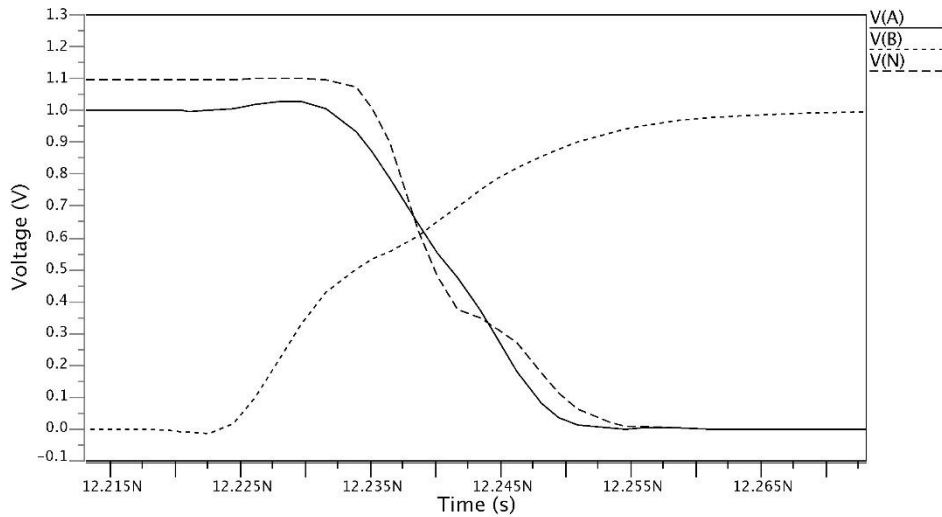
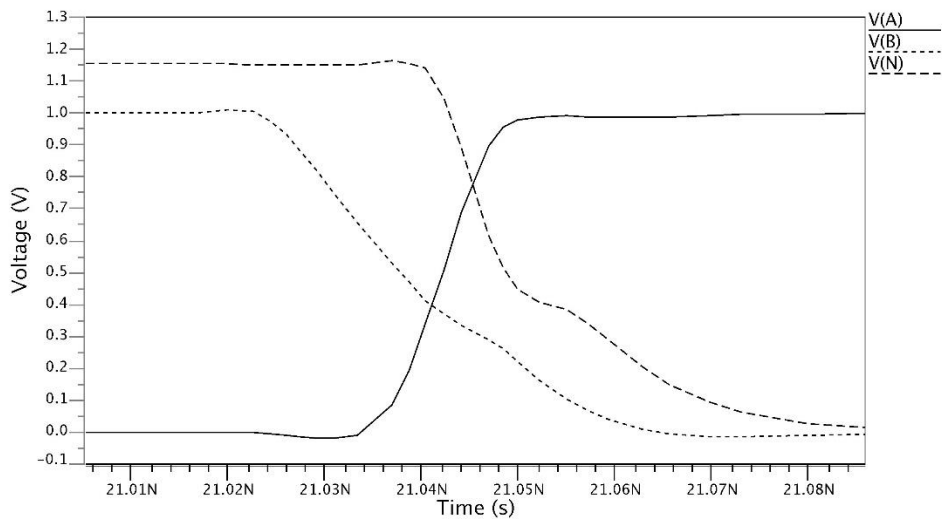


Figure 3.10 Simulated timing-error detection demonstration



(a)



(b)

Figure 3.11 Detailed view of rising and falling transition detection. (a) Falling transition detection by short circuit. (b) Rising transition detection by charge sharing.

The detection of a rising transition is shown in Fig. 3.11(b). The voltage of the intersection point of node A and B is below the NMOS threshold voltage, meaning that no short circuit is created. Charge sharing starts when the voltage of node A rises above the NMOS threshold voltage. The redistribution of charge brings down the voltage of node N. At 21.05 ns, the voltage drop rate slows down indicating that the redistribution

has finished. Then the discharge transistor starts to take over the discharge process until node N is fully discharged.

In a performance simulation we measured the detection delay for both rising and falling transitions and compared them to the D-to-Q delay and CLK-to-Q delay of the datapath latch. The results are shown in Table 3.2. The detection delay for falling and rising transitions are 21.04 ps and 32.00 ps respectively. The rising transition detection is 52.09% slower than the falling transition. The difference in performance is due to the delay of Inv2. Charge sharing starts when M6 is turned on in the detection of a rising transition while a short circuit starts when M7 is turned on in the detection of a falling transition. Note that M7 always switches earlier than M6 due to the delay introduced by Inv2, which explains why rising transition detection is slower than falling transition. Comparing the performance of TDCS and the datapath latch, TDCS is 0.19% faster than the latch CLK-to-Q delay and 11.99% faster than the latch D-to-Q delay. This shows TDCS performs quite fast and, at the same level as a standard latch. The transition detection is quite quick.

