AALTO UNIVERSITY
School of Electrical Engineering

Akshay Sridharan

A Neoteric Delay-Initiated Transition Detector for Subthreshold Processors

Thesis submitted for examination for the degree of Bachelors of Science in Technology

Otaniemi, 31.05.2011

Thesis instructor:	Supervisor:
Matthew J. Turnquist, Researcher	Lauri Koskinen, Senior Researcher

Aalto University School of Electrical Engineering abstract of Bachelor's Thesis

Author: Akshay Sridharan

Title: A Neoteric Delay Initiated Transition Detector for Subthreshold Processors

Date: 31.05.2011 Language: English Number of pages: 30

Faculty: Faculty of Electronics, Communications and Automation

Instructor: Matthew J. Turnquist

Supervisor: Lauri Koskinen

Keywords: Timing error detection, transition detection, subthreshold, low power, inverter chain.

Preface

I would like to thank my instructor, Matthew J. Turnquist for the support, guidance and patience during this work. I am very thankful to him for giving me this chance to work under his project. I am also very thankful to Erka Laulainen and Luari Koskinen for their guidance and motivation. I also extend warm thanks to all the people in Electronic Circuits Design Lab. I really enjoyed working with all of you.

I would also like to thank my parents especially my mother and all my friends for the great support and friendship you have given me through all these years. Without you my life wouldn't be as happy.

Otaniemi,

Akshay Sridharan

Contents

Abstract

Preface

Contents

List of symbols

Operators

Abbreviations

1. Introduction. 6
    - 1.1. TED Techniques 7
        - 1.1.1. Technique1-TFD. 8
        - 1.1.2. Technique2-Razor. 9
        - 1.1.3. Technique2-TDTB. 9
        - 1.1.4. Technique2-TDTBsub. 10
2. Proposed Design. 14
    - 2.1. Operation Principles 11
    - 2.2. Sizing of SRAMsub 13
    - 2.3. Delay Characterization in an Inverter Chain. 15
        - 2.3.1. Sizing and Stages. 15
            - 2.3.1.1. Delay and Width. 15
            - 2.3.1.2. Power and Width. 17
            - 2.3.1.3. Delay and Length. 18
            - 2.3.1.4. Power and Length. 19
3. Simulation Results and Comparisons 21
    - 3.1. Operation. 21
    - 3.2. Uncertainty Region. 22
    - 3.3. Power Consumption. 24
4. Conclusion. 26
5. Future work. 27
    - 5.1. Theoretical Delay. 27
6. References. 29

List of symbols and abbreviations

V_{dd}	Supply Voltage
V_T	Threshold Voltage
t_p	Propagation delay
PVT	Process, voltage and temperature conditions
TED	Timing error detection
DITD	Delay initiated transition detector
PoFF	Point of first failure
TDTB	Time borrowing transition detector
TFD	Transient fault detection
TD	Transition detector
CLK	Clock
W_n	Width of NMOS
W_p	Width of PMOS
l_n	Length of NMOS
l_p	Length of PMOS
t_{pHL}	High to low propagation delay
t_{pLH}	Low to high propagation delay
k	Sizing factor

1 Introduction

Ultra low power applications have pushed circuits to operate in the subthreshold region, where the supply voltage (V_{dd}) is below the threshold voltage (V_T) of the transistors. When supply voltage is scaled below threshold voltage we obtain minimum energy consumption in digital CMOS logic [2]. Sensor network devices and other portable devices that require low speeds of operation and are energy constrained can benefit greatly by employing subthreshold circuits.

In subthreshold the effect of variation in process, voltage and temperature (PVT), cause design uncertainties. The devices thereby require not only low energy of operation but also need to be adaptable to these variations. To identify and adapt to these variations, three adaptive DVS methods are typically utilized: critical path emulators that track variations [4], [5], independent monitoring of variations with individual sensors [6] and scaling supply voltage until failure [7]. The latter technique accounts for local and global variations where data is processed.

Timing error detection (TED) is a form of scaling voltage until the point of first failure (PoFF). At PoFF, data transitions become slow and timing error occurs between stages of a pipeline. As the timing error rate increases beyond the PoFF, the recovery energy begins to grow due to the effort required to correct the errors. The tradeoff is between the recovery energy required to fix the errors and quadratic reduction in energy past the traditional safety margin. Current TED methods include RazorII [8], TDTB [9] and TDTBsub [1]. Both RazorII and TDTB are incapable of subthreshold operation but TDTBsub was designed to operate in the subthreshold region.

This thesis explores a radical new method to enable the operation of TED in subthreshold. A new TED design called Delay Initiated Transition Detector (DITD) is presented along with many inherent design challenges. The operation of the DITD is verified using a 65nm design kit.

1.1. TED Techniques

TED Systems have been researched widely and are considered as an effective method to reduce power consumption. In a TED system latches are inserted onto all critical paths of a pipeline. Any *ERROR* signals are passed to an OR gate which then forwards it to the Voltage Control block. This block determines the acceptable error rate and adjusts Vdd and or frequency of operation within the pipeline accordingly. In the pipeline, Vdd is decreased until the PoFF for a given frequency.

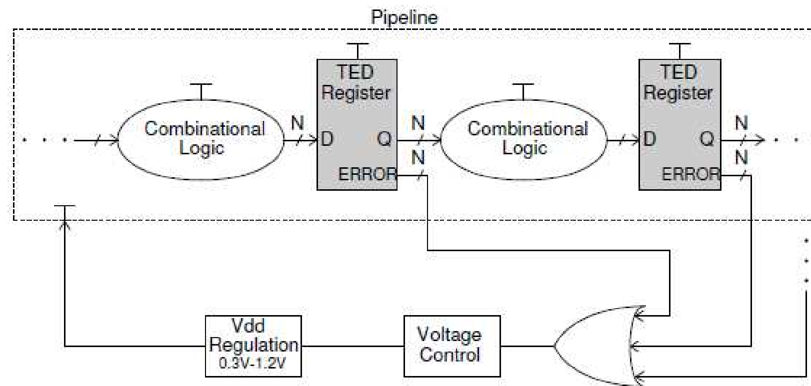

Figure1 System-level view of a generic pipeline system that uses TED register [1]

The TED latch has the ability to recognize a timing error and generate an error signal when data transition becomes too slow between combinational logic. For example, when the input data *D* to a TED latch transitions under a *CLK* low, the delay due to the value of Vdd is considered appropriate and no timing error signals are generated (Fig. 2). However, when *D* transitions under the time when *CLK* is high (i.e. the TED window), an *ERROR* is generated. This indicates that the delay due to the value of Vdd is too low for the combinational logic and thus timing errors will occur until Vdd is increased.

