11.7b Time-to-Digital Converter with 0.82ps resolution in 130nm CMOS Technology

Rodrigo Filipe Fernandes Granja

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisors: Prof. Nuno Cavaco Gomes Horta

Prof.Jorge Manuel Correia Guilherme

Examination Committee

Chairperson: Prof. Gonçalo Nuno Gomes Tavares

Supervisor: Prof. Jorge Manuel Correia Guilherme

Member of the Committee: Prof. Jorge Manuel dos Santos Ribeiro Fernandes

April 2018

Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

All the resources used in the development of this dissertation were provided by the Instituto de Telecomunicações.

ABSTRACT

This dissertation belongs to the scientific area of Integrated Circuits design and addresses the development of a 11.7-bit architecture Time to Digital Converter using a" 130nm UMC CMOS technology. The Time Converter will be developed with a Circular architecture using a pulse-shrinking cell-based measurement technique. Compared with conventional Pulse-Shrinking techniques, where the buffer is used to implement the delay element in the chain, here it is proposed the use of an inverter for this purpose. The sampling circuit will be based on TSPC dynamic flip flops, specially modified to fit this circuit, reducing the load imposed on each of elements in the delay chain enabling a better performance of the same. To perform the cycle counter, a 5-bit Asynchronous Counter or Ripple Counter has implemented, having a maximum operating frequency of 4.94 GHz. Finally, in order to convert the "Thermometer Code", returned by the time converter, in a binary word, a decoding based on the 7-bit Fat-Tree concept was proposed, using NOR and NAND ports. The TDC circuit will contain a 128-element chain, reaching a time resolution of 0.82ps with a dynamic range of 2.92ns. The DNL will vary between -0.98 and 0.97 LSB and the INL between -1.52 and 1.27 LSB. The Circuit will perform a consumption of 7.5mW, an area of 0.148mm2, and a FOM of 91.2 fJ/conv.

KEY-WORDS

- Time-To-Digital Converter
- Inverter based Pulse-Shrinking Technique
- High Resolution
- High Dynamic Range
- Low area Consumption
- Low Power Consumption

RESUMO

Esta dissertação pertence à área científica de projeto de Circuitos Integrados e aborda o desenvolvimento de um Conversor de Tempo para Digital com uma arquitetura de 11.7 bits usando uma tecnologia "CMOS UMC de 130nm". O Conversor de Tempo-para-Digital será desenvolvido com uma arquitetura Circular usando uma técnica de medição baseada em inversores de Pulse-shrinking, que em comparação com as técnicas convencionais, onde o buffer é usado para implementar o elemento de atraso na cadeia, aqui é proposto o uso de um inversor para o mesmo propósito. O circuito de amostragem será baseado em flip-flops dinâmicos TSPC, especialmente modificados para este circuito, reduzindo a carga imposta em cada um dos elementos na cadeia de atraso, permitindo um melhor desempenho do mesmo. Para executar o contador de ciclos, um modelo assíncrono, ou ripple-counter de 5 bits foi implementado, tendo uma frequência de operação máxima de 4.94 GHz. Finalmente, para converter o "Código-Termômetro", devolvido pelo conversor de tempo, em uma palavra binária, foi utilizado um conversor de 7 bits baseado no conceito de Fat-Tree, usando portas NOR e NAND. O TDC terá uma cadeia de 128 elementos, atingindo uma resolução de tempo de 0,82ps com uma gama dinâmica de 2.92ns. A DNL variará entre -0.98 e 0.97 LSB e a INL entre -1.52 e 1.27 LSB. O Circuito realizará um consumo de 7.5mW, uma área de 0.148mm2 e um FOM de 91.2 fJ/conv.

PALAVRAS-CHAVE

- Conversor de Tempo para Digital
- Técnica de Pulse-Shrinking com base em Inversores
- Alta resolução
- Alta Gama Dinâmica
- Baixo consumo de energia
- Baixo consumo de área

ACKNOWLEDGEMENTS

To all Non-Portuguese's that are reading this work I want to apologise, but I would like to write all my acknowledgments in my native language.

Com a Finalização desta dissertação concluo para já, não sabendo o que o futuro académico me reserva, o meu ciclo de Estudos, a minha jornada na Instituto Superior Técnico. e com isto o meu Mestrado em Engenharia Eletrotécnica e de Computadores. Foram árduos longos anos de trabalho e dedicação, que sem o apoio dos que me são mais próximos seriam impensáveis.

Em primeiro lugar gostaria de prestar um especial agradecimento à minha família, a minha Mãe Helena Granja, ao meu Pai Diamantino Granja, e ao meu Irmão Diamantino José Granja, por toda a incrível dedicação, esforço e apoio que me deram, sem vocês nunca teria conseguido chegar onde cheguei e ultrapassar todos os obstáculos que se foram apresentando a minha frente. Um grande, grande Obrigado!!

Quero também prestar um especial agradecimento aos meus orientadores Professor Doutor Jorge Guilherme e ao Professor Doutor Nuno Horta, pela oportunidade que me deram em realizar a minha tese numa das minhas áreas de maior interesse, por todo o apoio, motivação, orientação e porque tudo aquilo que me ensinaram ao longo ao longo de toda a realização desta dissertação.

Ao Engenheiro Mauro Santos queria agradecer, por todo o apoio, compreensão, paciência, ajuda, por todas as ferramentas que me mostrou e por tudo aquilo que me deu a conhecer e ensinou. Para mim um terceiro orientador na realização desta dissertação.

A todos os meus AMIGOS que mesmo estando longe, sempre me deram total e incondicional apoio, dedicação, ajudando-me a levantar e a ver a luz a fim do túnel mesmo em alturas mais complicadas e que demostraram que eu conseguia cada vez fazer mais e melhor, o mínimo que lhe posso dar é um grande obrigado!! Gostaria também de dar um especial agradecimento a D.Ruth Pearman, a minha amiga Catarina Nogueira pela ajuda que me deram e os erros que corrigiram na escrita desta dissertação.

Por fim gostava de expressar a minha gratidão ao Instituto de Telecomunicações, e ao Instituto Politécnico de Tomar, por todo o suporte e material que me disponibilizaram durante todo este percurso.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Δt_{max}	Maximum Time-measured
ADC	Analog-to-Digital Converter
C_L	Load Capacitance
CMOS	Complementary Metal-Oxide-Semiconductor
DLL	Delay-Locked-Loop
DNL	Differential non-linearity
ELD	Early-Late-Detectors
ENOB	Effective Number of Bits
FF	Fast – Fast Corner
FOM	Figure of Merit
f_s	Sampling Frequency
INL	Integral Non-Linearity
METDC	Multi-Event TDC
PLL	Phase-Locked-Loop
PVT	Process, Voltage, Temperature
SS	Slow – Slow Corner
TC-to-BC	Thermometer Code to Binary Code
TDC	Time-to-Digital Converter
T_{LSB}	Time Resolution
T_{pd}	Propagation Delay-time
T_{PHL}	Time for a High-to-Low Transition
T_{PLH}	Time for a Low-to-High Transition
TSPC	True Single-Phase Clock
VTH	Threshold Voltage

1. INTRODUCTION

Physical properties such as temperature, length, mass or time are behind the control of all known systems, and are usually processed through analog circuits, using electrical physical quantities like voltage, current, charge or frequency. In the analog literature, the most common signal processing techniques are voltage-mode followed by current-mode and charge-mode type, which translate their signals into voltage, current and charge respectively [1].

Technological advances in CMOS is always geared by the optimization of digital circuits, therefore analog approaches are continuously losing their advantages and benefits. Technology scaling results in the decrease of the voltage power supply. However, this decrease is not accompanied at the same rate by the threshold- voltage, V_{TH}, therefore the resulting voltage headroom will decrease. Hence, voltage-mode circuits, defined by the ratio of the minimum detectable voltage (typically set at the noise floor) and the maximum available headroom become deeply affected, scaling poorly with technology [2,3].

Current-mode circuits offer an alternative, representing the “data” by the branch current of electrical networks. These circuits have the ability to work with low nodal voltages, lowering the impedance nodes, offering also a low time constant at every circuit node. However, with lower nodes impedance gives rise to large branch currents, which results in higher power consumption when compared with the voltage-mode circuits. Therefore, current-mode circuits are more suitable for applications when high-speed is more critical than power consumption [1,3].

Since voltage and current are inherently related to each other via impedance or conductance, it is easy to conclude that, the characteristics of both types of circuits will not scale well with technology in their performance. There are some methods to compensate some of these issues, based on digital tuned systems, however this approach is costly in terms of silicon and power consumption [3].

Time-mode circuits use the time difference between two events to quantify a determined quantity. The amount of time difference can be linearly proportional to the amplitude of the analog signal that needs to the digitalized.

A time variable possesses a unique duality characteristic. It is an analog variable, with a continuous amplitude represented by the pulse duration and it is a digital signal because it only has two distinct values. [3] This characteristic enables it to conduct analog signal processing in a digital environment. The idea to process signals using timing events is also shared with the most powerful known existent computer, i.e. the Human Brain. It is an analog computer, but it does not transmit commands based on analog signals, instead its information is represented and transmitted using timing of asynchronous digital-timing pulses [1].

Time-mode circuits advantages become even more appealing since CMOS rapid scaling results in the sharp increase of time resolution as the intrinsic gate delay of each transistor decreases. As a result,

time-mode circuits offer a viable and technology-friendly mean to combat scaling-induced difficulties encountered in the design of mixed-mode systems [2].

1.1 **Motivation**

The evolution of technology comes hand in hand with technology scaling to achieve better performance, i.e. power, speed and mainly cost. However, only digital circuits can take full advantage of technological improvements. [2] Analog circuits on the other hand face disadvantages, because they do not rely on fast operation speed, but on internal characteristics of the transistor such as the output resistance, and the intrinsic gain of a transistor. [2] In addition, voltage operation of these devices is severely reduced, adding to the fact that the device threshold does not reduce at the same rate as the voltage operation. As one can see in Figure 1.1, it results in a small headroom voltage to get proper measurements and if a cascode topology is used, an additional voltage drop will be consumed thus aggravating this effect. The shrinking of the voltage headroom will lead to a decrease in the Sinal-to-Noise ratio (SNR), meaning that the devices become much more susceptible not only to receiving noise from inside or outside sources, but also to the inability of obtaining a proper processing signal [4]. Other approaches could be used in order to improve the results, but this will lead to an even more complex and power-hungry circuit, which is contradictory to what should be obtained with technology scaling. [3]

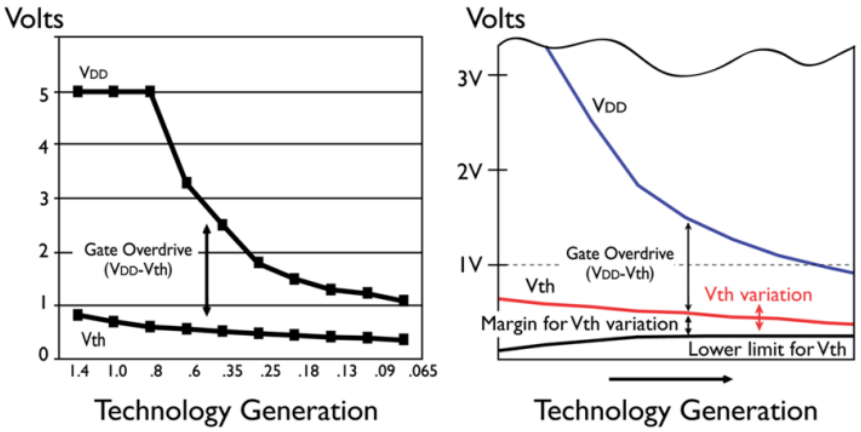


Figure 1.1 - V_{DD} and V_{TH} vs Technology Scaling [5]

Digital circuits on other hand do not present these disadvantages, having the capability of implementing their functions, except presenting the signal continually over-time, in a much smaller area, which results in a lower power consumption. Moreover, they present lower susceptibility in obtaining noise (caused by the increase in operation speed, and even from the type of signal itself), robustness to process variations, and cheaper production. However, there will always be functions that are better realized with analog approaches, so they should always be used when proved to be advantageous. However, if the analog circuits behaviour (i.e. performance, robustness, power consumption) could be increased with the assistance of digital circuits, it must be tried. [2]

Digital circuits cannot develop or represent any type of information in the analog domain, but they achieve a very high resolution in the time domain, due to consecutive gate delay reduction caused by technological scaling, offering exceptional time accuracy surpassing voltage resolutions in analog/mixed-signal approaches. “In a deep-submicron CMOS process, the time-domain resolution of a digital signal edge transition is superior to the voltage resolution of an analog signal”, Dr.R.Staszewski. [6] Considering the last statement, it can be alleged that time-mode circuits could be the answer to perform analog signal processing in the digital domain in deep-submicron technology because the performance of these circuits scales well with technology, they offer full programmability, low-power consumption and high-speed performance, and also the detrimental effect of technology scaling on the voltage-mode analog circuit processing disappears. [2,3,7]

This is where Time-to-digital converter (TDC) comes in. As a time-mode circuit, it represents nothing more than a precise stopwatch that converts continuous time domain information into a digital representation, providing a bridge between time domain and the digital world. In that case, it should be noted that only full digital Time-to-Digital converters get real advantage in time domain improvement, provided by technology scaling. Also, one thing to be aware of is the fact that TDCs are not a digital circuit. This seems contradictory with the last sentence, but what the TDCs really do is to rely on a full digital circuit to perform analog measurements exploiting all these advantages, but still behaving like a mixed-signal circuit.

Heretofore, TDCs were used as simple delay-lines mainly in all-digital phase-locked loops (PLL) [8,9] serving only as phase detectors, or even in the field of particles and high-physics, working as a chronometer where a precise time-interval measurement is needed [10]. TDCs However, driven by technology scaling problems in analog circuits and the potential of time-based circuits, micro-electronics communities have started to develop TDCs. Therefore, new applications started to emerge [11,12,13]. One of them is the main reason for developing this thesis, that is signal acquisition and processing. The technological scaling comes hand in hand with operation voltage reduction, so trying to produce analog-to-digital converters (ADCs) in the voltage domain becomes incredibly complex or impractical.