TDCS Performance	<i>TDCS H-L transition/ps</i>	<i>TDCS L-H transition/ps</i>	<i>Detection delay/ps</i>
Delay	21.04	32.00	26.52
Latch CLK-to-Q delay	<i>Latch H-L CLK-to-Q/ps</i>	<i>Latch L-H CLK-to-Q/ps</i>	<i>Latch CLK-to-Q/ps</i>
Delay	25.88	27.26	26.57
Latch D-to-Q delay	<i>Latch H-L D-to-Q/ps</i>	<i>Latch L-H D-to-Q/ps</i>	<i>Latch D-to-Q/ps</i>
Delay	25.22	22.14	23.68
TDCS performance relative to Latch CLK-to-Q delay		-0.19%	
TDCS performance relative to Latch D-to-Q delay		11.99%	

Table 3.2 TDCS performance

At this point, we have verified that the exploitation of charge sharing by our transition detector is feasible under ideal conditions. However, in reality, the circuit must be able to work under a certain range of process and environmental variations. In the next chapter, our goal is to test the robustness of TDCS

3.5 Evaluation of area overhead

The major advantage of the proposed TDCS over conventional transition detectors lays in its low area overhead. TDCS employs only 10 transistors, which is 60% fewer than TDTB and 62.3% fewer than ARM's transition detectors, while still retaining the same functionality as the Intel and ARM's designs. Here, we did two area overhead estimations, one for latch-based design and the other is for flip-flop-based design. Suppose flip-flops and latches take up 30% and 20% of total area in a microprocessor respectively. Additionally, we assume a flip-flop employs 21 transistors and a latch employs 11 transistors according to typical implementation of flip-flop and latch. The area overhead can be calculated by the following equation:

$$A_{overhead} = \left(\frac{N_{TD}}{N_{DFF/Latch}} \right) \cdot Area Usage_{DFF/Latch}$$

Where N_{TD} is the transistor count of the target transition detector and $N_{DFF/Latch}$ is the transistor count of the flip-flop or latch on the datapath. $Area Usage_{DFF/Latch}$ is the percentage of area that flip-flop or latch takes up.

Name of Design	Transistor count	Area overhead estimation (Latch based design)	Area overhead estimation (Flip-flop based design)
TDCS	10	18.2%	14.3%
Intel's TDTB	25	45.5%	35.7%
ARM's transition detector	27	49.1%	38.6%

Table 3.3 Area overhead comparison

The results in Table 3.3 show that the area overhead of TDCS is only 18.2% for latch-based design and 14.3% for flip-flop-based design, smaller than both TDTB and ARM's designs.

Chapter 4

TDCS DESIGN ROBUSTNESS

4.1 Introduction

Performance and functional goals are not the end of design. A good, reliable, design must ensure billions of transistors get to all functions for quintillions of consecutive cycles. Transistors are so small that printing errors below the wave length of light and variation in the number of dopants atoms have major effects on their performance. During the operating lives of transistors, the environment temperature can vary from freezing to boiling. Additionally, the fluctuation in supply voltage can have great impact on performance of integrated circuits.

Integrated circuits are designed to work for a range of temperatures and voltages, rather than a single temperature and voltage. These have to work under different environmental conditions and different electrical setup and user environments. Conventional static CMOS circuits are exceptionally well suited to the task because they have great noise margins, are minimally sensitive to variations in transistor parameters, and will eventually recover even if a noise event occurs. However, in the proposed design of TDCS, the error detection process relies on a short circuit current which is sensitive to process and environmental variations. Thus, the design robustness of TDCS must be tested.

So far, we only tested TDCS under ideal condition. Different kinds of variations must be introduced to test the robustness of TDCS. In general, there are three different sources of variation—two environmental and one manufacturing:

- Process variation
- Supply voltage
- Operating temperature

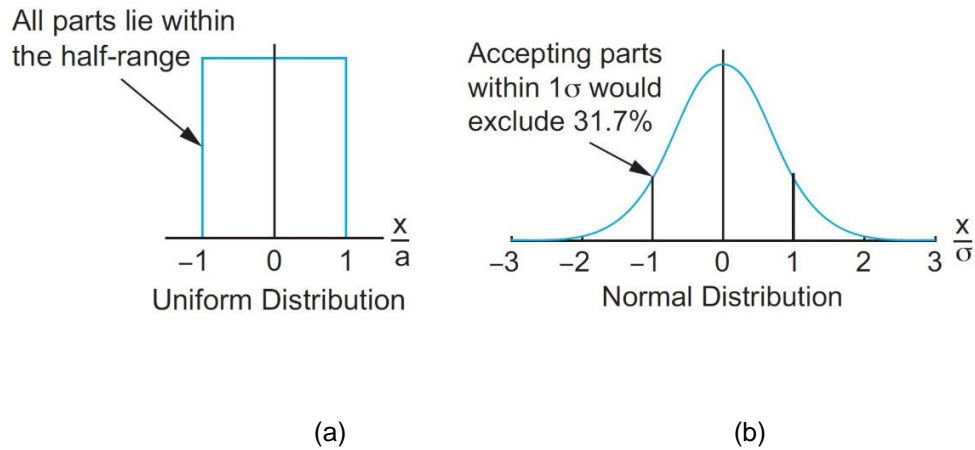


Figure 4.1 a) Uniform distribution b) Normal distribution [15]

The variation sources are also known as Process, Voltage, and Temperature (PVT). Variations are usually modeled as uniform or normal (Gaussian) statistical distributions as shown in Figure 4.1. Uniform distribution is helpful in describing environmental variations such as temperature and supply voltage. Process variations are usually modeled as normal distributions.

4.2 Process variation

Process variation accounts for deviations in the semiconductor fabrication process. Variations in the process parameters include impurity concentration densities, threshold voltage, oxide thicknesses, dimension of devices and diffusion depths. These variations are caused either by non-uniform conditions during depositions and/or diffusions of the impurities or limited resolution of the photolithographic process.

In this particular work, we take variations in devices' dimension, oxide thickness and threshold voltage into consideration.

4.3 Supply voltage

Systems are designed to operate at a nominal supply voltage, but this voltage may vary for many reasons including tolerances of the voltage regulator, IR drops along supply rails,

and di/dt noise. Therefore, we must ensure that our design works under the highest possible supply voltage variation range. Typically, a digital system must be able to work robustly under +/-10% supply voltage variation.

4.4 Temperature

Similar to supply voltage, temperature cannot be always the same on a chip; some parts of a chip can be hotter than the others due to more activities in these parts. Temperature can affect the performance of a chip. Typically, the drain current of a transistor increases as the temperature increases, thus the performance increases as well. The temperature of a chip is determined by the power consumption and package thermal resistance.

Standard	Minimum	Maximum
Commercial	0 °C	70 °C
Industrial	-40 °C	85 °C
Military	-55 °C	125 °C

Table 4.1 Ambient temperature ranges [15]

Table 4.1 lists the ambient temperature ranges for parts specified to commercial, industrial, and military standards. Since our design is meant for commercial use, the acceptable temperature variation range is set to 0-70°C

4.5 PVT Corners

In order to account for PVT variations, traditionally designers verify the circuit functionality and performance under extreme process, voltage and temperature conditions, assuming that a circuit that functions and performs adequately at the extremes should perform properly at nominal conditions [16, 17]. This method is known as corners analysis, where different SPICE model decks are used to determine circuit response at extreme PVT conditions.