There have been two fundamental techniques in which the TED has been implemented. They differ in architecture at the circuit and system level. These are discussed in detail in the next sections 1.1.1-1.1.4

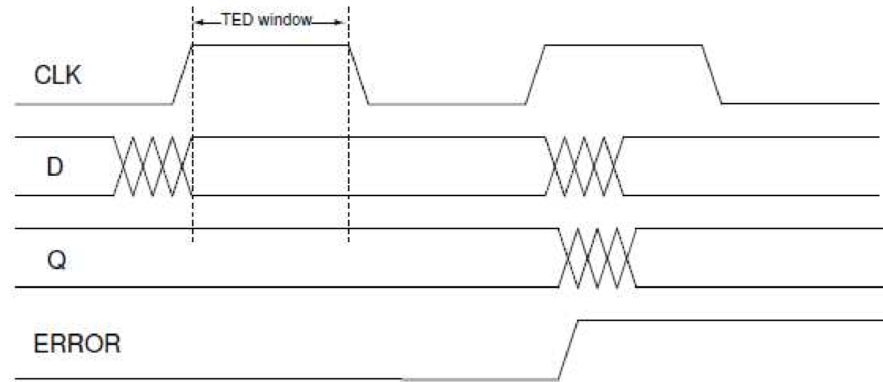

Figure2 TED timing diagram [1]

1.1.1. Technique 1: TFD

TED was first introduced in [10] and consists of two latches and a comparator. If data (*out*) transitions occur during the *CLK* period, an error signal (*err*) is obtained. Two latches (LATCH, Extra LATCH) have delayed clock signals *CLK* and *CLK+X*, they successfully latch the data (*out*) to produce latch outputs Q1 and Q2. A comparator detects the difference in Q1 and Q2 which are also delayed by X and thereby generating an error signal (*err*) (Fig.3). It [10] proposes the most economical solution of combining TED technique with an architectural replay procedure that places the data again through the combinational logic every time an error signal (*err*) is generated.

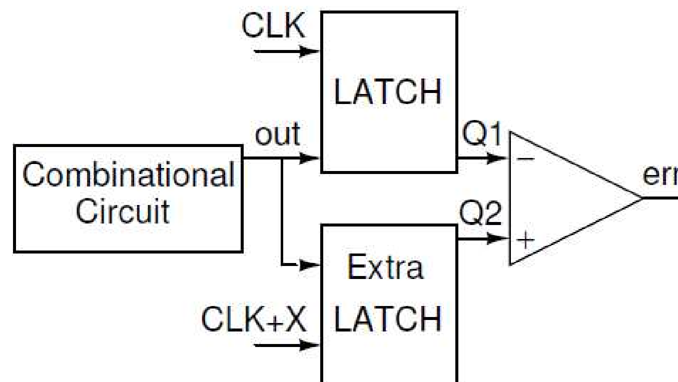

Figure 3 TFD [10]

1.1.2. Technique 2: Razor

The architecture of Razor II is shown in Fig. 4. It uses a single positive level-sensitive latch, with the addition of a transition-detector (TD) controlled by a detection clock (DC). Data is considered on time when D transitions prior to the rising edge of the CLK. However, if D transitions after the rising edge of CLK, during transparency, then the transition of latch node N occurs and when TD is enabled and an ERROR is generated. At the rising edge, DC provides a short pulse to disable TD for at least a delay of the clock-to-Q delay of the latch.

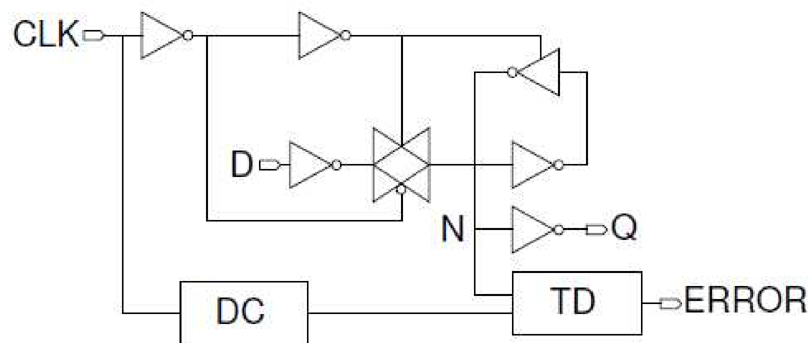

Figure 4 Razor II [11]

1.1.3. Technique 2: Time Borrowing Transition Detector (TDTB)

The TDTB monitors the location of input data *D* transitions with respect to the *CLK*. For each *D* transition, a pulse is generated at the output of the XOR gate. When *CLK* is low, node K is driven high by P1 thus keeping *ERROR* low. The pulse from the XOR does not affect the *ERROR* signal during *CLK* low. A late arriving *D*, during the time when CLK is logic high, discharges node K since both N1 and N2 are ON. This enables the ERROR node to pull up to logic high. The transition detector has metastability issues. Metastability may occur if D arrives close to a CLK edge, which is the boundary of a timing failure. The LATCH is transparent during the CLK high and thus *D* still propagates to the next stage. The error buffer will be driven high or low during metastability, therefore, either case still maintains correct functionality [7].

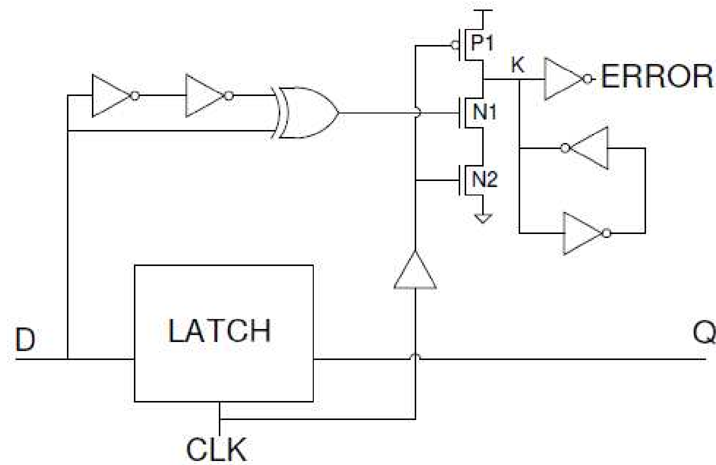

Figure 5 TDTB [7]

TDTB when compared with other designs was found to have least energy overhead. This concept was extended in [1] to achieve subthreshold operation and was known as *TDTBsub.*

1.1.4. Technique 2: TDTBsub

The pulse generator consisting of inverters and XOR gate produce a narrow pulse for every data transition. The pull down network controlled by CLKd allows node k to transition low if D transitions at the time CLK is high. The Keeper is used to prevent a floating node and it switches states if node K is driven low through the pull-down network. The output is then held till the next clock by the latch2. (Fig.6)