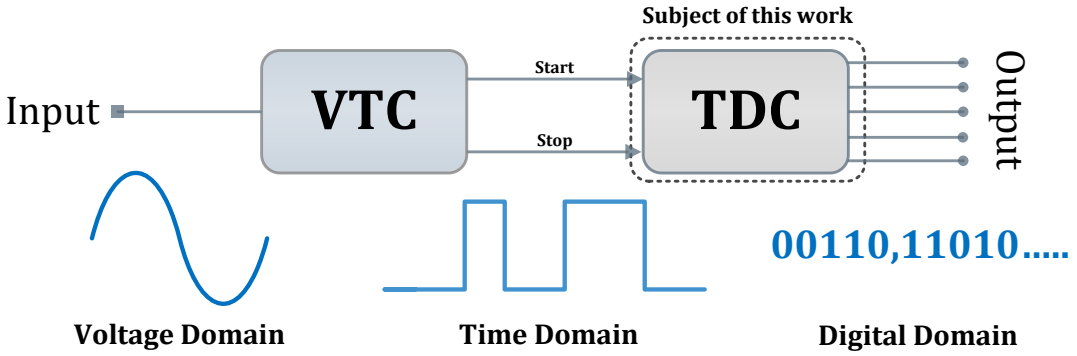


Figure 1.2 – Analog-to-Digital Conversion based in Time-Mode Circuits

Thus, provided that the continuous reduction of the gate delay increases the time resolution, better measurements in time domain can be obtained, with a smaller, simple, faster and less expensive circuit, maybe the answer could be the conversion of voltage signals into equivalent time intervals and using

TDCs to enable the time-domain processing, Figure 1.2. Therefore, with this key block the use of the well-known mixed-signal circuits to process voltage signals becomes possible in sub-micron technologies.

1.2 Goals

The main goal is to design a Time-to-Digital Converter with a range higher than 10 bits, able to measure time pulses higher than 2ns. Additionally, and due to the developed state-of-art projects know so far it is intended that the minimum resolution achieved by the Time-to-Digital Converter would be lower than 5ps.

1.3 Document Structure

The Present document is organized by the following chapters:

- Chapter 2 – State of the Art on Time-to-Digital Converters and State of the Art on Decoding Logic for Thermometer Code. Presents a brief study with the basic principles and concepts behind the most known TDC techniques so far. Additionally, introduces the most common architectures used to decode the Thermometer Code received by outputs of sampling units into a binary word.

- Chapter 3 – Time-to-Digital Converter Architecture, starts by describing the proposed TDC Architecture, giving an overview of the purpose of each chosen block in the system. Then for each block that was introduced before, an architecture is presented in detail and dimensioned.

- Chapter 4 – Results, describes all the obtained results from the proposed technique, including the Corners analysis. Additionally, it is proposed a Layout of the system, and the respective extracted results.

- Chapter 5 – Conclusion, describes the achievements of the proposed work and resumes and suggests possible suggestions for future work to improve the system performance.

2. STATE-OF-THE ART ON TIME-TO-DIGITAL CONVERTS

A time-to-digital converter maps a well-defined time interval into a digital word, and as one of the most important time-mode circuits they have become one of the main ways to overcome the challenge of low-voltage circuits in analog to digital signal conversion. Hence, time-domain offers a better solution compared to voltage-based methods, when implemented in deep sub-micron processes, taking advantage of the reduction in gate delay. [2,3,7,9]

With the growing interest in the development of this technology, various new architectures emerge increasing the resolutions, dynamic ranges, conversion times and lowering the power consumption.

This section will provide an overview of the existing architectures of time-to-digital converters, mainly focusing on its basic concepts, emphasizing all the advantages and disadvantages compared to other topologies. In addition, some specific performance parameters of the presented architectures will be summarily presented.

2.1 Analog Time-to-Digital Converters – The First Generation

In the first approach, the method Time-to-Digital Converters use is based on the conversion of time into voltage, quantizing that voltage into a digital word using a conventional ADC. [2,14]

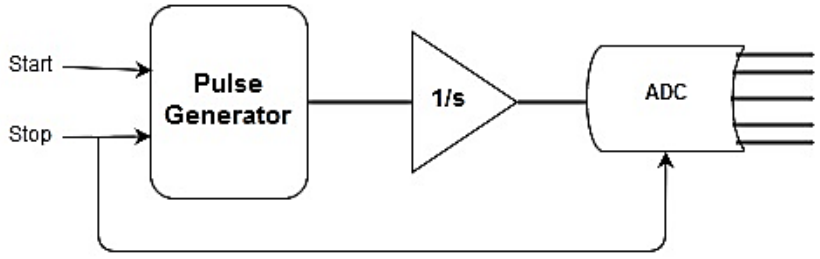


Figure 2.1- Single-slope Analog-to-Time-interpolation

The start and stop commands are used to form a pulse with a width corresponding to the time interval that needs to be measured. This pulse is then introduced to an integrator that transforms the pulse into voltage, which is given to the ADC that returns a digital word. [3]

The number of bits provided by the ADC will dictate the maximum and minimum time that the converter could measure, in other words, it defines the dynamic range.

$$DR = 2^N . T_{LSB} \tag{2.1}$$

As seen above, the minimum time that the ADC can achieve is T_{LSB}, which is dependent on the resolution of the ADC, and on the dynamic range measurement. If it is intended to measure long time intervals it will result in low resolutions, bigger T_{LSB}; if the goal is low time measurements, high resolution can be achieved.

One of the main problems of this architecture comes with the fact it must assure that all the blocks must have a linear response, and their elements suffer minimum mismatches. If these elements present deviations it will result in errors in the measurement.

One other approach is based on a dual slope ADC, Figure 2.2, [2,4,14] that instead of having an unknown voltage integrated in a well-defined time, it will have exactly the opposite, i.e. it will integrate a well-defined voltage during an unknown time. This technique gives us the possibility of overruling the effect caused by the variation of the absolute value of the components, as it does in the ADC.

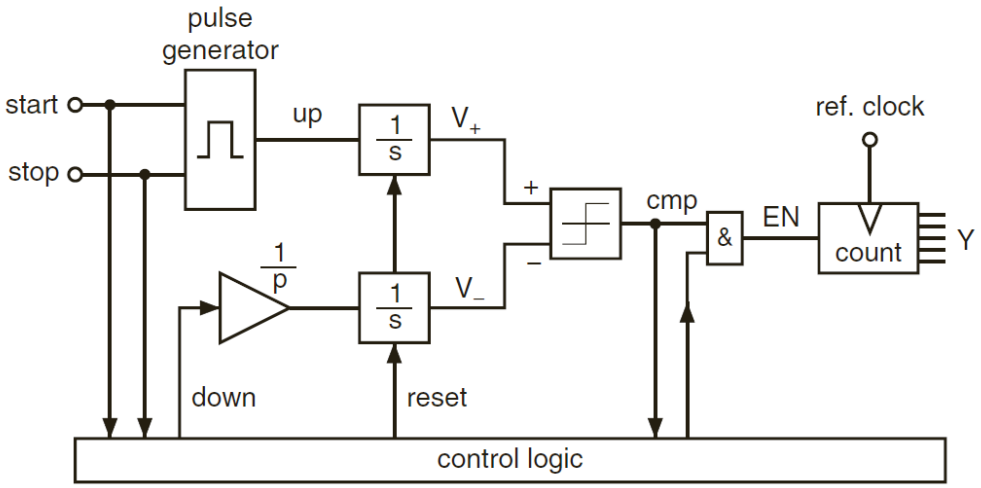


Figure 2.2 - Dual-slope Analog-to-Time-interpolation [2]

As well as in the previous architecture, the start and stop signals define the length of the time that is needed to be measured with the support of a pulse generator. The pulse generated is then integrated, achieving the voltage level V^+, so far nothing new. The difference in this technique is in the method of achieving the digital output when the stop signal arrives. With the stop signal, a second integrator is triggered, and only stops when the voltage V^- provided by the second integrator becomes equal to V^+. The integrator that makes the second integration has a reduced integration constant that is given by 1/p, which is related to the integration ratio of both integrators. [2]

The digital word is provided by a digital counter controlled by a reference clock, that counts the “time interval” where V^+ is bigger than V^-, so it can be concluded that the resolution of this architecture is a trade-off between the reference clock, the integration constant *p* and the dynamic range, as in the previous one is defined by the counter.

In both cases the mechanism that is used is mainly based on a conventional ADC, therefore the performance provided by this type of technique is greatly dependent on the linearity and mismatch of their elements [3,15], and also on the resolution of the voltage-mode ADC itself. Additionally, and as it has already been seen, the voltage mode ADC has a negative scaling effect with technology, therefore it is possible to conclude that analog-based TDCs will not also scale well with technologies.

2.2 Digital Time-to-Digital Converters – The Second Generation

The analog implementations that were discussed so far could implement precise timers if that is necessary. However, if it is intended to establish a mixed-signal interface with an analog timer, it will not benefit from the full advantages of time domain signal processing, even more so in scaled CMOS Technologies. Time domain signal processing can only be fully exploited if there is no analog conversion step in the time-to-digital conversion, being fully digital.

2.2.1 Counter Based TDC

The first generation of truly digital Time-to-Digital converters is based on a reference clock and a counter of the number of cycles that occur between the start and stop command [8,14], an example of the operation principle is showed in Figure 2.3.

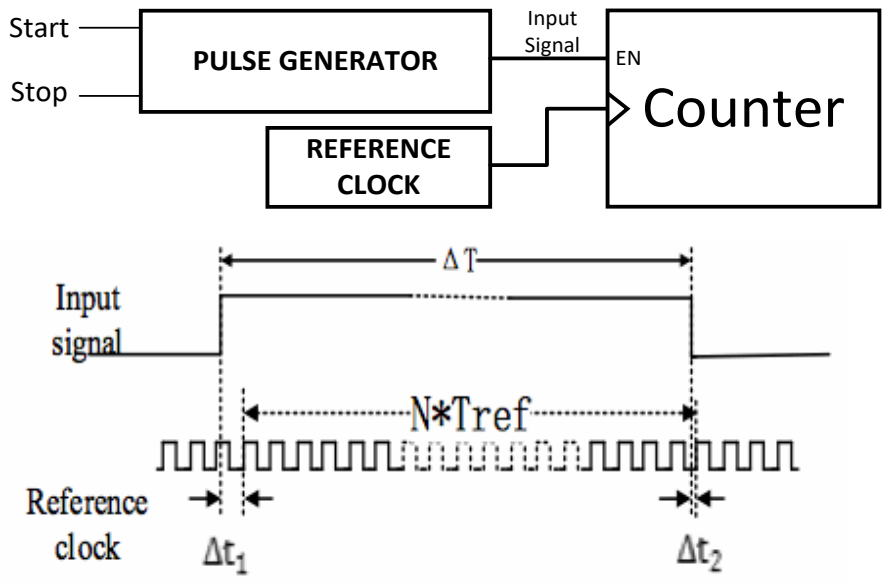


Figure 2.3 - Counter Based TDC

$$\Delta T = N.T_{ref} - (T_{ref} - \Delta_{t1}) + (T_{ref} - \Delta_{t2}) = N.T_{ref} + \Delta_{t1} - \Delta_{t2} \quad (2.2)$$

$$\varepsilon = \Delta t1 - \Delta t2 \in [-T_{ref}; T_{ref}] \quad (2.3)$$

It is simple to conclude that this approach has a resolution that is associated with the clock period, and the dynamic range limited by the range of the counter. As the start and stop commands are completely asynchronous with the reference clocks, it leads to quantization errors, $\Delta t1$ and $\Delta t2$, as both commands arrive. These quantization errors are limited between *-Tref* and *+Tref*, leading to a maximum error of *2.Tref* given that the clock is jitter-free. [2] One of the greatest advantages of this type of circuits is the large dynamic range and good linearity as the last one is determined by the stability of the frequency of the reference clock.

The accuracy could be increased by higher clock frequency. However, increasing the clock frequency comes with higher power consumption, and above a certain clock frequency CMOS oscillator are not available [2]. Another issue comes with the fact that the counter is also limited in frequency in order to

provide a proper measurement. One solution for obtaining high resolution with a low clock frequency can be achieved by subdividing the period asynchronously. The mechanism that performs this task is called a fully digital delay-line TDC.

2.2.2 Delay Line Time-to-Digital Converter

2.2.2.1 Basic Digital Delay Block: Inverter

The basic block behind all the full digital delay-line time-to-digital converts is the inverter. Therefore, before introducing the different TDCs architectures, some time should be spent in understanding the components and parameters that affect the behaviour of its main component.

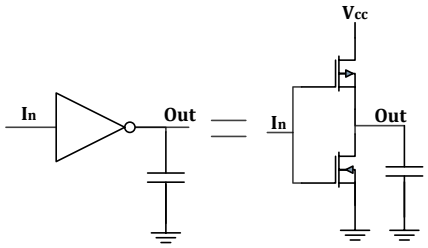


Figure 2.4 - CMOS Inverter Circuit

An inverter is the fastest block in CMOS technology. The switching speed is measured by its propagation delay-time, t_{pd}, which results from the average between the time for a high-to-low, t_{PHL}, and a low-to-high, t_{PLH}, transition that is limited by the time needed to charge and discharge the load capacitance C_L. Before proceeding, there are some terms that need to be defined: [16]

- Rise time, t_r, represents the time for a waveform to rise from 10% to 90% of its steady-state value.
- Fall time, t_f, represents the time for a waveform to fall from 90% to 10% of its steady-state value.
- Propagation Delay, t_{pd}, represents the time difference between the input transition (50%) and the 50% output level. In other words, it represents the time difference in a logic transition passing from the input to the output.

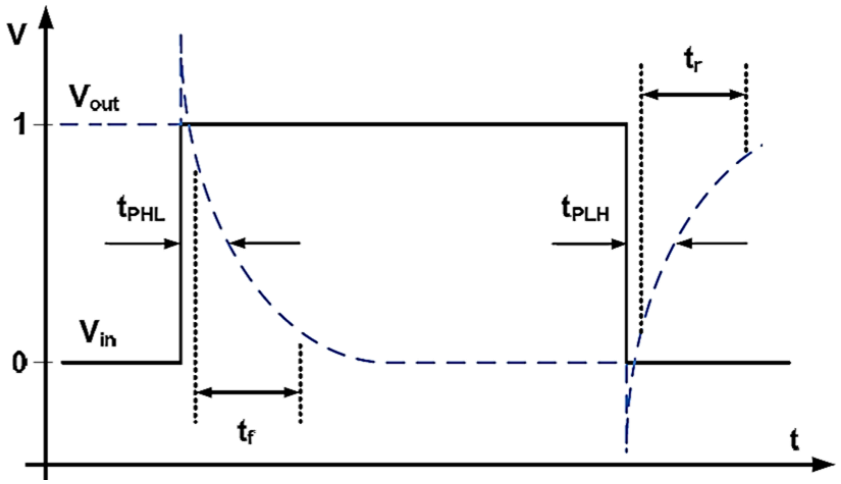


Figure 2.5 - Switching time diagram [9]

2.2.2.2 Falling-Time

The following Figure 2.6(a) shows a simple circuit with a capacitive Load C_L, connected to another inverter gate. Figure 2.6(b) represents the characteristic operation of a n-type transistor when a step voltage is applied at the input of the first inverter. Let's consider initially that the n-type transistor is cut-off and the capacitor C_L is fully charged, point X1 in Figure 2. (b).When voltage is applied at its gate, $V_{GS}=V_{DD}$, the NMOS transistor starts to conduct, point X2 Figure 2. (b), discharging the capacitor, resulting in the characteristic curve ending at point X3. To obtain the respective equation, two intervals must be considered:

- t_{f1} period during which the capacitor voltage, V_{out}, drops from $0,9V_{DD}$ to $(V_{DD}-V_t)$.
- t_{f2} period during which the capacitor voltage, V_{out}, drops from $(V_{DD}-V_t)$ to $0,1V_{DD}$.