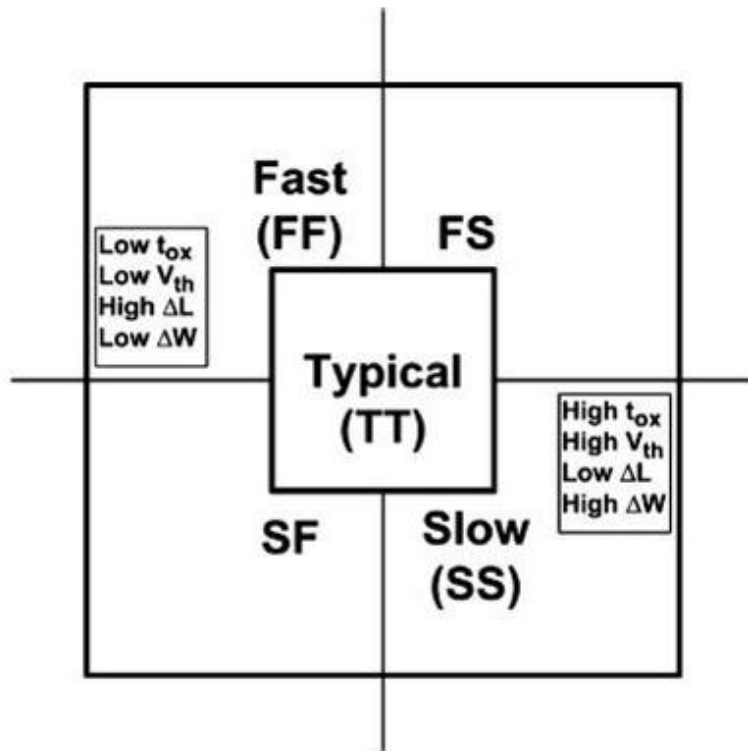


Figure 4.2 Five different types of process corners

Figure 4.2 illustrates the five different corners when only process variation is taken into consideration. F stands for fast, meaning that all the variations are in the same direction of making the device faster. In opposition, S stands for slow, meaning that all the variations are in the same direction of making the device slower. Notice that here are two acronyms in the definition of a particular corner. The first acronym represents n channel transistors and the second one represents p channel transistors

Since process variation information of 45nm technology is not accessible for academic use, we explored the maximum process variation tolerance of TDCS. We assume the maximum process variations of width, length, threshold voltage, and oxide thickness are within +/-5%.

Process corners								
	NMOS				PMOS			
	Width	Length	Vth/V	Tox/ns	Width	Length	Vth/V	Tox/ns
F	W+ Δ	L+ Δ	0.45	1.19	W+ Δ	L+ Δ	-0.47	1.23
T	W	L	0.47	1.25	W	L	-0.49	1.3
S	W- Δ	L- Δ	0.49	1.31	W- Δ	L- Δ	-0.52	1.36

Table 4.2 Process corners

Environmental Corners		
Corner	Voltage/V	Temperature/°C
F	1.1	0
T	1	25
S	0.9	70

Table 4.3 Environmental Corners

Corners			
NMOS	PMOS	Voltage	Temperature
F	F	F	F
F	F	F	S
F	F	S	F
F	F	S	S
F	S	F	F
F	S	F	S
F	S	S	F
F	S	S	S
S	F	F	F
S	F	F	S
S	F	S	F
S	F	S	S
S	S	F	F
S	S	F	S
S	S	S	F
S	S	S	S
T	T	T	T

Table 4.4 PVT corner list.

Corner Analysis	TDCS H- L /ps	TDCS L- H /ps	Detection delay/ps	Latch H-L CLK-Q/ps	Latch L-H CLK-Q/ps	Latch CLK- Q/ps	Latch H-L D-Q/ps	Latch L-H D-Q/ps	Latch D- Q/ps	TDCS performance relative to Latch CLK- to-Q delay	TDCS performance relative to Latch D-to-Q delay
FFFS	10.94	17.76	14.35	17.86	19.39	18.63	15.13	15.07	15.10	-22.95%	-4.97%
SSSF	62.94	71.59	67.27	48.30	48.02	48.16	51.83	39.97	45.90	39.67%	46.55%
TTTT	21.04	32.00	26.52	25.88	27.26	26.57	25.22	22.14	23.68	-0.19%	11.99%
Variation			-45.89%			-29.90%			-36.23%		
			153.64%			81.26%			93.83%		

Table 4.5 Corner analysis results.

Tables 4.2 and 4.3 show the process and environmental corners that are simulated in this work. For process variation, we allow $\pm 5\%$ variation in minimum feature size, threshold voltage and oxide thickness. Δ in Table 4.2 represents the maximum variation in width and length, which is 5% of minimum feature size. For voltage corners, the maximum allowed variation is $\pm 10\%$. In terms of temperature corner, $0-70^{\circ}\text{C}$ is the typical range for commercial ICs. The nominal voltage and temperature are 1.0V and 25°C , respectively. For supply voltage, we allow 10% variation in supply voltage. Temperature is allowed to vary from $0^{\circ}\text{C}-70^{\circ}\text{C}$. Table 4.4 shows all the 17 corners (including the 'TTT' corner), which were simulated.

4.6 Corner analysis results

A total of 17 corners including a typical corner are simulated in HSPICE. The D-to-Q delay and CLK-to-Q delay of the datapath latch are also measured in this simulation to compare the influence of PVT variation. Table 4.5 shows the results of two worst-case corners and the typical corner.

Simulation results have confirmed the error detection feature of TDCS functions properly at all 17 process corners. The detection delay of TDCS varies from -45.89% to 153.64% relative to the typical case. Although the performance variation of TDCS is larger than the variation of the D-to-Q delay and CLK-to-Q delay of the latch, this is expected since short circuit discharge and charge sharing are transient dynamic effects, which makes them more sensitive to PVT variations than static CMOS gates.