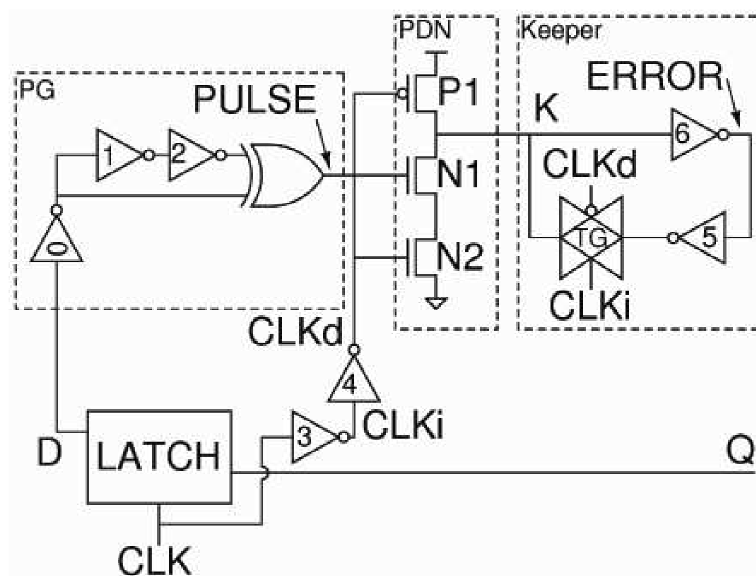

Figure 6 TDTBsub [1]

The design of TDTBsub in [1] requires a lot of design time. The sizing of inverters 8 and 9 of keeper are crucial and the sizing the circuit for robustness eventually leads to a higher power consumption. Also the delay inverters in the pulse generator need to be carefully sized. They should be capable of producing narrow and detectable pulses. Also the transition to delay detection is important. These constraints were carefully tackled in the design of DITD.

2 Proposed Design

DITD has been carefully designed by studying the existing TED latches: TDTBsub, TDTB and Razor. The working principles and the sizing of transistors which are crucial for subthreshold operation are discussed in this section.

2.1 Operation Principles

DITD is new type of TED latch, its design as shown in Fig.7 consists of a Transition detector similar to [1]. The inverters in the transition detector are the most intricate. The number of inverters in the inverter chain determines the power consumption. Also the output of the XNOR gate depends on delay produced by them. Too less propagation delay (t_p) between the input (D) and $\overline{D}$ and the transition will go undetected. A high propagation delay ensures the transition detection but increases power consumption. A detailed discussion about the inverter chain is carried out in section 2.3.

It is noted that the previous designs TDTB and TDTBsub use an XOR gate in the transition detector but the DITD uses an XNOR gate. However, the same functionality is achieved by using an inverter before the transition detector which is seen in [1].

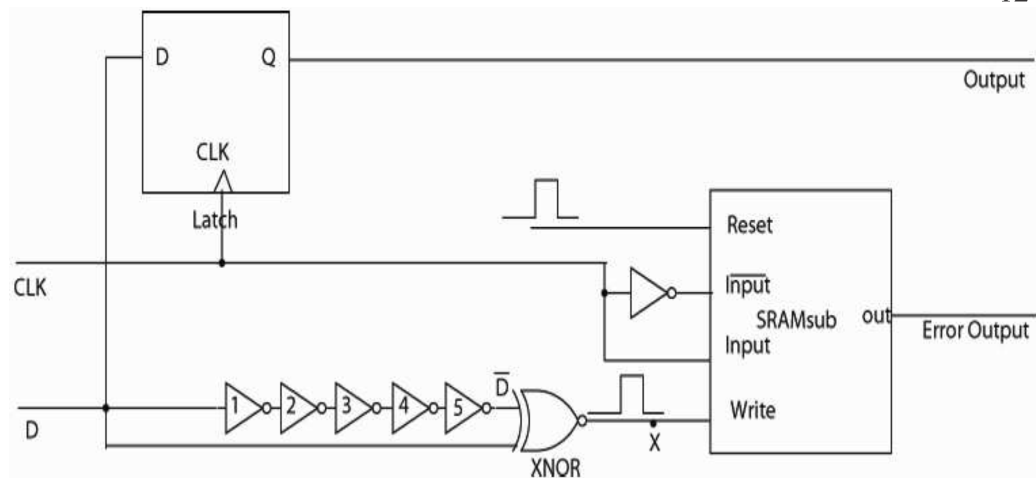

Figure 7 Design of DITD.

When there is a transition in the data input D, the inverted version $\overline{D}$, due to finite propagation delay (t_p) of the inverters 1-5, arrives at the input of the XNOR gate later. The XNOR gate therefore generates a narrow pulse at node X. This narrow pulse is then fed to the write input of a SRAMsub. The $input$ and $\overline{input}$ of the SRAMsub are connected to the clock signal CLK and $\overline{CLK}$ respectively. Thus when there is a data (D) transition and when the clock is high simultaneously, a '1' is written in the SRAMsub. If no data transitions occur in the CLK high the output of SRAMsub is held in the previous state.

Once the error is detected, a reset pulse *RESET* is used to reset the SRAMsub. When a *RESET* pulse is given to the SRAMsub, a '0' is written. The *RESET* pulse must be generated from the processor once the error detection is done. However, the SRAM has been designed such that it can RESET automatically if an input transition occurs in clock low, otherwise the RESET pulse must be generated.

It must be noted that the DITD uses only one latch whereas the designs TDTB [7] and TDTBsub [1] use an additional latch (2 latches in total) to hold the error high in CLK low.

2.2 Sizing of SRAMsub

A Static RAM has been used for keeping the error high once a transition of the input is detected. The use of SRAM is a radical new approach in DITD. The output is held high once an error is detected without much complexity as compared to RazorII or TDTBsub. Also, the lowered power consumption is achieved due to this reduction in complexity.

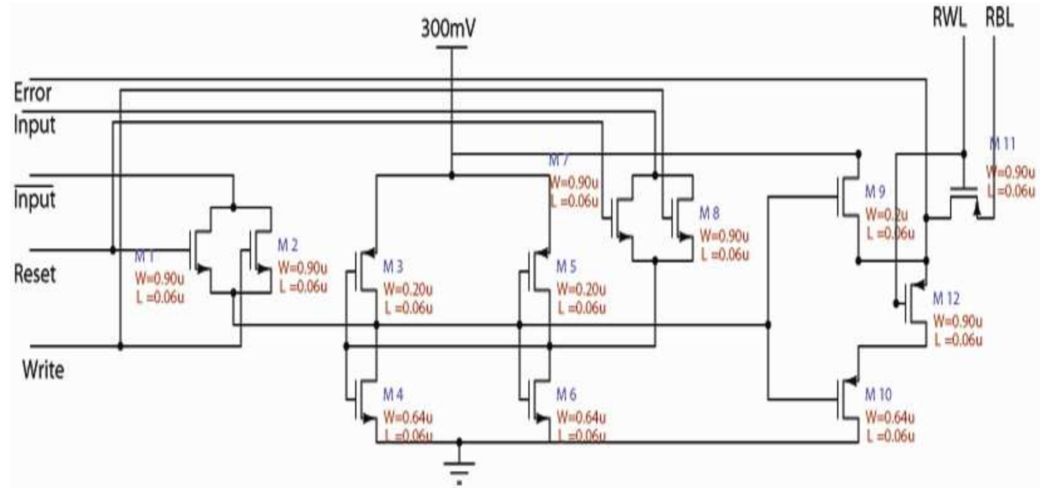

Figure 8 SRAMsub circuit.