The equivalent circuits for both cases are represented respectively in Figure 2.(c).

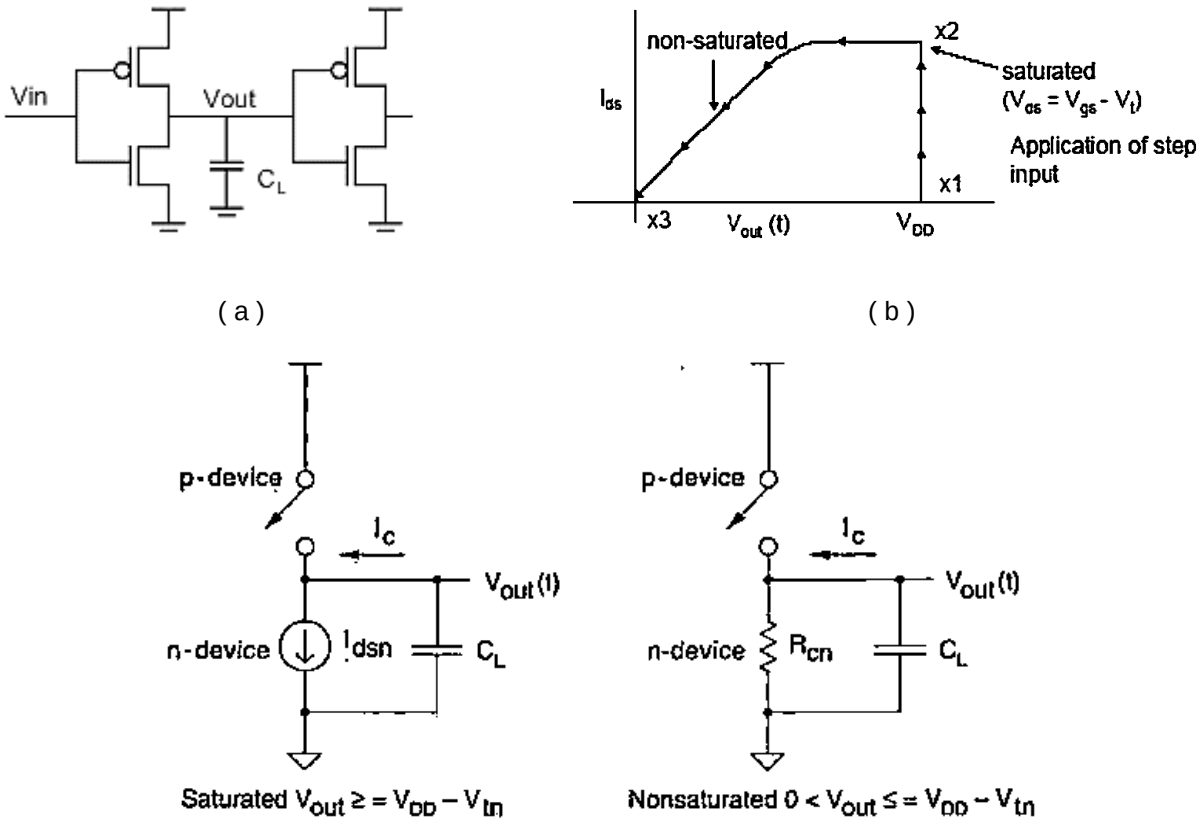


Figure 2.6 - Switching test circuit (a), trajectory of n-transistor operating point during switching (b), equivalent inverter falling circuit (c) [16]

Starting with the circuit represented in Figure 2.6(c) that corresponds to the equivalent circuit of the transistor working at the saturation region, the following equation can be assumed:

$$C_L \frac{dV_{out}}{dt} + \frac{\beta_n}{2}\left(V_{DD} - V_t\right)^2 = 0 \qquad \beta_n = \mu C_{ox} \frac{W}{L} \tag{2.4}$$

Considering the integration interval between t_{f1} and t_{f2}, the following expression is obtained:

$$t_{f1} = 2\frac{C_L}{\beta_n}\left(V_{DD} - V_t\right)^2 \int_{V_{DD}-V_t}^{0,9V_{DD}} dV_{out} = \frac{2 \times C_L(V_t - 0,1V_{DD})}{\beta_n\left(V_{DD} - V_t\right)^2} \tag{2.5}$$

When the voltage of the capacitor passes V_{DD}-V_t the n-type transistor leaves the saturation mode, entering the linear region. Within this area, the current passing through the transistor is no longer constant and it starts to behave like a resistor as is illustrated in Figure 2.6(c). The time, t_{f2}, taken to discharge the capacitor to $0,1V_{DD}$ can be obtained [17]:

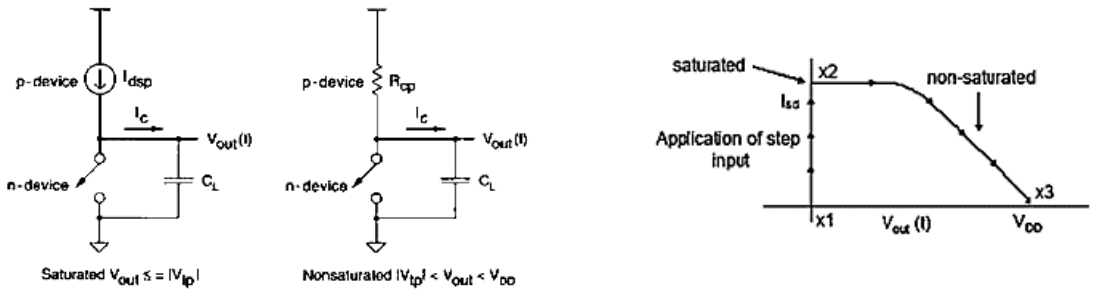
$$t_{f2} = \frac{C_L}{\beta_n\left(V_{DD} - V_{tn}\right)} \int_{0,1V_{DD}}^{V_{DD}-V_t} \frac{dV_{out}}{\frac{V_{out}^{\ 2}}{2\left(V_{DD} - V_{tn}\right)} - V_{out}} = \frac{C_L}{\beta_n\left(V_{DD} - V_{tn}\right)} \ln\left(\frac{19V_{DD} - 20V_{tn}}{V_{DD}}\right) =$$
$$= \frac{C_L}{\beta_n\left(V_{DD} - V_{tn}\right)} \ln\left(19 - 20n\right), \tag{2.6}$$

The complete expression for the falling time is given by:

$$t_f = t_{f1} + t_{f2} = 2\frac{C_L}{\beta_n V_{DD}(1-n)}\left[\frac{n-0,1}{1-n} + \frac{1}{2}\ln(19-20n)\right], \qquad n = \frac{V_{tn}}{V_{DD}} \tag{2.7}$$

2.2.2.3 Rising-Time

A similar approach can be made to obtain the equation for the rising time, Figure 2.7. Consider now that initially the p-type transistor is cut-off and the capacitor C_L is fully discharged, point X1 Figure 2.7 (b). When the gate is connected to ground, V_G=*0V*, PMOS transistor starts to conduct, point X2 Figure 2.7 (b), resulting in the characteristic curve ending at point X3.



(a) (b)

Figure 2.7 – Equivalent rising inverter circuit (a), Trajectory of p-transistor operating point during switching

$$t_r = 2\frac{C_L}{\beta_p V_{DD}(1-p)}\left[\frac{p-0,1}{1-p} + \frac{1}{2}\ln(19-20p)\right] \qquad p = \frac{\left|V_{tp}\right|}{V_{DD}} \tag{2.8}$$

Both equations are pretty much similar. However, due to the fact that the carrying mobility of the n diffusion is higher than the p diffusion, *N-MOS* transistors present a faster switch corresponding to a

faster fall-time. Thus, if it is intended to make a balanced inverter, it must be ascertained that both fall time and rise time are equal, resulting in: [16]

$$\frac{\beta_n}{\beta_p} = 1 \tag{2.9}$$

To make this possible, the channel width of the p-mos device must be increased, and by rule of thumb it is known that the increase is around two to three times the n-device. [16]

$$W_p = 2 \sim 3W_n \tag{2.10}$$

Up to now the only aspects that have been considered to obtain these equations were the effects caused by the device itself and the capacitive load connected to the output. However, if the input signal is not perfect, it will also have its own rise and fall time which affects the invert.

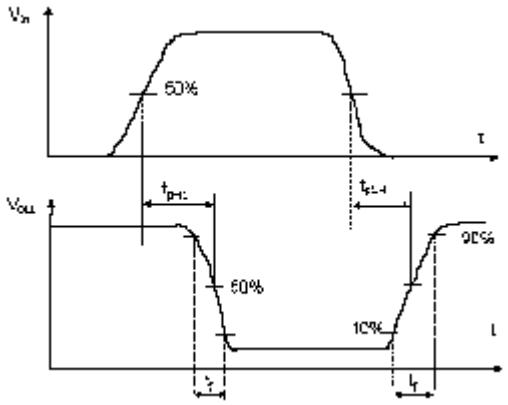


Figure 2.8 - Rise and Fall time graph of Output affected by the Input [17]

Therefore, if these times are considered in order to obtain a result closed to reality the t_{PHL} and t_{PLH} will be given by:

$$\tau_{PLH_{real}} = \sqrt{(\tau_{PLH})^2 + \left(\frac{t_f}{2}\right)^2} \qquad \tau_{PHL_{real}} = \sqrt{(\tau_{PHL})^2 + \left(\frac{t_r}{2}\right)^2} \tag{2.11}$$

2.2.2.4 Propagation Delay-Time

Propagation delay-time is dominated mainly by t_{PHL}, and t_{PLH}. These two fully depend on the rising and falling time of the device. Mathematically it can be expressed it by the following equations,

$$t_{PLH} \approx \frac{t_r}{2} \approx \frac{C_L}{\beta_p V_{DD}(1-p)}\left[\frac{p-0{,}1}{1-p} + \frac{1}{2}\ln(19-20p)\right], \qquad n = \frac{V_{tn}}{V_{DD}} \tag{2.12}$$

$$t_{PHL} \approx \frac{t_f}{2} \approx \frac{C_L}{\beta_n V_{DD}(1-n)}\left[\frac{n-0{,}1}{1-n} + \frac{1}{2}\ln(19-20n)\right], \qquad p = \frac{|V_{tp}|}{V_{DD}} \tag{2.13}$$

Accordingly, as said before, the propagation time is the average between the low-to-high and high-to-low transition time, turning out to be, [16,17]

$$t_{pd} = \frac{t_{PLH} + t_{PHL}}{2} \tag{2.14}$$

Taking a closer look at the equation it can be concluded that t_{pd} is mainly affected by the capacitive load, transistor size, threshold Voltage and Supply Voltage. [9]

The inverter propagation delay has a large dependence on the load capacitance that consists of the total amount of inverter intrinsic capacitance at the input and output nodes and also their parasitic capacitances. To achieve high speed, the capacitance seen by the output node of the inverter must be minimized. It could also be noted particularly by the β constant that the t_{pd} is inversely proportional to the *W/L* of a transistor. So, if the width of the transistor is increased, keeping the length low, it will increase the amount of current that passes through the channel, decreasing the time taken to charge and discharge the load. However, increasing the transistor width will have a drawback. It will increase the parasitic components attached to it, so it must be verified that the decrease in the propagation time will justify the size change the transistor will suffer. [9]

Threshold Voltage also has influence in the propagation delay. The lower it gets, the lower t_{pd} will be. However, this factor is influenced by the technology itself and by the conditions where the transistor is operating, so there is not much that can be done to improve it.

Finally, the supply Voltage has a big impact on result time, and is highly related to the threshold voltage. Higher values of voltage, within technological standards, will lead to more driving strength, therefore a lower delay is achieved, as can be seen with the factors *n* and *p* in the equations (2.12) and (2.13), respectively.

2.2.3 **Buffer Based Line**

The principle behind a TDC based in a delay-line is the consecutive delay of a reference clock through a line with N elements. In this method, the reference clock is nothing more than a command that triggers the delay chain, and the delay imposed by those elements is what defines the resolution of the TDC. Each element in the delay-line can be achieved with a simple buffer, and the sample elements with type D flip-flops. [2,14]

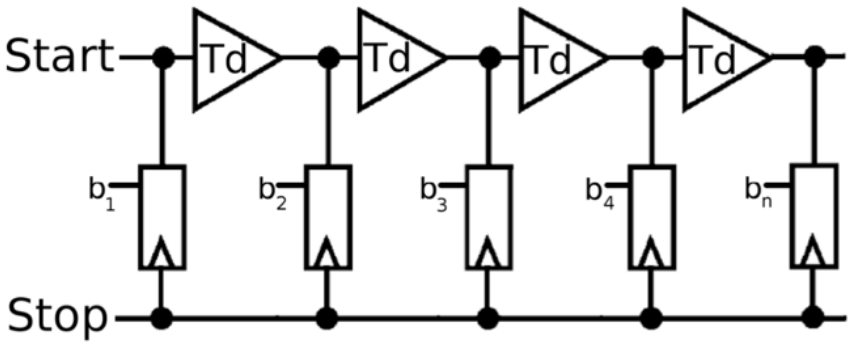


Figure 2.9 - Basic delay-line TDC

The working principle behind this architecture is very simple. A start command enters the delay line and goes through a series of buffers, whose outputs are connected to the data inputs of each flip-flop.

With the arrival of the stop command, each one of the versions of the start command is sampled in parallel. The result, also known as Thermometer code, says how many delay elements the signal passed through by exposing the level logic high at the output. [2,3]

The time interval ΔT, can be obtained by:

$$\Delta T = N.T_{LSB} + \varepsilon \tag{2.15}$$

Where *N* is the number of elements with a HIGH output, T_{LSB} is the delay imposed by each element in the delay-line, which is the TDC resolution, and ε describes the quantization error associated with the process.