Chapter 5

CONCLUSION AND FUTURE WORK

An efficient transition detector is proposed which exploits charge sharing for the detection of rising switching transitions, along with traditional detection through short circuit discharge of the falling transition. Our approach employs a common circuit structure to detect both transitions and reduces area overhead to less than half compared to traditional designs. HSPICE simulations of this TDCS design show that it reliably achieves the same functionality as state-of-the-art published designs with 60% fewer transistors. Furthermore, detailed corner analysis shows that our TDCS design is robust under extreme PVT variations. The proposed low cost design appears well suited for the increasing number of applications that are beginning to incorporate an online error detection capability to protect circuit flip-flops from metastability and timing errors due to single event upsets and PVT noise in highly scaled technologies.

Our transition detector design is optimized for BTWC design in this thesis. In fact, the transition detector can also be used in memory chips to sense changes in the address line in order to signal the memory and associated circuitry that information is going to be written. We could optimize TDCS for memory arrays, which can give us the advantage in both performance and area overhead.

TDCS is sensitive to supply voltage variation because it employs charge sharing and short circuit discharge to form the transition triggered logic. If we can redesign the transition detector into a state triggered logic circuit, the robustness of the transition detector would improve. For example, we have two logic functions. The first logic can detect logic “1”s, the second can detect logic “0”s. If we “and” the results of the two logic circuits, the transition detection function is generated. The potential benefits of this idea are higher tolerance to PVT variations and less strict requirement on the size of transistors.

Bibliography

- [1] Austin T, Bertacco V, Blaauw D, et al. Opportunities and challenges for better than worst-case design[C]//Proceedings of the 2005 Asia and South Pacific Design Automation Conference. ACM, 2005: 2-7.
- [2] Bentley B. Validating the Intel (R) Pentium (R) 4 microprocessor[C]//Design Automation Conference, 2001. Proceedings. IEEE, 2001: 244-248.
- [3] Allan A, Edenfeld D, Joyner W H, et al. 2001 technology roadmap for semiconductors[J]. Computer, 2002, 35(1): 42-53.
- [4] Ziegler J F. Terrestrial cosmic rays[J]. IBM journal of research and development, 1996, 40(1): 19-39.
- [5] Mukherjee S S, Weaver C T, Emer J, et al. Measuring architectural vulnerability factors[J]. Micro, IEEE, 2003, 23(6): 70-75.
- [6] Ernst D, Blaauw D, Austin T, et al. Razor: circuit-level correction of timing errors for low-power operation[J]. IEEE Micro, 2004, 24(6): 10-20.
- [7] Bull, David, et al. "A power-efficient 32 bit ARM processor using timing-error detection and correction for transient-error tolerance and adaptation to PVT variation." Solid-State Circuits, IEEE Journal of 46.1 (2011): 18-31.
- [8] Burks A W, Burks A R. First general-purpose electronic computer[J]. Annals of the History of Computing, 1981, 3(4): 310-389.
- [9] Metropolis N, Howlett J, Rota G C. A history of computing in the twentieth century[J]. 1980.
- [10] Moore G E. Cramming more components onto integrated circuits[J]. 1965.
- [11] Moore G E. Excerpts from a conversation with Gordon Moore: Moore's Law, 2005[J].
- [12] Makimoto T, Sakai Y. Evolution of low power electronics and its future applications[C]//Proceedings of the 2003 international symposium on Low power electronics and design. ACM, 2003: 2-5.

- [13] Gelsinger P P. Microprocessors for the new millennium: Challenges, opportunities, and new frontiers[C]//Solid-State Circuits Conference, 2001. Digest of Technical Papers. ISSCC. 2001 IEEE International. IEEE, 2001: 22-25.
- [14] Uppu R T. A High Throughput Multiplier Design Exploiting Input Based Statistical Distribution in Completion Delays[D]. Auburn University, 2013.
- [15] Weste N, Harris D. CMOS Vlsi Design[J]. A Circuits and Systems perspective, Pearson Addison Wesley, 2005.
- [16] Hanson D A, Goossens R J G, Redford M, et al. Analysis of mixed-signal manufacturability with statistical technology CAD (TCAD)[J]. Semiconductor Manufacturing, IEEE Transactions on, 1996, 9(4): 478-488.
- [17] Nardi A, Neviani A, Zanoni E, et al. Impact of unrealistic worst case modeling on the performance of VLSI circuits in deep submicron CMOS technologies[J]. Semiconductor Manufacturing, IEEE Transactions on, 1999, 12(4): 396-402.
- [18] Bowman K A, Tschanz J W, Kim N S, et al. Energy-efficient and metastability-immune resilient circuits for dynamic variation tolerance[J]. Solid-State Circuits, IEEE Journal of, 2009, 44(1): 49-63.