The sizing and leakage become very important in subthreshold. The SRAM used here is a modification of the SRAM in [2] that uses 10 transistor style. The RESET functionality has been added to the SRAM found in [2]. The leakage reduction is achieved by using stacked devices. Transistors M11 and M12 constitute a buffer are therefore needed for operation in sub threshold in addition to a normal 6T SRAM design.

The transistors in SRAMsub are sized to ensure robustness in 100 Monte Carlo simulation runs. The lengths of all the transistors were sized at minimum for a 65nm technology (i.e. at 0.06μm). It was found that sizing the NMOS and the PMOS transistors to the minimum widths yielded very poor results in Monte Carlo simulations. So they were all sized bigger. It was noted from [2] that the variation is given by

$$\sigma(\Delta P) = \frac{A_p}{\sqrt{WL}} \tag{1}$$

From equation (1) it is clear that variation is reduced by using wider and or longer transistors.

Robustness in SRAMsub is achieved by sizing the transistors bigger with the tradeoff being high energy. Initially a dynamic RAM (DRAM) that uses a lesser number of transistors and that is therefore lesser in power consumption was suggested. Later on further investigation revealed that the design needs to employ only static logic due to higher leakages during subthreshold operation.

The width of PMOS M3, M5, M9 were swept from a minimum value of 0.12µm to 0.84µm in steps of 0.12µm. 50 MC were run at 0.3V for each sweep value. It was found that these can be minimum sized, but were sized higher at 0.20µm as the NMOS needed to be sized higher than the minimum sizes so as to have symmetric high to low and low to high propagation delays. The symmetric propagation delays are not vital here.

The widths of NMOS M1, M2, M4, M6, M7, M8, M10, M11, M12 was also swept from 0.12µm to 2.68µm in steps of 0.12µm. 50 Monte Carlo at 0.3V were also run at each sweeps. It was found that for proper functionality these NMOS transistors need to be sized greater than 0.60µm. In [2] the suggested sizing is 5.33 times the minimum widths for robustness. So these were sized at 0.64µm.

After these sizings were obtained a 100 Monte Carlo simulation was run finally and the number of failures was 2, an optimized and a robust design of the SRAMsub was obtained.

2.3 Delay Characterization in an inverter chain

The delay must be precise and is very crucial in the operation of the DITD. Methods to increase delay with lesser power consumption were investigated and relation between delay, power consumption, number of stages and sizing needed was studied. The minimum delay through an inverter chain is given by [3]

$$t_p = t_{p0}.N(1 + \frac{\sqrt[N]{F}}{\gamma}) \tag{2}$$

In (2) N is the no. of stages, F the fan out factor, t_{p0} the unloaded delay of an inverter and γ the proportionality factor of technology. The delay depends on γ is inversely, but the dependence on N is unclear. There exists an optimum no. of stages such that the delay is minimum [3]. However there is no standard equation or method in literature to *increase* the delay in an inverter chain as it is always required to be minimum.

2.3.1 Sizing and Stages

The number of stages of inverters was varied from 1 to 11. It was found that the propagation delay (t_p) due to 1 or 3 stages was too small and hence the detection of data transition was impossible, more stages were needed. So the stages 5, 7, 9 and 11 were studied in detail. The delay as a function of sizing was studied for each stage. The power consumption for each sizing and stage was simultaneously measured.

2.3.1.1 Delay and Width

There are many ways to obtain the optimum sizings within an inverter, as the widths of PMOS can be swept or the NMOS can be swept. The method employed in this section was to sweep widths of both PMOS and NMOS such that the ratio of $\frac{\left(\frac{W}{L}\right)_p}{\left(\frac{W}{L}\right)_n}$ was always a constant and was equal to 1.4. It is important that the ratio is maintained as the symmetric propagation delays are required. Increasing the widths of PMOS or NMOS arbitrarily

increases the delay but the t_{pHL} *and* t_{pLH} increase differently. Even [1] states a ratio of 1.5 where the widths of PMOS and NMOS have been swept separately. The lengths of the PMOS and NMOS were at a minimum of 0.6μm. The width of the PMOS was $\boldsymbol{W_n = 0.28 * k}$ and the width of NMOS was $\boldsymbol{W_n = 0.2 * k}$. The widths of the minimum sized inverter x2 in 65nm library are 0.28μm and 0.2μm. The value of sizing factor 'k' was then swept from 1 to 5 in steps of 0.1. The delay is shown in Fig.9 and the power consumption is shown in Fig.10.

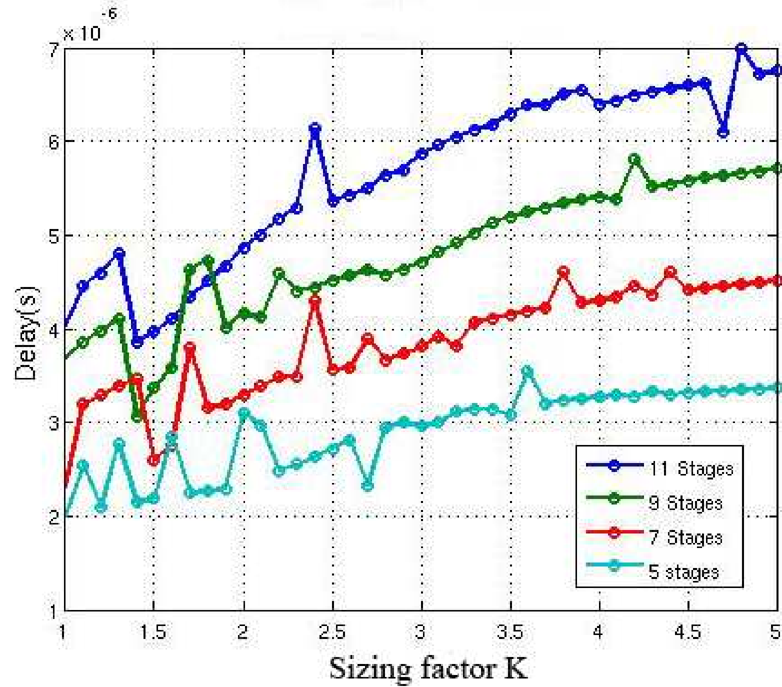

Figure 9 A plot of delay t_p for various sizing and number of stages.