The structure itself is the simplest one in this kind of applications, with an easy embedded control; low conversion time, $T_{in} + \tau_{Conv}$ where τ_{Conv} is the time used to present the proper code, [3] and its dynamic range will only be limited by two factors, the first one is the power/area consumption because the bigger your delay-line becomes, the more power and area it will use. The second fact is related to the mismatch that could occur in the delay elements due to the process parameter variation. If this happens it will be a source of differential nonlinearity (DNL), which is accumulative as the signal goes along the delay line. [18] The maximum time resolution that this technique could achieve depends entirely on the technology used to implement the delay-line.

2.2.4 **Inverter Based Line**

The buffers used as a delay-element are no more than two inverts connected in series, so if in the delay-line all the buffer elements are exchanged with simple inverters, the delay of each element will be reduced by half. Accordingly, it will double the time resolution despite any noxious effect that may exist in the system.

The first difference that needs to be considered is that the returned code from the delay-line no longer dictates the number of elements that the start command has gone through. Unlike the previous one, it does not return a sequence of HIGH level logic to determine the time interval, instead, it returns a code with a sequence of alternating ones and zeros. Hence in order to obtain the time measure, the current state needs to be compared with the previous one, looking for a phase change in the alternation HIGH-LOW sequence. [2,8]

One issue is that the rising and falling transitions of the CMOS inverters are different (even if partially correlated due to common process steps in the production of PMOS and NMOS devices), leading to non-linearity in the characteristic converter [2]. In nominal process conditions, these asymmetries can be annulled, but any variation that occurs in the system leads to new asymmetries in the devices.

Hence, to obtain a high resolution TDC based on an inverter delay seems to be impossible. However, there is a solution, as is shown by Figure 2.10.

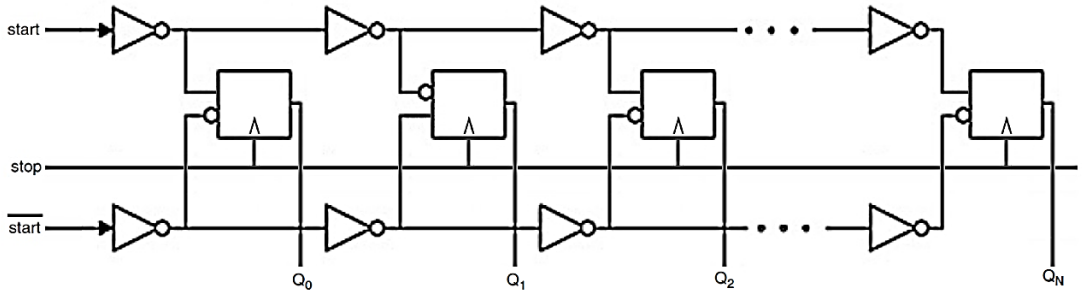


Figure 2.10 - Inverter Based delay-line TDC

This solution is based on a second inverter-based delay-line triggered by an inverted start signal, interconnected by fully symmetrical differential flip-flops. Both lines will provide differential data which will compensate possible asymmetries in the setup time of the flip-flops and in the rise and fall delays of the inverters. To complement, the inverters must be alternately connected to the inputs of the flip-flops. It is important to be aware that the process variations could make a faster signal propagation in one delay-line when compared to the other one, so in order to prevent this effect, cross coupled pairs of inverters could be used between both delay lines, thus reducing the difference. [2]

Considering both topologies, delay line and inverter-based delay line are the basic and easiest full digital TDC topologies, and since neither attempt to achieve sub-gate delay resolutions, the demands of the delay elements are not so stringent, making them very easy to control and implement. Moreover, they are also very “robust” to mismatch errors in future technology scaling. However, in some applications the resolution achieved by these topologies is not enough, and the fact that they do not possess sub-gate resolution becomes a fundamental drawback, when compared with other known types of time-to-digital converters.

2.2.5 **Variations effects in Delay Line TDCs**

Considering the previous statement about the propagation delay of an inverter, one must consider that the elements presented in the delay lines are not perfectly linear or equal. In reality they may suffer some variations affected by Process, Voltage and Temperature, PVT, variation effects. These factors have independent and uncorrelated effects causing a variance in the propagation delay that could be expressed as $\sigma^2_{tpd,PVT}$ [9].

One other cause of propagation delay variation is the mismatch of the delay stages rising the nonlinear behaviour of the TDC. Mismatching of elements may cause uncertainties in the threshold voltage, V_{TH}, device trans-conductance, β, and drain current due to possible change in the transistors sizes. These fluctuations will trigger the increase of propagation delay mismatch, contributing to the output skew between consecutive elements.

$$\sigma_{tpd,charmis} = \frac{t_{pd}^2}{4}\left(\frac{2\sigma_{C_L}^2}{C_L^2} + \frac{\sigma_{\beta_n}^2}{\beta_n^2} + \frac{\alpha\sigma_{V_{TN}}^2}{(V_{DD}-\sigma_{V_{TN}})^{\alpha}} + \frac{\sigma_{\beta p}^2}{\beta_p^2} + \frac{\alpha\sigma_{V_{TP}}^2}{(V_{DD}-\sigma_{V_{TP}})^{\alpha}}\right) \quad (2.16)$$

Where, $\sigma^2_{V_{TP}}$, $\sigma^2_{V_{TN}}$, $\sigma^2_{\beta_n}$, $\sigma^2_{\beta p}$ and $\sigma^2_{C_L}$ are the variances of V_{Tp}, V_{Tn}, β_p, β_n, and C_L respectively. Note that to make this equation (2.16) valid, the t_{PHL} and t_{PLH} are assumed to be equal, making the element symmetrical, choosing the proper nMOS and pMOS widths. [19]. Besides, if it is assumed that each element varies independently, with a standard deviation $\sigma^2_{tpd,charmis}$, these delay variations will be accumulated through the line. Therefore, the maximum mismatch variance will be at the end of the inverter chain [9].

$$\sigma_{tpd,mismatch} = \sqrt{N} \times \sigma_{tpd,charmis} \tag{2.17}$$

Lastly, jitter mismatch is another source of timing error, contributing to the variation of the propagation delay. Luckily these timing errors tend to be much lower than the ones caused by mismatch of the delay elements. There are many sources that contribute to jitter, they include switching noise, cross talk and device noise. These effects have already been deeply analysed in [20,21].

$$\sigma_{tpd,jitter} = \frac{4kT\gamma t_{pd}}{I_{L,sat}(V_{DD} - V_{TH})} + \frac{kTC_L}{I_{L,sat}{}^2} \tag{2.18}$$

With *k* representing Boltzmann's constant, temperature *T*, noise factor γ (typically equal to 2/3), output Capacitance C_L, threshold voltage V_{TH}, and discharge current $I_{L,sat}$.

As in the device mismatch, jitter accumulates through the chain so, the bigger the delay line, the bigger the jitter error, arising from the fact that any error occurring at an early transition of the delay line will affect all subsequent transitions of the delay line.

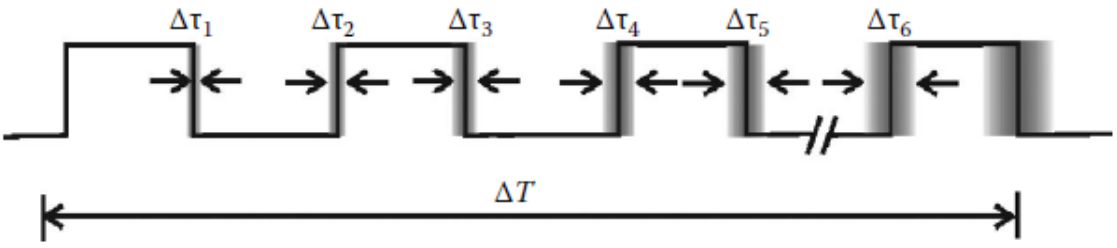


Figure 2.11 - Jitter accumulation through the delay-line [3]

2.2.5.1 Meta-stability in Time-to-Digital Converts

Metastability is one of the undesired effects the sampling element stages of a time-to-digital possess. This has a huge impact on the performed measurement, which may cause errors at the output binary decoded word. This side effect has greater relevance in asynchronous systems than in synchronous systems, because while in synchronous systems it is known how often the sampling process is performed; in asynchronous systems, this is no longer true. It will not be known for sure if the clock signal will or will not violate the setup and hold time of our sampling units. So, if one of these two conditions is spoiled, the output will end at an intermediate level. This can cause the system to fail if it has not been resolved before a certain time, because of its inconsistent logical state. The possibility of

entering into the metastable state will depend mainly on 3 factors, transition data rate, clock frequency, and the intrinsic characteristics of the data circuit. However, if the relation between the data and clock varies, this effect is unavoidable (asynchronous systems). [3]

Let's take a deeper look in order to really understand this effect, looking at the circuit represented at Figure 2.12.

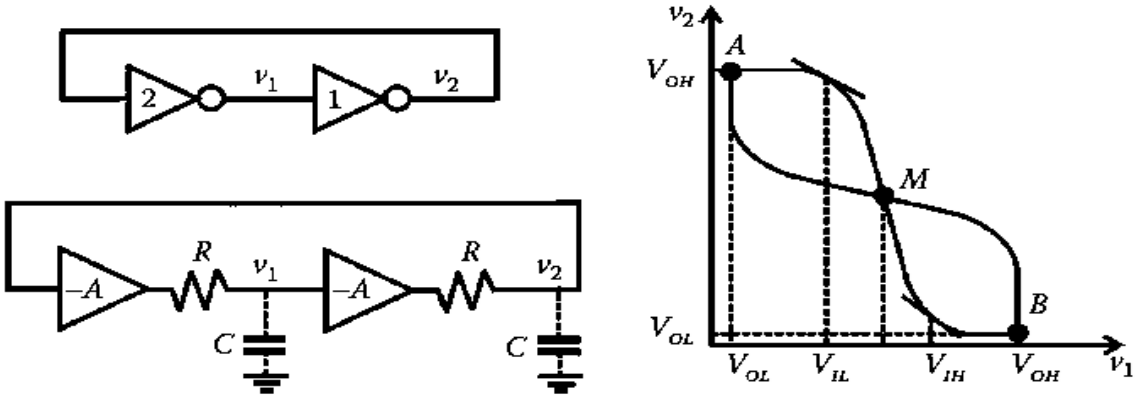


Figure 2.12 - Metastable state of bistable circuits [3]

Metastable state is one of three states that exist in a bistable circuit, usually located at an unstable equilibrium voltage of usually $\frac{V_{dd}}{2}$, considering that the inverter is balanced. In the graphic on the right, state A and B represent the stable states and to move them a voltage bigger than V_{IL} is needed. At the M state, both PMOS and NMOS transistors of the inverter are at saturation state, so in a perfectly matched and no noise environment, they will remain there indefinitely. The ability to overcome this condition is mainly dependent on the characteristics of the circuit itself, such as its gain and parasitic capacitances, although the surrounding environment may help in overcoming this state [3].

2.3 Advanced TDC Design Techniques

Before taking a step into the deep sub-gate resolutions some techniques that can be used with the majority of the TDCs architectures in other to improve some of their faults or weaknesses thus improving their performance without great effort.

2.3.1 Bipolar Time-to-Digital Converter

In most applications, the start command is the first to arrive, meaning that normally all measures are positive. However, if for some reason the stop command arrives first, the measured time returned by the TDC would be zero, and with the arriving of the start signal, the converter will continue to measure time until it overflows.

For some applications, the time interval that needs to be measured could be "negative", which means that the order of arrival is reversed, one possible application might be type II PLL's, Phase Locked Loop, more specifically in the phase detector, thus it could detect positive and negative differences.

One of the first approaches to solve the issue is based on the delay of the stop command, allowing the start command to arrive first. The main problem comes with the unknown offset in the converter's characteristic response, which is susceptible to process variations leading to changes in the circuit operation.

Another more advanced possibility uses 2 TDCs, where one of them measures the time interval between the start and the stop command (forward TDC) and the other one measures the complementary time interval, that is between the stop and start signals (reverse TDC). Therefore, none of the converters will measure negative time intervals. The bipolar output is obtained by subtracting the forward TDC from the reverse TDC.

This architecture exhibits a huge drawback, which is the offset. This nonlinear characteristic is a huge disadvantage compared with other TDC techniques in many applications, particularly in feedback systems where a nonlinear element inside the loop may cause instability or break the system's performance. Also, the delay offset may be highly susceptible to PVT variations, and it could generate a positive output even though the STOP signal arrives before the START signal, in a "negative time" measurement case. [2]

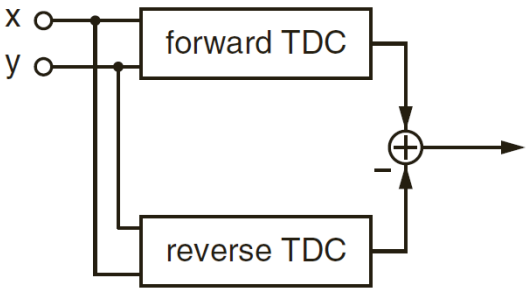


Figure 2.13 - Basic Bipolar TDC

2.3.2 **Looped Time-to-Digital Converter**

All the TDCs that have been discussed so far are based on a single delay-line, with a maximum time that can be measured, ΔT_{max}, dependent on the number of the elements in the line. Hence, if it is necessary to measure times that are bigger than ΔT_{max}, the only possibility so far could be adding more elements to the line, which leads to huge area and power consumption in spite of the difficulties in the layout and process variations. This could be avoided with a loop architecture, which counts the number of times that a signal travels through all the line, making it possible to measure huge time intervals with a few elements in the delay-line and a counter, reducing drastically the area and power consumption.

Therefore, the time measured with the loop architecture will be given by:

$$B = M.N_{count} + B_{TDC} \tag{2.19}$$

where M is the number of elements in the delay line, N_{count} is the number of cycles completed and registered by the counter and B_{TDC} is the actual position of the signal when the stop signal arrives.

Unlike the previous implementations, the timing event that goes through the delay line will be a pulse, and it must be, as if the time to measure needs a second lap, there must be a way to check the position in the delay-line where the stop occurs. [2]

The simplest implementation for the looped TDC technique is shown in Figure 2.14.

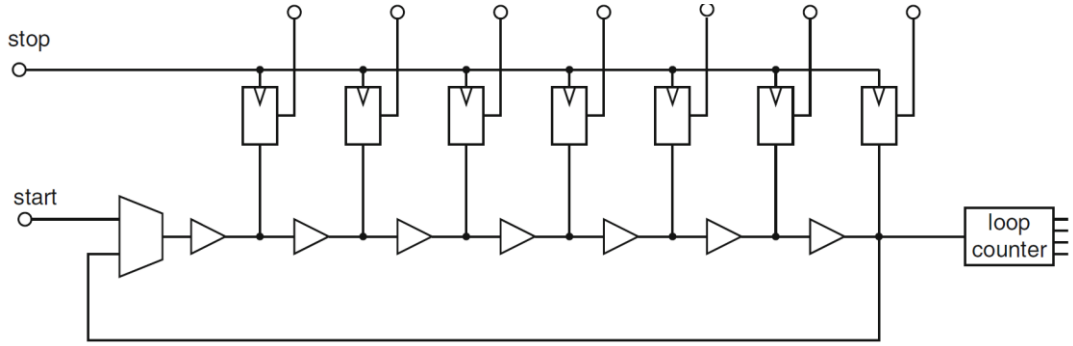


Figure 2.14 - Looped Time-to-Digital Converter [2]

A start event is injected into the delay-line through a multiplexer. The control unit switches the multiplexer input to the end of the delay-line and wait there for the arrival of the start event. As the start event emerges and with no stop signal detected, the multiplexer puts the start event at the beginning of the delay-line for a new cycle. Each time the start event reaches the end of the line, the counter is triggered and counts one more cycle completed. When the stop signal occurs, both delay-line and counter are sampled, and the circuit is prepared for a new incoming measurement (the multiplexer switches the input, and both delay-line and counter are cleaned).

One of the major problems with this architecture is the detection and processing of the events associated with the arrival of the stop command. The stopping command needs to have a certain width to be correctly detected, and it must be ascertained that the command arrives at the same time to all sampling elements in order to avoid errors in the measured time, which in turn could be of at least 1 T_{LSB}, and if the pulse is circulating near the count could lead to a extra count.

To prevent this last effect, a new approach, Figure 2.15, can be used, where a technique is used that only enables the counter when the timing event approaches its input. To perform this task, two extra flip-fops are used to check the relative position of the event. With this extra information plus the thermometer code, it will be possible to turn the counter on and off and even correct it if for some reason an unexpected delay occurs, and an extra count is performed, preventing huge errors in the returned result (N.LSB).

So far, area consumption is the main reason to go for looped TDCs, assuring long time measurement times with high resolutions. However, there is another advantage that is some-how correlated with the previous one. The error grows through the delay-line, meaning bigger delay-lines introduce bigger errors in the final result. Therefore, implementing loop architectures reduce drastically the measured error, because in short delay-lines only a few variations can accumulate, making it possible to compensate either by digital calibration or by a controlled-delay line, like a Delay-locked loop TDC.

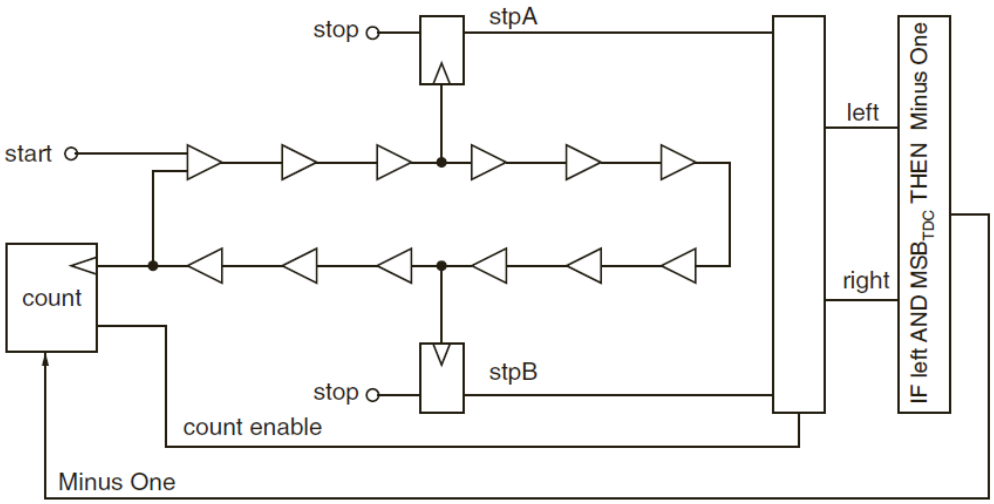


Figure 2.15 - Stop signal detection improvement in Looped TDC [2]

However, this approach has its own negative points that need to be pointed out. The layout is very difficult, because it needs to assure that the loop geometry has minimum asymmetries to promote linearity in the measurement. Other problem is the return of the loop that if not done correctly leads to non-linearities. Another source of non-linearity is the multiplexer itself because the delay that it performs is different from the ones in the delay-line, and changing all the elements to multiplexers is not the answer. It will only increase area and power consumption as well as it degrades the resolution. Another major problem is related to pulse shrinking (or growing). This affects all types of architectures, specially the loop architectures, which become the most impaired, limiting the time that can be measured [2].

2.3.2.1 Linearly Extended Time-to-Digital Converter Loop

This architecture is an improvement of the classic looped TDC. The start point comes from the need to delete the asymmetries caused by the layout and coupling multiplexer in an attempt to increase the linearity response of the system. The idea of the following architecture is to make use of the TDC itself to measure the asymmetries of the feedback loop, and later correct the result.

This method presented in Figure 2.16 consists of two TDCs interconnected, assuring that neither disturbs the other, a main TDC and extended TDC. The main TDC is designed like a normal looped TDC, and the extension will be a simple delay-line, connected to the end of the main TDC.

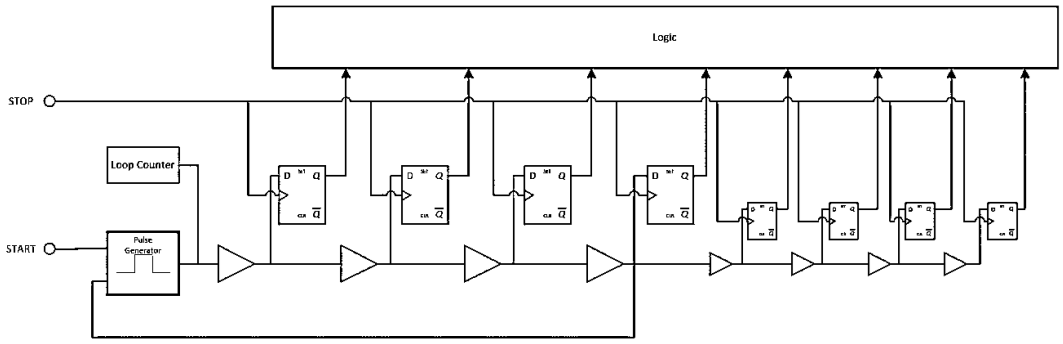


Figure 2.16 - Linearly Extended TDC Loop

The objective of the extended TDC is to continue the measurement while the timing event travels through the feedback loop, allowing the process to quantize the delay imposed by the feedback loop, enabling for example a proper result to be obtained if the stop command appears while the pulse is in the feedback loop. The only rule that must be ascertained is that the total delay time is longer than the time the signal takes to travel through the feedback loop of the main TDC.

So what type of improvement does this type of architecture provide? Firstly, the asymmetries from the feedback loop are annulled by the extended delay-line. The layout becomes easier to implement, there is no need to have a very careful design, so automatic layout generation could fulfil it. Secondly, the design does not have problems with pulse shrinking or growing because it has a pulse generator that provides a new pulse in each loop cycle, improving the accuracy of the measurement. The drawback for using this architecture is the increased area and power consumption as well as the more complex control logic required in the process. [2]

In summary, there are 3 possible scenarios that could happen when the stop signal appears:

- The timing event is still in the main TDC. Here the measurement is quantized similarly to a simple looped delay-line according to,

$$t_d = N \times t_L + N_1 \times T_{LSB} + N \times T_{fb} \tag{2.20}$$

where t_L is the complete loop time of the main TDC, N is the number of loops that the time event made, N_1 is the current position in the main TDC where the stop signal occurred and T_{fb} is the time taken for the signal to travel through the feedback loop.

- The event signal was detected in the extended TDC, but not in the main TDC. This means that the event signal is travelling through the feedback loop. This is where the extended TDC comes into play, so the measurement result is

$$t_d = N \times t_L + (max(N_1) + N_2) + N \times T_{fb} \tag{2.21}$$

where N_2 is the measurement value from the extended delay line and $max(N_1)$ represents the full length of the main TDC.

- Finally, it is when the event is detected in both delay-lines, main and extended. Here the event is in the transition from the feedback loop to the main TDC. In this case, it could use either expression from the last cases. However, this event has a special meaning for this circuit, it can be used to calculate the feedback loop time, T_{fb} in the circuit.

2.3.2.2 Multi-Event Time-to-Digital Converter

Until now, the concepts that were referred previously were about single individual time measurements. But what if consecutive measures need to be performed?

One of the main reasons that makes this operation impossible with the current techniques is the dead-time that exists at the end of each measurement in order to get the results and prepare the TDC for the

next one. And what about multiple time events at the same time? That would be pretty much impossible with the techniques discussed so far.

A first approach considers multiple delay-lines in parallel, selected by a multiplexer. However, this results in huge area and power consumption and if there are mismatches between elements and between the delay lines, it will result in errors in the measurements, as well as a non-linear characteristic response. [2]

A multi-event time-to-digital converter is a new approach that allows multiple and consecutive measures using a single looped delay-line.

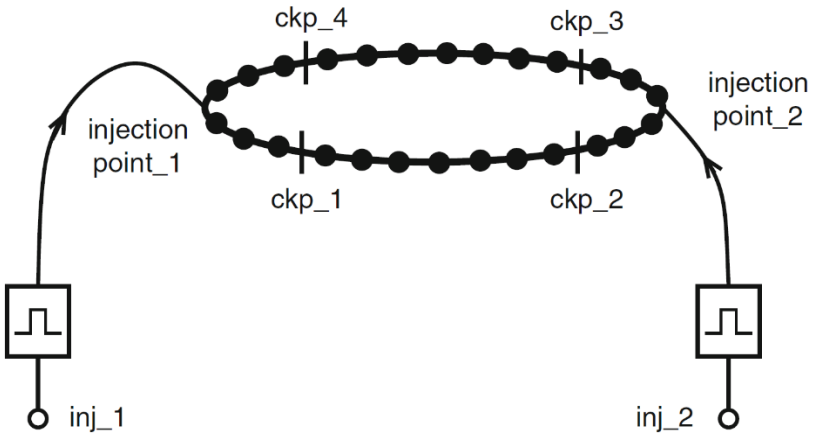


Figure 2.17 - Principle of multi-event TDC [2]

The delay-line is a closed-loop not much different from the classic ones, although it has a few extras. Unlike others, METDC, Multi-Event-TDC, has more than one input typically named “injection point”. The timing event is made by a pulse generator connected to each injection point.

The delay-line also comprises check points before and after each injection point to choose the exact injection point where a new timing event should be introduced. So, when a timing event that is circulating in the loop comes near an injection point, the checker before sends a signal command that blocks the respective injection point. When the command passes through the consecutive checker the block signal is removed and the injection point is now ready to receive a new event. This strategy is important because it prevents a new event being injected above the others which were already circulating in the loop.

The main challenge of this type of converter is obviously the control logic behind it, because in addition to choosing the right spot to introduce the event, it must be ascertained that all the measures are kept in track and each pulse is easily identified and related to its measurement. Also, if the event gives more than one lap to the circuit, the control must correctly count the number of loops each event does.

Even if the converter processes multiple events at the same time, there is a problem that arises with the injection point. What if there were timing events near all the injections? It will be impossible to perform a new measure, leading to information loss in some cases. So, there is a maximum number of events that can be processed. That depends mainly on the length of the loop, width of the pulse, number of

trackers and of course the delay done by each element.

Another challenge arises when a timing event measurement ends. The pulse cannot continue to circulate in the loop forever, so the control logic must create a way to clean the respective pulse. The difficulty in cleaning is that the pulse's right location needs to be known, in order to force down the output where it is currently passing through.

In conclusion, the advantages of this architecture are the reduced area, multiple measures and consecutive measures with relatively low asymmetries. The main disadvantage/difficulty is the complex control logic behind the whole process.

2.3.3 **Hierarchical Time-to-Digital Converter**

The architectures that have been discussed so far use a full resolution technology to measure all the time intervals independently of their size. So, if a long-time interval needs to be measured with a high resolution, a huge power will be consumed, and a high number of bits will be needed to present the result.

Hierarchical TDC gives a solution that reduces the power consumption in long time intervals and measures it with a high resolution using the same number of bits. For this, it uses two TDC, the first performs a coarse quantization, using low power and area. The second performs a fine quantization with a low error, having a small dynamic range.

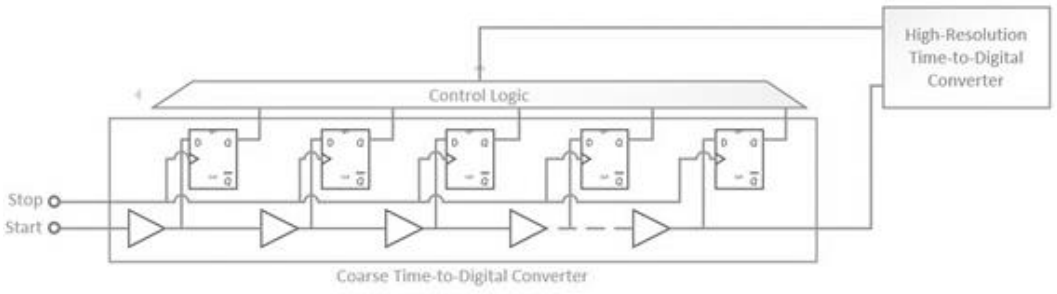


Figure 2.18 - Hierarchical TDC

The coupling of Both TDCs is made by a control logic and a coupling multiplexer and that is where this architecture shows its drawback, due to the time latency generated. Other problem is the layout project and wiring of all the components and the multiplexer itself. Even if it has a symmetrical built, the wiring may cause some asymmetries, so it must use the minimum connections possible, to take full advantage of this type of time converter.

2.3.4 **Delay-locked-Loop Time-to-Digital Converter**

DLL, Delay-Locked-Loop, TDC is one of the most important and well know architectures. It can be associated with almost any time of TDC, giving a huge improvement to the linear response of the system, and making it work as it was a synchronous system. The resolution and performance of the TDCs architectures depend mainly on the delay provided from the inverters, however there are always PVT variations [3,4].