The delay was found to decrease initially but then increases gradually as the value of k is increased. This is due to the fact that increasing the size of the devices increases the capacitance. The propagation delay t_p is obtained in [3] as (3)

$$t_p = \int_{v_1}^{v_2} \frac{C_L(v)}{i(v)} dv \tag{3}$$

The delay depends directly on the capacitance and inversely on the saturation current. The saturation current is given as (4)

$$I_{DSAT} = \frac{k'W}{L}\left((V_{DD} - V_T) - \frac{V_{DSAT}^2}{2}\right)$$

(4)

Equation (4) implies that the current depends on the width W. Increasing width increases current therefore the delay is reduced as seen from (3). However increasing W also increases the load capacitance C_L as all the capacitances- overlap capacitance, bottom and sidewall capacitance and gate capacitance are scaled proportional to W. As the capacitance is increased the inverter becomes self-loaded and the delay is increased drastically. So the simulation agrees to theoretical predictions.

2.3.1.2 Power and Width

It was found from simulations that power increases almost linearly as expected. The bigger the capacitance the bigger is the switching energy. If the widths of the transistors are increased, higher delay can be achieved only with higher power consumption seen from Fig.10.

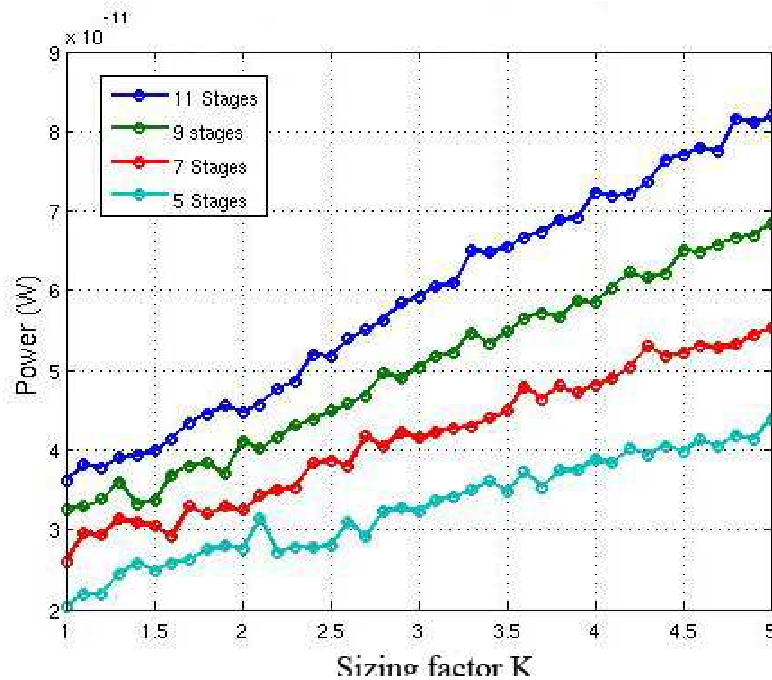

Figure 10 Power consumption of delay chains for various widths and number of stages.

2.3.1.3 Delay and Length

The same method used to sweep the width in section 2.3.1.1 was employed in this section. The lengths were swept with widths of both PMOS and NMOS kept constant at 0.28um and 0.2um. The lengths were $\mathrm{L_p} = 0.06 * \mathrm{k}$ and $\mathrm{L_n} = 0.06 * \mathrm{k}$. 'k'

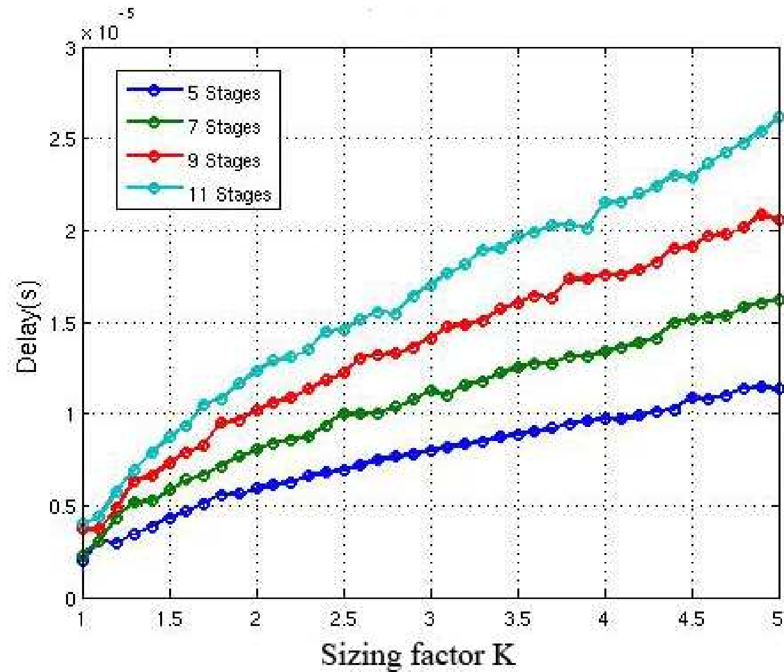

Fig. 11. A plot of delay t_p for various lengths and number of stages

was swept from 1 to 5 as previously. Note again the ratio $\frac{\left(\frac{W}{L}\right)_p}{\left(\frac{W}{L}\right)_n}$ was always a constant and was equal to 1.4. Equation (4) suggests that the current reduces as L is increased. The delay t_p (3) therefore increases rapidly as its inversely proportional. The capacitance C_L is also scaled when length of the transistors is increased as the gate capacitance increases (seen from Fig. 11).

2.3.1.4 Power and length

As the lengths of the transistors are increased the current is decreased but as the capacitance increases the power consumption is increased ultimately as seen from fig.12.

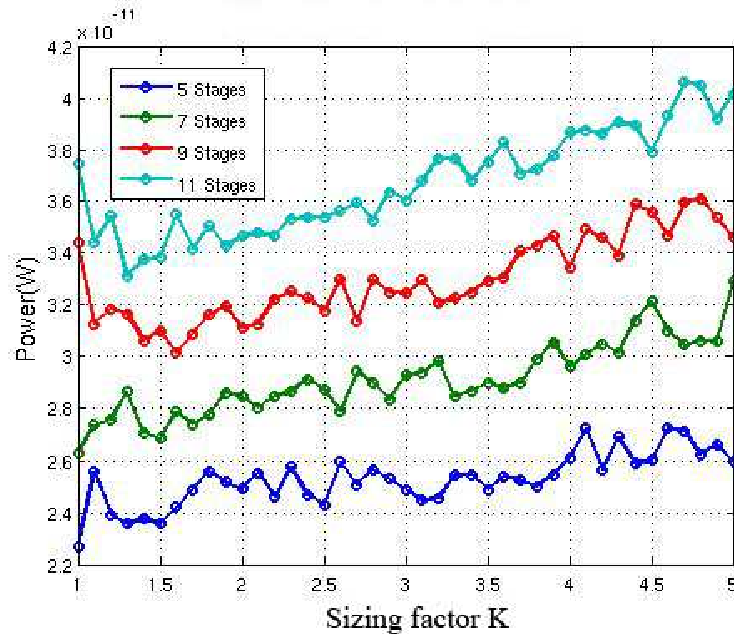

Fig. 12. Power consumption of delay chains for various lengths and number of stages.

It was observed that the delay is more sensitive to changes in length. It increased more rapidly for increase in lengths than to width. This was observed by plotting the delay for different width and lengths simultaneously for 11 stage inverter chain (fig.12). Also for this chain it was observed that sizing the widths bigger leads to more power consumption than sizing the lengths bigger (fig.13).

It can be concluded clearly from the simulations that in order to obtain maximum delay and minimum power consumption one needs to choose least number of possible stages. The transistors must be of minimum widths but of longer lengths.

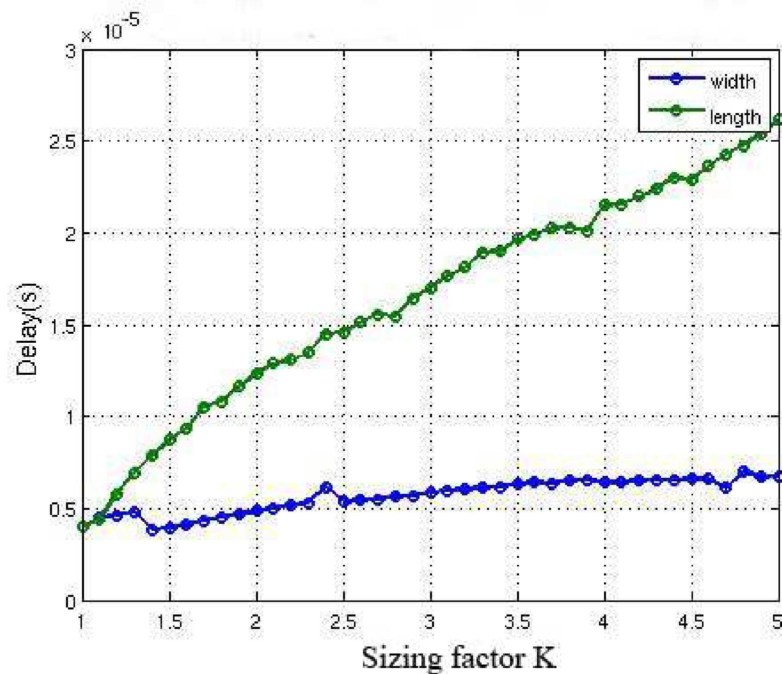

Fig. 12. A plot of delay t_p for various lengths and

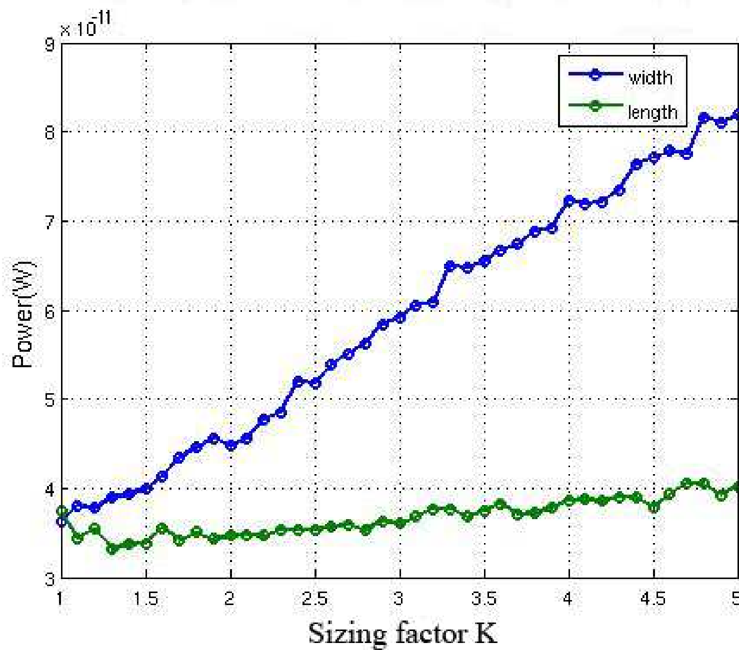

Fig. 13. A plot of delay t_p for various lengths and number of stages.

3. Simulation Results and Comparisons

3.1.Operation

The DITD was simulated and the output waveforms have been exported from Ezwave and plotted using MATLAB (Fig.7). The clock period used were 130F04 which at 0.3V=83.5µs. The

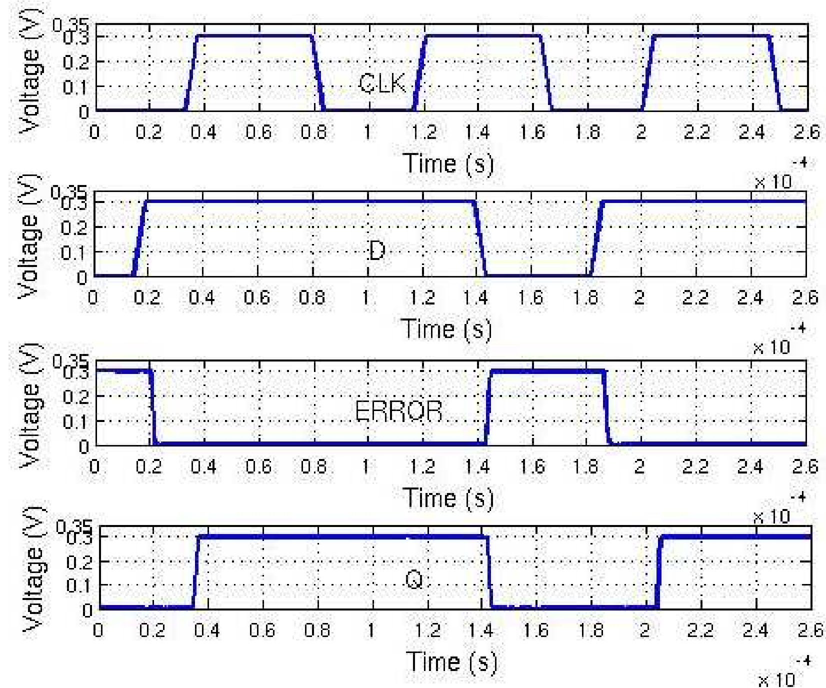

Fig.14 A plot of Error Detection. TCLK= 130FO4.

DITD was working as predicted. Whenever there was transition under CLK high the ERROR was high and is held till the next CLK. The latch input D propagates to output Q under CLK high and is held to the previous value sampled under the falling CLK.

3.2. Uncertainty Region

There exists an interval before the rising and falling edge of the clock during which the error signal goes high unexpectedly. This region is called the uncertainty region. This is concept is illustrated in Fig. 9. Care should be taken to make sure the data transitions do not occur here. The smaller the uncertainty region the better the design.