The delay-locked-loop Based TDC is an architecture that is designed to provide absolute time measurement, based on a reference clock. So, what it does is to subdivide the reference clock, in a given number of sub-intervals, calibrating them. The number of intervals depends in the number of elements in the delay-line. [2]

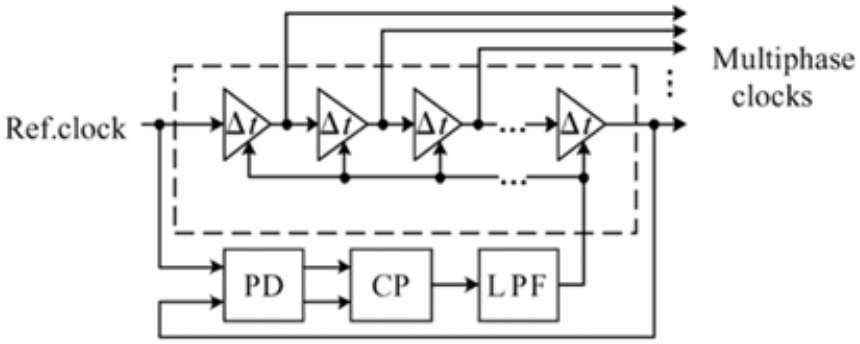


Figure 2.19 - Delay-locked-Loop TDC

In order to calibrate the delay-line, periodic multiple start commands are injected into the delay-line, and after each element a skewed copy of the signal is created (sub-intervals). At the end of the delay-line a sample is compared with the input, checking the phase difference between both signals to adjust the delay-line. If the phase difference is not zero, the delay provided by each element is changed until the phase of the input and the output becomes the same. To change the delay of each element, additional blocks must be inserted with the intention of changing the slew-rate of the inverters, and normally what is done is tuning the delay elements via a controlled supply voltage or current limitation [2].

Most of the DLL TDC are configured to hold a single period clock in the delay-line, but it can also hold multiple periods. However, with this comes a problem that a basic phase detector cannot solve: false-locking. Let's imagine that it has a clock period configured in the delay-line but, there is a signal that frequency is a multiple of the reference clock, making it possible that the DLL locks in the wrong flank. To avoid this effect a phase-frequency detector can be used.

In general, this architecture makes a huge contribution to the performance of the TDC, but a considerable part of this circuit is from the analog domain, considering all the advantages of having a full digital TDC, as previously discussed. There is an emergent digital approach that measures a known reference signal during idle periods of the system, which uses this result to correct all the next measurements during a digital post-processing. Anyway, the analog implementation is still predominant. One other drawback is the fact that it loses resolution in each element of the delay-line due to the tuning of the system.

2.4 Time-to-Digital Converters with Sub-Gatedelay Resolution – Third Generation

All types of architectures that have been reviewed so far, have their maximum resolution limited to the gate delay, which depends on the specifications of the technology that is being used.

The third generation focuses on techniques that can achieve resolutions under a gate delay, without changing the technology. However it will still be limited by asymmetries and process variations that may occur. In the deep-submicron regime, the consequent increase of the leakage current precludes further constant field scaling. Hence, if it does not have new concepts, TDC resolution would not increase as it supposes for the next few technology nodes. [4]

2.4.1 **Parallel Scaled Delay Element Time-to-Digital Converter**

The delay achieved by the inverters depends on the technology that it is using, but also scales linearly with the transistor width and load capacitance.

This relation can be expressed by the following equation:

$$t_{pd} = p + g.h \tag{2.22}$$

where p is the parasitic delay, g the logic effort and h the fan-out of the gate ($h = \frac{C_{load}}{C_{in}}$).

Hence, this technique has the capability of manipulating the *h* parameter either by modifying the width of the transistor or scaling the output load capacitance. In both cases the delay of each gate element will change independently. [4,8] In a regular delay-line the coefficient h is not considered, because each gate has an identical gate load.

The following Figure 2.20, presents an example of an implementation for this architecture.

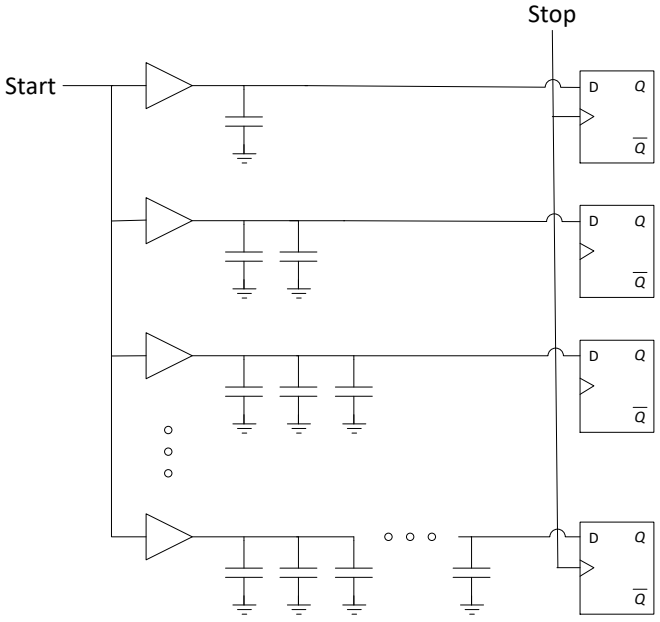


Figure 2.20 - Parallel Scaled TDC

This consists of a parallel chain of gates each one with a different load capacitance. The load is changed by *N* equal capacitances in parallel to each delay element, where *N* is the number of delay elements in the chain, resulting in a very linear system response. The parasitic delay and the effort that is made to drive the sampling elements contribute to the offset of the system but fortunately does not affect the measurement accuracy. [2]

So, with the arrival of a start signal, it will propagate through all the delay elements in parallel. On the rising edge of the stop signal, the output of all delay elements is sampled, just like an ordinary delay-line.

Merging the previous information with the first equation, the time delay of each element can be described by,

$$T_{dn} = {t_d}^0 + \Delta t_d . n \tag{2.23}$$

where ${t_d}^0$ is the technology resolution, Δt_d is the variation introduced by each capacitance and n is the number of capacitances in series with the delay element. Hence, it can be concluded that the resolution of this architecture is given by $T_{LSB} = \Delta t_d$. The maximum range for this architecture is given by the number of parallel branches, *N*, and the minimum time that it can measure is ${t_d}^0$, which is quite a big offset, meaning that it is impossible to measure a time interval lower than the element delay and local variations can even increase it. [8]

By tuning the delay output of each delay element, it can achieve almost a quasi-continuously response, however as well as the offset, this response is also affected by process variations, thus resulting in loss of resolution and bubble errors. Another point is that theoretically it can have any possible resolution intended, because any capacitor size can be built, but in practice this is limited by process parameters.

This technique also presents very restricted rules in the layout concept, especially with high dynamic ranges, that is because it must be ascertained that the start and stop signal get to all the branches “at the same time” to have a precise measurement, and the respective pulse generator has enough power to provide current to the circuit. [2]

2.4.2 **Vernier Time-to-Digital Converter**

Vernier topology is the most popular technique in the TDC sub-gate resolution world, not only because it has a huge versatility, meaning that it can be implemented in various ways (simple delay-line, looped structure, hierarchical, fractional, two-dimensional), [2,3,8] but also because of its simple principle of operation and very high resolution. Hence, in order not to make it exhaustive this work will only focus in the principle behind all the Vernier TDCs known.

The Vernier TDC structure consists of two parallel delay-lines, one to receive the start command and the other to receive the stop command, both with the same number of delay elements although with a slightly different propagation delay. The basic configuration and operation is shown in Figure 2.21. [22,23,24]

The delay elements on the upper chain, where the start command is inserted, are slightly bigger than the elements on the lower chain, where the stop command is inserted.

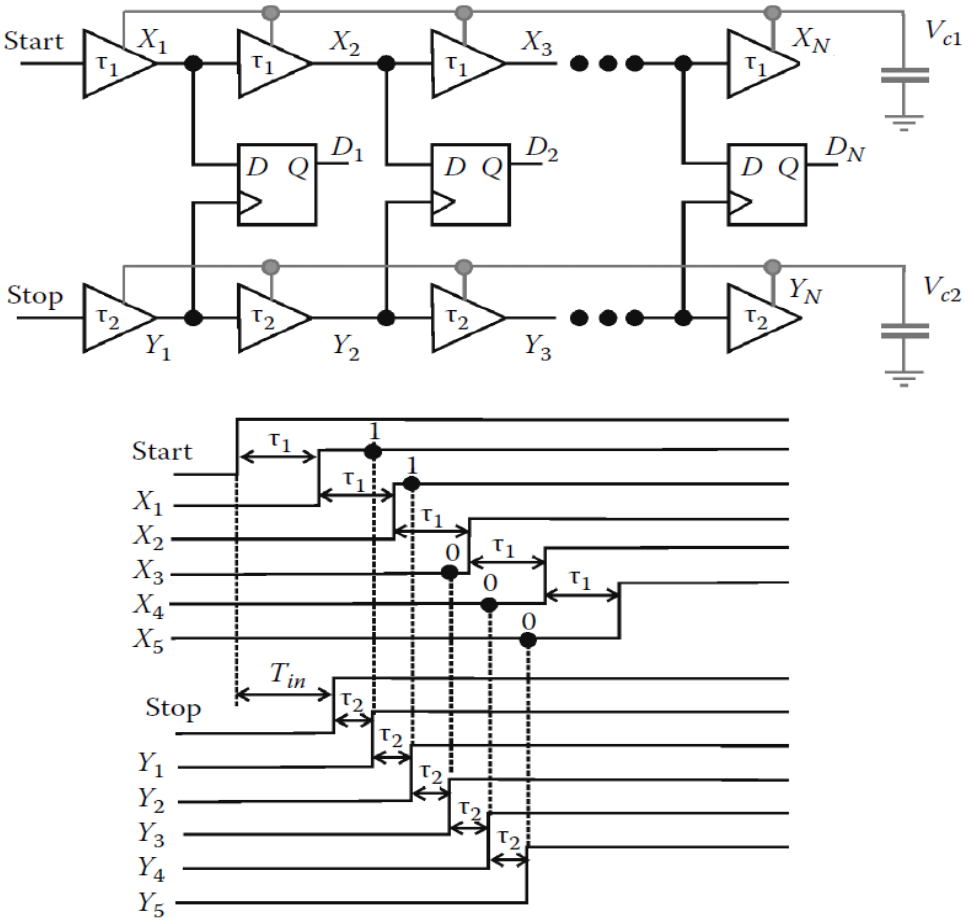


Figure 2.21 - Cut-out of a Vernier delay-line TDC and its operation principle [3]

This means that the signal that travels through the stop chain is faster, therefore in each stage the stop command will catch up with the start command, $T_1 - T_2$ being the difference between both delay-lines and consequently the resolution of Vernier TDC, T_{LSB}.

At a certain time, the stop signal passes through the start signal, marking the end of the measurement. This instant time is detected by Early-Late-Detectors, ELDs, normally implemented with flip-flops and determined by: [2]

$$T_{catch} = N\tau_1 = N\tau_2 + T_{in} \tag{2.24}$$

where N is the number of delay elements that the Stop and Start signal passed through before the catch up occurs, and T_{in} is the input signal that needs to be measured.

As the skew between both commands is reduced by T_{LSB} in each stage, the number N of elements that is needed in both chains to perform a certain time measurement, T, is given by,

$$N = \frac{T}{\tau_1 - \tau_2} \tag{25}$$

Note that each stage is composed of two delay elements, one from each chain, and an early late detector.

The dynamic range is upper- bound by $N(\tau_1 - \tau_2)$ and lower-bound by $(\tau_1 - \tau_2)$, however, similarly to the classic delay-lines, the effects of the mismatches and PVT variations have a huge impact on the dynamic range of Vernier delay line TDCs, preventing it from scaling linearly with the number of delay elements. So, a Vernier loop structure has a huge advantage here because there could be big time intervals using few delay elements. [2]

As an upgrade, if the delay-line is big enough, and if the measured samples become quite large compared to the resolution of the converter, it may be possible to inject a new measurement with a previous one still running. However, it will require additional control effort to provide proper measurement results, similar to the multi-event time-to-digital converter that was referred previously.

2.4.3 **Pulse - Shrinking Time-to-Digital Converter**

Pulse-Shrinking architecture [25,26,27] is another major method of achieving sub-gate delay resolution, based on unequal sized inverters. As in the architecture presented earlier, pulse-shrinking TDC can be implemented in numerous ways using different techniques and sometimes merged with other architectures, making it a very versatile system.

The basic principle behind this technique is based on changing the width of a pulse over the delay-line, instead of having a constant command, as in the previous techniques. The measurement pulse is generated by the rising edges of the start and stop event [2,4].

The structure of the delay-line is very similar to a basic buffer delay-line, however, in opposition to what happened in the previous techniques, where equal sized elements were used at the same delay line, here each element has an intentional asymmetry that causes a change in the width of the pulse So at a certain point of the delay-line, depending on the measure time, the pulse vanishes completely and the measure is complete. The basic structure is shown at in next Figure 2.22: [3]

The pulse-shrinking TDC consists of a chain of homogenous stages, where the elements present a similar rising and falling response, and inhomogeneous stages, where intentional asymmetries makes the inverter possess different rising and falling times. When a pulse passes through these stages the width reduction is controlled by the charge and discharge of the load capacitor, which is controlled by the size of the devices which in turn will change the respective high-to-low and low-to-high propagations delays.