In order to measure this region the input D was skewed relative to the clock edge in all corners TT, FF and SS. The skewing step in every corner was always 1 FO4. Standard FO4 values were used for 0.3V (table 5). 100 MC simulations were run at each skew position.

It was found that for a time period t_{a2} before the rising CLK the percentage of false errors generated was 0. If the input transitions fall in the interval $t_{a2} < D < t_{edge2}$ the percentage of false errors increases from 0% and finally it becomes 100% for an interval t_{edge2}. The setup time for the latch is 0 so for the duration t_{edge2} before the CLK the error is high falsely.

There always exists local variations between the different TED latches which may be present in various critical paths of a processor. So there can be errors generated falsely due to the data transitions in several of them. So this is a key system level constraint in the design of TED latches.

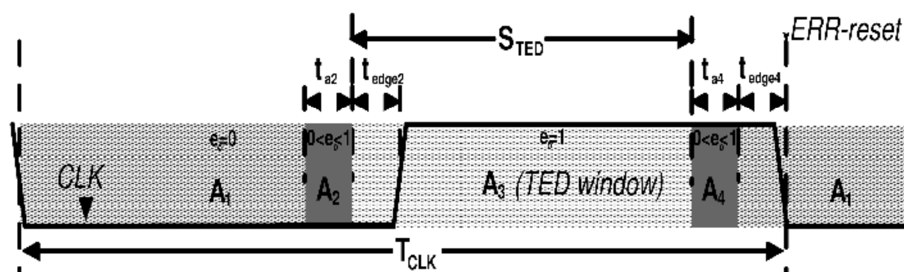

Figure 9. A plot of Delay for various sizing and number of stages

The above discussion applies to the falling edge. The values of t_{a2}, t_{a4}, t_{edge2}, t_{edge4} are tabulated in table 1. The uncertainty regions of TDTBsub are also presented in table2. It is seen that the value of the intervals agree closely. The TED window S_{TED} is also equal.

TABLE I
UNCERTAINTY REGION FOR DITB 0.3V

Corner	FO4 Delay (μs)	Duty Cycle	T_{MIN} (FO4)	f_{MAX} (kHz)	t_{a2} (FO4)	t_{a4} (FO4)	tedge2 (FO4)	tedge4 (FO4)	$t_{r,f}$ (μs)	S_{TED} (FO4)
TT	**0.643**	**50%**	**130**	**11.96**	**12**	**14**	**3**	**6**	**4.17**	**48**
FF	**0.23**	**50%**	**130**	**33.44**	**11**	**13**	**3**	**5**	**1.495**	**50**
SS	**1.72**	**50%**	**130**	**4.472**	**15**	**16**	**2**	**7**	**11.18**	**44**

Table1 Uncertainty region for DITD

TABLE II
UNCERTAINTY REGION FOR TEDSC 0.3V

Corner	FO4 Delay (μs)	Duty Cycle	T_{MIN} (FO4)	f_{MAX} (kHz)	t_{a2} (FO4)	t_{a4} (FO4)	tedge2 (FO4)	tedge4 (FO4)	$t_{r,f}$ (ns)	S_{TED} (FO4)
TT	**0.643**	**50%**	**130**	**12**	**15**	**15**	**4.5**	**4.5**	**252**	**50**
FF	**0.23**	**50%**	**130**	**31.6**	**20**	**20**	**10**	**10**	**86**	**45**
SS	**157-268**	**50%**	**130**	**4.5**	**13**	**13**	**3**	**3**	**656**	**52**

Table2 Uncertainty region of TEDsub

3.3.Power Consumption

The DITD was simulated using several inverter chain elements. Initially it was simulated using a chain of 11 inverters where the widths of PMOS and NMOS transistors were sized at 5 times that of an X2 inverter. It was seen that a chain of 11 inverters were required when the widths were sized bigger or else there were many failures in 100 MC simulations.

Design	Description	Sizing
Initial Design	Chain of 11 Inverters with l=0.06um $W_n = 1um\ , W_p = 1.5um$	S=5(5*X2) S=$\frac{\left(\frac{W}{L}\right)_p}{\left(\frac{W}{L}\right)_n} = 1.5$
Modified Design I	Chain of 5 inverters with l=0.18um $W_n = 1um\ , W_p = 1.5um$	S=5(5*X2) S=$\frac{\left(\frac{W}{L}\right)_p}{\left(\frac{W}{L}\right)_n} = 1.5$
Modified Design II	Inverter chain from [1]	-

Table 3 Description of various simulated designs

Parameters	Initial Design	Modified Design I	Modified Design II	TEDsc Measured
Current	223.12pA	162.35pA	524.08pA	300pA
VDD	300mV	300mV	300mV	300mV
Power	**66.94pW**	**48.71pW**	**157pW**	**210pW**
Monte Carlo (All corners)	2 failures in 50 MC SS 1 failure in 50 MC	0 failures for 100 MC in FF 2 failures for 100 MC in SS	0 failures in 100 MC SS 0 failure in 100 MC	0 failures in 100 MC SS 0 failure in 100 MC

Table 4 Power Comparisons

Note: MC simulations were done at 130FO4. To provide a direct comparison, the power was calculated at 150FO4 with the pad ring driver in place ($W_n = 12um\ , W_p = 7.2um, L_p =$

$0.08um, L_n = 0.08um$) as measurements of TEDsc were made for 150FO4.

Corner	Clock=130 FO4	Frequency
FF	29.9x e-6	33444.8
SS	223.6x e-6	4472.27
TT	83.5x e-6	11976.0

Table5 Clock values for various corners at 300mv.

A modified design where the lengths of the PMOS and NMOS transistors were sized bigger at 0.18um was also simulated. As indicated in section 2.3.1 the power consumption was less and this design was found to provide enough delay. The robustness was also better than the initial design as there were lesser failures in MC simulations. This design is the least power consuming of all.

A second modified design with the delay chain of TEDsc [1] was simulated to provide an even more straight forward comparison. It was found that this circuit was very robust with no failures but was consuming more power.

4. Conclusion

This thesis gave a brief introduction to a TED system. Previous TED implementations including the previous subthreshold design were then provided. Based on these the new design of DITD was developed and then simulated.

Simulations proved that the goal of designing a new subthreshold TED was successfully accomplished. This work also discussed about the subthreshold delay and its power requirements. A clear sizing for subthreshold was obtained. The innovative idea to use of SRAM to keep the error high proved effective and eliminated the need for another latch. As a consequence additional inverter requirements from the previous designs were tackled removed. The uncertainty region, which is a very important system level constraint was also discussed. However, now the DITD design at the system level requires a new reset pulse that must be generated after the error, in order to RESET the ERROR signal.

Monte Carlo Simulations were run at all corners at 300mV and all the designs presented proved to operate robustly under each corner.