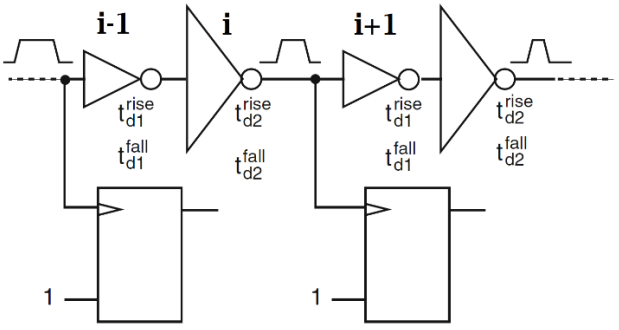


Figure 2.22 - Pulse-Shrinking TDC [2]

As it is shown in the last schematic, each one of the inhomogeneous stages, *i*, stands between two homogenous stages, i-1 and i+1. This mechanism is made in such a way that it is possible to accurately control the pulse shrinking done by this member. To quantify the shrinking time made in the pulse, the falling and rising expressions of the inverters need to be considered, as already shown in equation (2.12) and (2.13). According to the equations and transposing them to this case, the capacitance load is in fact the input capacitance of the next element, then at the i-1 to i transition, assuming that NMOS and PMOS have the same threshold voltage, in order to make this effect possible it must be assured that $\tau_{PLH,i-1} > \tau_{PHL,i-1}$, therefore the reduction in the width of the pulse when it passes from the stage (i-1) to the stage i is given by, [28,29,30]

$$\Delta\tau_{(i-1)\to i} = \tau_{PLH,i-1} - \tau_{PHL,i-1} = \alpha C_i \left(\frac{2}{\beta_{p,i-1}} - \frac{2}{\beta_{n,i-1}} \right) \tag{2.26}$$

where,

$$\alpha = \frac{2V_T}{\left(V_{DD} - V_T\right)^2} + \frac{1}{V_{DD} - V_T} \ln\left(\frac{19V_{DD} - 20V_T}{V_{DD}} \right) \tag{2.27}$$

Looking at the second stage where the pulse crosses from the element i to the element i+1, it will have a similar expression,

$$\Delta\tau_{i\to(i+1)} = \tau_{PHL,i} - \tau_{PLH,i} = -\alpha C_{i+1} \left(\frac{2}{\beta_{p,i}} - \frac{2}{\beta_{n,i}} \right) \tag{2.28}$$

The total pulse-shrinking time is the result of the sum of both factors previously showed,

$$\begin{aligned} \Delta\tau &= \Delta\tau_{(i-1)\to i} + \Delta\tau_{i\to(i+1)} \\ &= \alpha \left[C_i \left(\frac{2}{\beta_{p,i-1}} - \frac{2}{\beta_{n,i-1}} \right) - C_{i+1} \left(\frac{2}{\beta_{p,i}} - \frac{2}{\beta_{n,i}} \right) \right] = T_{LSB} \end{aligned} \tag{2.29}$$

As it has been said before, the inhomogeneous stage stands between two homogeneous stages, and if it is established that these elements are proportionally related, meaning $W_i = kW_{i-1} = kW_{i+1}$, (keeping the length of all elements equal), the last expression can be simplified to,

$$\Delta\tau = \alpha C_{i-1} \left(\frac{2}{\beta_{p,i-1}} - \frac{2}{\beta_{n,i-1}} \right) \left(k - \frac{1}{k} \right) \tag{2.30}$$

From the last equation, it may be concluded that it will always have pulse-shrinking effect for k≠1, despite PVT effects. The structure itself has a good linearity because the way it is constructed and controlled makes the system perform under the parameters previously calculated.

So, with this simple structure, when compared to the Vernier way, high resolutions can be achieved with a basic principle and without having very complex control logic behind it. To achieve a good dynamic range a large number of stages are needed. Consequently, the possibility of jitter and mismatch effects

in the elements increases with it deteriorating the linearity of the converter in a similar way to the classical delay line TDCs. Similarly, also the PVT effects on the system will affect the rising and falling time of the gates, consequently changing the performance of the systems. [3]

However, the critical negative point is the minimum pulse that needs to exist in order to be sampled by the flip-flops, creating a big offset to the system. Before the pulse totally vanishes, it becomes very small, being unable to reach a detectable logic level. Therefore, the sampling systems have a big role to play in the behaviour of the systems and consequently in the final result. There are techniques to calculate and mitigate this offset error, one of them will be presented at a later stage.

2.4.4 Time Amplifier Time-to-Digital Converter

In this last section, a technique that enables all types of time-to-digital converters to achieve sub-gate resolution will be presented.

All the TDCs that have been discussed so far achieve their high single shot resolution based on the way that their architecture is built and, on the way, that the logic control behind them coordinates it. Time-Amplifiers on the other hand, they do not achieve high resolution and neither performs time-measurements. What they really do is to improve the measurement quality performance by the actual TDCs in use; they “grab” the time interval inserted to measurement and “stretch it along the time axis”. [2]

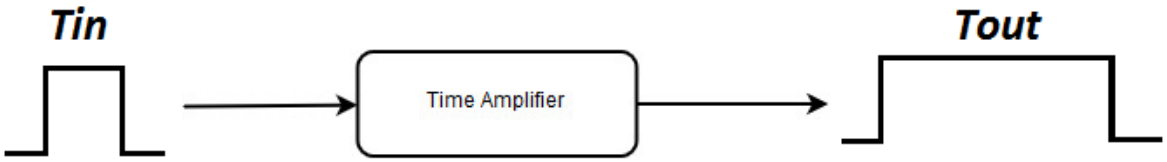


Figure 2.23 - Time Amplifier Principle

There are several techniques to perform these tasks, some of them based on analog principles like the dual-slope approach of Nutt [3]. Based on the several factors that have been discussed before and on fully digital approaches, such as the pulse-Stretching delay-lines, they are not the most recommended. [3].

Besides increasing the resolution of the standard techniques, it also helps to increase the linearity of the circuit itself, decreasing the offset, for example. However, the maximum amplification that the interval can support is limited mainly by the dynamic range of the TDC.

The time-amplifier circuitry itself also needs to be carefully designed and implemented. Any type of variation that the circuit has or gets will be reflected as a time error that will be introduced in the sample, leading to an error in the measurement result.

So, there are definitely huge advantages to use this circuit alongside the main techniques discussed so far; just by using a simple delay-line it is possible to achieve a huge resolution. However, there is a trade-off between the resolution and the intended dynamic range.

2.5 Time-to-Digital Converters Summary

Day by day, precise time measurements are becoming more useful in many applications, especially with technology scaling. Precise results require advanced electronic techniques. In the best traditional methods, the lowest single shot measurement precision is around 50ps. The TDC performs a conversion of a well-defined time interval into a digital (binary) word, and due to the rapid development of these techniques it could easily get single shot resolutions lower than 20ps.

Due to the advantages mentioned before, the digital methods are preferred to the analog ones. However, there is an exception, which is the Time-Amplification method, because of its easiness to implement in integrated circuit technologies, and its very low sensibility in being disturbed by external sources.

However, the possibility of being affected is not out of the question, causing issues in the performance of the TDCs. The main causes are process and environmental parameters such as PVT variations, changing the internal characteristics of each element. The best way to prevent these effects is to pay special attention to the construction rules of the layout itself.

In this chapter, a brief discussion of the principles behind the main architectures and design techniques were provided. Furthermore, there were presented some performance metrics in order to get a full picture of the techniques advantages and limitations.

Delay-lines TDCs (buffed or inverter based) were the first full digital method existent, providing a good resolution, limited by the technology itself, and low power consumption. However, a small dynamic range, and increasing the length of the delay-line will only lead to the increase of possible nonlinearities in the system's performance. Note that this phenomenon is not exclusive to this technique, all other architectures that use a simple delay line can be affected.

Parallel scaled delay-line is one of the simplest mechanisms to achieve sub-gate resolution, highly increasing the resolution of the TDC, based on scaled capacitors connected to the output of a parallel delay element. The drawback is related to the huge area consumption, the possible parasitic effects that could lead to the change of the modulated capacitors, and mainly the balance of the start net signal that need to get to all branches at the same time making the layout construction a very critical process.

A delay-line based on the Vernier principle is known for its high capability to measure time intervals with a very high resolution, based on 2 differently sized independent delay-lines. However, this good resolution and linearity holds a high cost in silicon and power consumption.

The Pulse-shrinking comes somewhere between the classic delay-line and the Vernier technique because its design is basically a delay-line with different sized elements built into it, capable of producing sub-gate resolutions, getting some of the advantages and disadvantages of both cases. The area and power consumption is much smaller when compared to the previous one, achieving equivalent resolutions, but suffers from long conversion times, and it has a considerable offset time that makes it not so appealing in measuring very short times, compared to the propagation delay of the inverter.

Table 2-1 – Comparison study on Time-to-Digital Converter

	TECHNIQUE USED									
	Buffer Based Delay-Line TDC		**Inverter Based Delay-Line TDC**	**Parallel Scaled Delay TDC**	**Vernier TDC**			**Pulse Shrinking TDC**		
Advantages	• Simple Structure • Low Power • Low Latency • Easy control and embedding		• Simple Structure • Low Power • Low Latency • Double Resolution	• Sub gate-delay resolution • Low Conversion and Latency Time • Good Linearity • Simple structure	• Sub gate-delay resolution • Modular Structure • High Dynamic Range • Good Linearity			• Sub gate-delay resolution • Modular Structure • High Dynamic Range • Good Linearity • Simple Structure • Low Area		
Disadvantages	• Low Resolution, limited by technology • Big offset		• Alignment of delay-lines • Resolution limited by technologic • Doubled number of components for the same dynamic range	• Big Offset • Susceptible to variations • No loop structure, leading to low dynamic ranges • Huge area consumption • Careful layout design	• Needs two delay-lines • High Conversion and latency time • High Area and Power consumption • Difficult control and layout			• Minimum pulse offset • High Conversion and latency time • Difficult Layout		
Reference	[31]	[32]	[33]	[34]	[24]	[35]	[36]	[37]	[27]	[25]
Tech [nm]	90	130	90	350	130	65	90	800	350	180
Resolution [ps]	20	60	15-20	68	8	4,8	13,6	20	47	1,8
No. of Bits	-	-	-	7	12	7	-	-	-	9
INL [LSB]	0,7	7	0,7	-	1	3,3	0,73	0,5	0,11	8,7
DNL [LSB]	0,6	6	0,6	-	-	<1	0,96	-	0,1	1,2
Power [W]		2,86m	1,69m	-	7,5m	1,7m	0,23m	-	-	3,4m
Area [mm^2]	0,01	0,116	0,01	0,03	0,26	0,04	0.01	2	-	0,07

2.6 **Decoding Logic for Thermometer Code**

In the Last Chapters, some of the main techniques used to properly measure a well-defined time-interval using a Time-to-Digital converter, have been discussed. However, there is a final step needed to obtain the result. This consists of the conversion of the TDC code into a binary word. As an ADC, the parallel outputs of the core time-to-digital converters do not generate a binary word, but rather a sequence of 'high' and 'low' bits that indicates the position where the measurement ended in the delay-line. [2,38] That sequence is known as Thermometer Code, which is an entropy encoding that represents a natural number, *N*, by a sequence of *N* 'ones' followed by a zero (considering that the count start at zero), or *N-1* 'ones' followed by a zero (considering that the count start at one). [39]

Table 2-2 - Truth table for Thermometer to binary code converter

Decimal Number	Thermometer Code	One-hot code	Binary Code
0	0000000	0000000	000
1	0000001	0000001	001
2	0000011	0000010	010
3	0000111	0000100	011
4	0001111	0001000	100
5	0011111	0010000	101
6	0111111	0100000	110
7	1111111	1000000	111

As a result of the development of the ADCs, specially the flash ADC, due to its high speed, TDCs can be included here. Choosing the proper design for a decoder is very important in order to obtain a non-error result with the minimum delay possible. Additionally, it is intended that our design achieves low area and power consumption to increase the performance parameters of the final circuit. [40,41] Finally, it is very important to take into account that the performance of the measurement is deeply affected by invalid transitions, or their lack, in the thermometer code. These types of errors are known as bubble errors, which consist of the presence of *'zeros'* in the N ones sequence. The main reason that could lead to this type of errors, are the uncertainty introduced in the sampling moment, due to metastability, offset, and cross talk. [42,43]

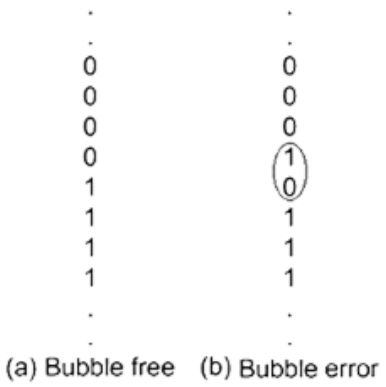


Figure 2.24 - Bubble error example [31]

The Thermometer code to binary code, TC-to-BC, converter is traditionally implemented with a two-stage method [44], as shown in Figure 2.25. The first step consists of the conversion of the thermometer code provided by the sampling circuit to a one-out-of-N (also known as one-hot code), or grey code, depending on the type of converter chosen.

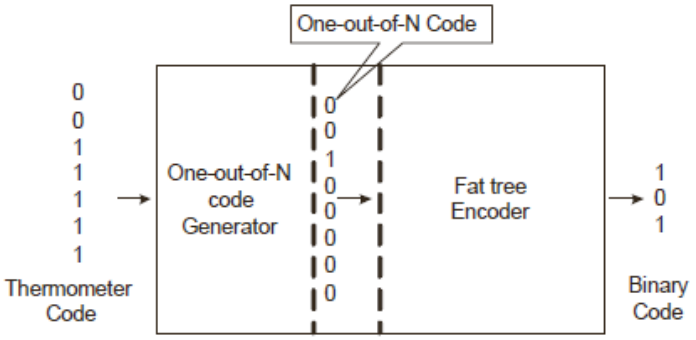


Figure 2.25 - The two-stage fat tree TC-to-BC encoder [38]

The second and final stage consists of the conversion of that last code into a binary work. There are various types of techniques to achieve this purpose, direct conversion architectures using multiplexers and intermediate decoding tech techniques using basic gates or digital ports.