5. Future Work

The delay of the inverter chain is a key design parameter as stressed many times and must therefore be analyzed theoretically. Achieving higher delays with lower power consumption is the most compelling task and is very hard to achieve in subthreshold using static CMOS. A small theoretical approach is presented here which can be extended for future work to obtain a truly optimized design.

5.1 Theoretical Delay

A delay Chain with N inverters each having a Sizing of S_1, $S_2 \ldots S_N$ was considered. The total capacitance at each node is represented as C_1, $C_2 \ldots C_N$.

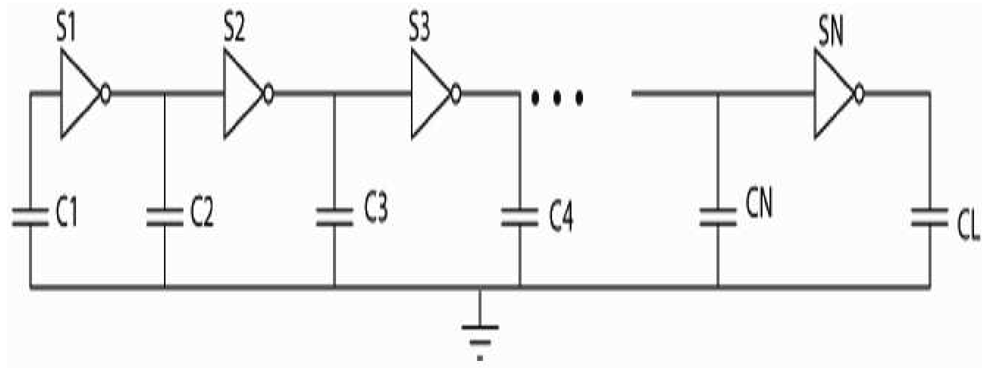

Fig.10 An inverter chain

The capacitance at every node consists of the intrinsic capacitance C_{int} and the extrinsic or load capacitance C_{ext}. Every inverter in the chain sees the gate capacitance of the next stage as the extrinsic capacitance.

$$C_{in}(g, N) = S_N . C_g \tag{5}$$

Where C_g is the input gate capacitance of a minimum sized inverter (X2).

The intrinsic capacitance due to self-loading is given as

$$C_{int}(g, N) = \gamma . C_{in}(g, N) = \gamma . S_N . C_g \tag{6}$$

Therefore the propagation delay through the inverter is given as

$$t_p = 0.693 . R . (C_{int} + C_{ext}) \tag{7}$$

$$t_{p1} = 0.693.R.(S_1C_g + S_2C_g) \tag{8}$$

$$t_{p2} = 0.693.R.((S_1C_g + S_2C_g) + S_3C_g) \tag{9}$$

$$t_{p3} = 0.693.R.((S_1C_g + S_2C_g + S_3C_g) + S_4C_g) \tag{10}$$

$$t_{pN} = 0.693.R.((S_1C_g + S_2C_g + S_3C_g + \cdots + S_NC_g) + C_L) \tag{11}$$

Therefore the total delay is the sum of the delay through each inverter and is given by

$$T_p = t_{po}.(N.S_1 + N.S_2 + (N-1).S_3 + \cdots + 2.S_N + \frac{C_L}{C_g}) \tag{12}$$

Now in order to maximize the equation (8) we can choose sizing as

$$S_1 = S_2 = \left(1 - \frac{1}{N}\right)S_3 = \left(1 - \frac{2}{N}\right)S_4 = \cdots = \frac{2}{N}S_N \tag{13}$$

$$\text{Yielding } T_{pmax} = t_{p0}(N^2S_1 + \frac{C_L}{C_g}) \tag{14}$$

[12] Gives the expression for the inverter delay chain as

$$TIP = \frac{2V_{DD}C_1}{\frac{C_g\mu_nW_n}{L_n}(V_{DD} - V_{tn})^2} + \frac{2V_{DD}C_2}{\frac{C_g\mu_pW_p}{L_p}(V_{DD} - V_{tn})^2} \tag{15}$$

As observed in section 2.3.1 the delay is inversely proportional to the widths of the transistors and directly proportional to the lengths and ratio of output to input load capacitances.

Equations (14) and (15) need to be analyzed carefully and this kind of holistic approach will lead to the most optimized design.

6. References

[1] Matthew J. Turnquist, "*Sub-threshold Operation of a Timing Error Detection Latch*".

[2] Alice Wang, Benton Highsmith Calhoun, Anantha P. Chandrakasan, "*Sub Threshold Design for ultra low power systems"*.

[3] Digital Integrated Circuits A design Prespective Second Edition, Jan M. Raeby, Anantha Chandrakasan, Borivoje Nikolic.

[4] Y. Ramadass and A. Chandrakasan, "*Minimum energy tracking loop with embedded dc-dc converter delivering voltages down to 250mv in 65nm cmos*",in Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2007, A. Chandrakasan, Ed., 2008, pp. 64.587.

[5] M. Najibi, M. Salehi, A. Kusha, M. Pedram, S. Fakhraie, and H. Pedram, "*Dynamic voltage and frequency management based on variable update intervals for frequency setting,*" in Proc. IEEE/ACM International Conference on Computer-Aided Design ICCAD '06, M. Salehi, Ed., 2006, pp. 755.760.

[6] J. Tschanz, N. S. Kim, ET. AL., "*Adaptive frequency and biasing techniques for tolerance to dynamic temperature-voltage variations and aging"*, in Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2007, N. S. Kim, Ed., 2007, pp. 292.604.

[7] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Wilkerson, S.-L. L. Lu, T. Karnik, and V. K. De, "*Energy-efficient and metastability-immune timing-error detection and instruction-replay-based recovery circuits for dynamic-variation tolerance*", in Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2008, 2008, pp. 402.623.

[8] S. Das, C. Tokunaga, S. Pant, W. H. Ma, S. Kalaiselvan, K. Lai, D. M. Bull, and D. T. Blaauw, "*RazorII: In situ error detection and correction for pvt and ser tolerance*", IEEE J. Solid-State Circuits, vol. 44, no. 1, pp. 32.48, 2009.

[9] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Wilkerson, S.-L. L. Lu, T. Karnik, and V. K. De, "*Energy-ef_cient and metastability-immune timing-error detection and instruction-replay-based recovery circuits for dynamic-variation tolerance*", in Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2008, 2008, pp. 402.623.

[10] L. Anghel and M. Nicolaidis, "*Cost reduction and evaluation of a temporary faults detecting technique*", in Proc. Design Automation and Test in Europe Conference and Exhibition 2000, 2000, pp. 591.598.

[11] D. Blaauw, "*Razor ii: In situ error detection and correction for pvt and ser tolerance*", ISSCC 2008, 2008.

[12] B. R. Blaes and M. G. Buehler, "*Inverter Propagation Delay Measurements using Timing Sampler Circuits*", in Proc. Int. Conference on Microelectronic Test Structures,Vol 2, No. 1, March 1989.