2.6.1 ROM based Thermometer to Binary Converter

The most popular technique used to perform a TC-to-BC conversion is the ROM circuit, mainly because of its fully parallel decoding structure. ROM is basically a programmable logic device, so it is not restricted to these types of applications. There are two types of ROM to perform this task, one based on a binary structure, the other based on a grey code structure. Grey code type is commonly used because of its resistance to bubble errors. [41,42] The appropriate row, *m*, in the ROM is selected with the output of the sampling unit *m* plus the output of the sampling unit *m+1*, both connected to a 2-input AND gate where one of its inputs is inverted. If our circuit is susceptible to bubble errors, a 3-input AND Gate could be used to prevent this from happening. [40,45]

The main advantages of this type of decoder are its simple construction and well-known design. However, its speed conversion is low and has considerable power consumption due to the constant static current that is used to perform its reset. Another problem is derived from the increased data speed, that makes it more susceptible to bubble errors. With this, a more complex bubble error correction circuit is needed to increase even more the area, power consumption and delay in the circuit. [44]

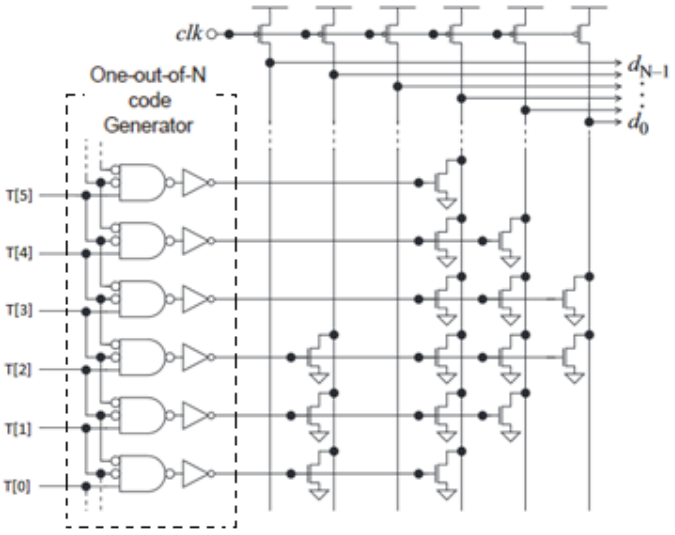


Figure 2.26 - ROM Based Thermometer-to-Binary Decoder

2.6.2 **Multiplexer based Thermometer to Binary Converter**

This architecture is well known among ADC decoding topologies mainly because of its low power and high speed performance. This technique presents a versatile output because it can directly convert the thermometer code to binary code, or it can convert to grey code for posterior analyses and processing. Additionally, it offers some bubble error suppression, and it has a regular structure that benefits the layout of this circuit. [40]

The main principle behind the building of the circuit design is based on the basic principle used to obtain logical circuits, i.e. translation of the truth table into a subsequent finite state machine, which results in Boolean equation that are translated into a multiplexer circuit.

To an N-bit ADC, the output binary word will be obtained with (N-1) stages of 2:1 multiplexers arrays. The most significant bit (MSB) of the binary code output will get a high logic level if more than half of our thermometer inputs are 'ones'. Therefore, this bit will be equal do the 2^{N-1} thermometer input. The remaining thermometer inputs are connected to the first multiplexer stage, controlled by the MSB. The second most significant bit will be the 2^{N-2} output from the multiplexer array. The third most significant bit will follow the same principle, as it is the result of the 2^{N-3} multiplexer output from the second stage controlled by the second most significant bit. This sequence will continue until only one multiplexer reaming, resulting in the last bit of the binary word. [41,46]

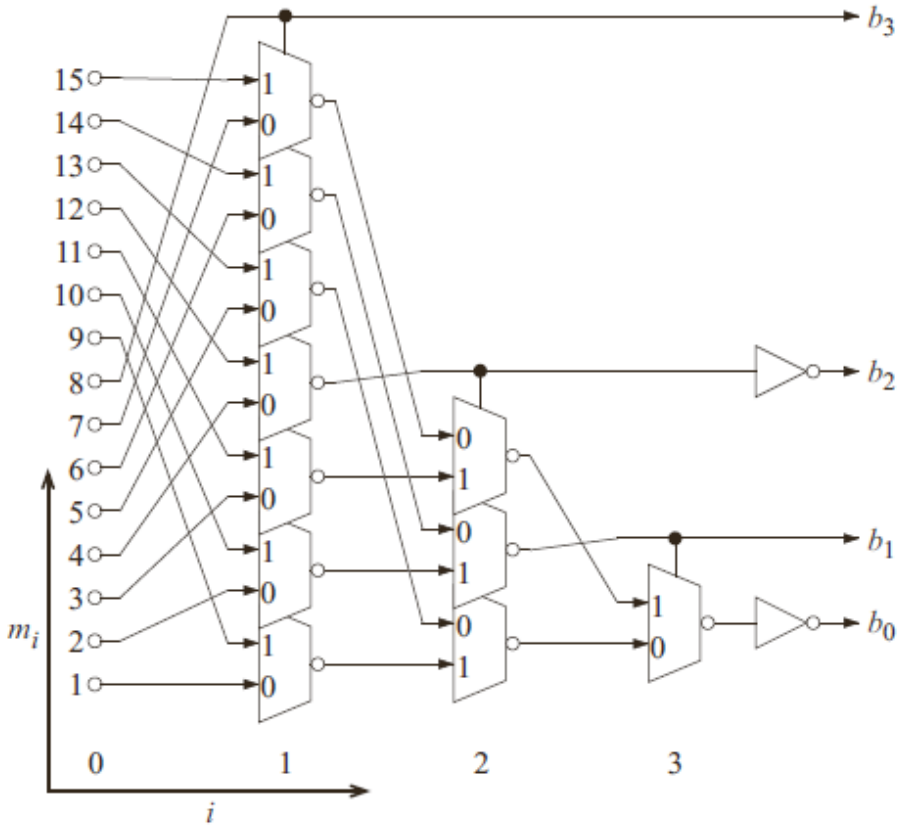


Figure 2.27 - MUX Based Thermometer-to-Binary Decoder [40]

2.6.3 **Wallace Tree Based Thermometer-to-Binary Converter**

Recalling what has been referred earlier, Thermometer Code represents a natural number N by a sequence of N ones followed by a zero, so it is like a one by one count to a certain number N. This is the main principle behind the Wallace Tree Thermometer to Binary Converter, also known as One-counter decoder. It simply counts the number of one's presented in the thermometer code. This technique uses as a basic building block, a full adder. A 4-bit decoder is presented as an example in the next Figure 2.28.

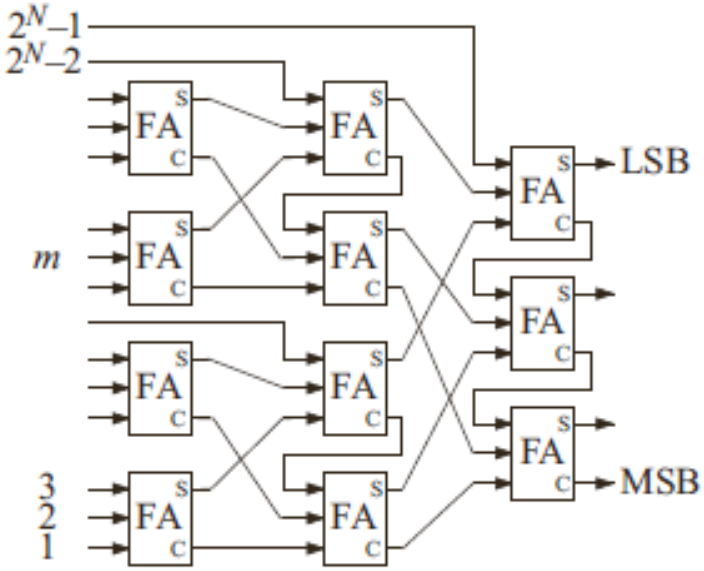


Figure 2.28 - Wallace-tree Based Thermometer-to-Binary Decoder [40]

This technique offers good error correction, and prevents the appearance of bubble errors, because the worst thing that could happen is an extra input being added or subtracted from the result. Another advantage is the high speed that this circuit can achieve, mainly because of its basic principle of operation.

However, it has a huge area and power consumption based on the number of transistors needed to build a full adder. There are some techniques used to improve the consumption of the system based on splitting the inputs of the thermometer code into multiple sum intervals, with the help of multiplexers, introducing them to a smaller Wallace tree. This will reduce drastically the size and power consumption of the circuit, increasing just by a small amount the delay operation, although an extra effort is needed to control the behaviour and operation of the system. Hence, it will end up with a trade-off between speed and power/area consumption that needs to be analysed and optimised based on our needs. [38,40,46].

2.6.4 **Fat-Tree based Thermometer to Binary Converter**

The fat tree-based decoder is the simplest structure to achieve this conversion. As well as the last topology, this technique can be used either to produce direct conversion from the thermometer code to a binary word or to produce a grey based code for posterior analyses.

Simple OR gates are the main ingredient to its structure, and by checking its topology, it is easily seen that there are always 3 OR gates from any leaf node to any root node. This makes it a very uniform circuit, enabling high speed operations.

The thermometer-to-binary is carried out in two stages, the first one is always the same, regardless of the output code type. It converts the thermometer code to a 1-of-N code. The fact that this type of technique needs this intermediate step makes this architecture automatically noise tolerant and bubble error resistant. The second and final stage converts the last code to a binary word, using multiple tree of *OR* gates, as is shown in Figure 2.29. [44]

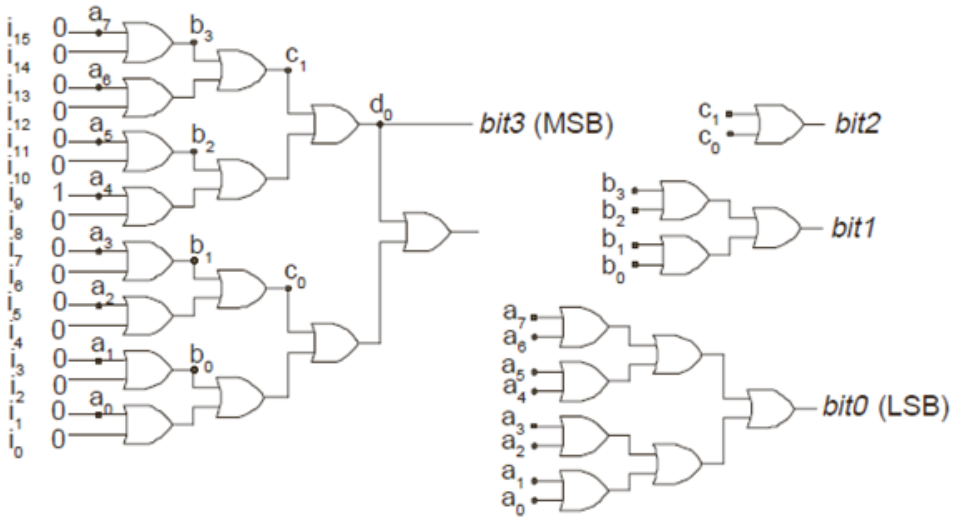


Figure 2.29 - Fat-tree Based Thermometer-to-Binary Decoder [44]

The simplicity behind the concept of this circuit is what makes it so fast, and less power and area hungry than other solutions. To improve power consumption CMOS implementation of the OR gates eliminates any static power consumption. [46] However, its layout design is a bit challenging mainly because of its tree structure.

2.6.5 **Thermometer Decoding Techniques Summary**

Nowadays, Thermometer-to-binary decoding has become one of the most important parts in any type of ADC, especially flash ADCs where speed and accuracy represent a key performance specification. It is useless to have very high-end ADCs that produce high resolution results if the decoding circuit structure present errors in the final code, or if it is too slow to keep up with the ADC. In addition, power and area consumption must be taken into account in the choice between different architectures.

Some of these techniques have already been presented, where the ROM and Multiplexer based decoders are the most common approaches to this end. ROM techniques are the simplest ones, compared with all others mainly because it is a Programmable Logic circuit, although a slow, high power and area consuming technique.

MUX based decoders on the other hand, are known for high speed and low power operation, presenting also low vulnerability to bubble errors, which in high speed ADCs become key characteristics. However,

they are lacking in the high non-equal input charge introduced in the outputs of the ADC itself.

Another method is based on a Wallace tree concept, one of the first decoding concepts used, working on a tree architecture using Full adder cells. Compared to the previous techniques it presents a very low vulnerability to error based on its concept, but if it is intended a low power/area circuit it will not be the best option.

Finally, a Fat Tree concept is presented, which initially converts the thermometer code to 1 of N code, reducing drastically the probability of bubble errors, and then to binary code using a tree architecture with basic logic gates. Compared to the other proposed techniques it uses more transistors to produce the same binary code, so it may be supposed that it has bigger area/power consumption and lower speed performance. Actually, it happens to be the other way around, considering that this circuit is based on basic logic gates. The way they are connected enables very low area consumption, also this circuit results from logic equations specially made with this in view, resulting in high speed performance. One other point in its favour is the fact that this technique possesses an equal input charge in all nodes, contributing to the linear response of the system.

Table 2-3 – Comparison study on Thermometer-to-Binary Decoder

	TECHNIQUE USED							
	ROM Decoder		**MUX Decoder**		**Wallace Tree Decoder**		**Fat-Tree Decoder**	
REF.	[46]	[44]	[46]	[40]	[46]	[40]	[46]	[44]
Technology [nm]	180	180	180	130	180	130	180	180
Nº of Bits	5	6	5	6	5	6	5	6
AREA [mm^2] \ Nº Devices	-/215	0,0094 \ 222	-/108	0,4 \ -	-/572	0,7 \ -	258	0,0094 \ 485
Speed [GSPS]	-	1,11	-	-	-	-	-	2
Power [W]	233,4u	54,41m	254u	182m	790u	175m	134,7u	38,65m

3. TIME-TO-DIGITAL CONVERTER ARCHITECTURE

In Chapter 2, the principal techniques that cover the Time-to-Digital conversion were reviewed. Although the conversion was not complete, the output of the TDC is not a binary code. To convert the obtained code into a binary word, a Thermometer-to-Binary Converter technique will be used based in the ones that were reviewed in before

Therefore, in this chapter, an overall view of the global system is going to be made, describing each block used. Then each system block is going to be exploited individually starting with the Time-to-Digital Converter itself, the Thermometer Decoder and its sampling system which is responsible for obtaining the output code from the TDC itself and converting it into a binary word.

3.1 System Overview

Before starting to show and explain the overall system, there is an expression that has been taken into account in all the work developed in this thesis “Keep it Simple”, meaning that the system built will only use the most indispensable and essential components to make the circuit work properly, taking it to the transistor level. In other words, if there is a transistor that may even be part of the circuit as it is known but is not actually necessary for the purpose for which this circuit is intended, it will be automatically excluded.

In the end, the result target will be a simpler and direct circuit, which will be translated in low area and power consumption. In the future, this could extremely useful, because it will easily allow new functions just by adding or excluding one simple system block.

The Following Figure 3.1, presents the proposed system Overview Architecture. The system is composed of 6 basic blocks, however two of them (Control Logic and Thermometer-to-Binary Decoder) cannot work independently as it is going to be seen later on.

- **Coupling Circuit** – This simple block has three objectives, deliver a clean and strong Pulse to the TDC delay line, close the delay-line loop, and reset the delay-line. The Input and Reset Pulse comes from an outside source connected to the system.

- **Delay-Line** – “System heart”, containing the delay-line itself. This is responsible for the time measure, and the delivery of the time code to the thermometer converter.

- **Loop Counter** - Responsible for counting the number of loops that the pulse gives to the delay chain, delivering a binary output code which represents the course time measurement.

- **Control Logic** – The “System Brain” is responsible for the control and command of the system, from the insertion of the pulse into the chain, to its reset.

- **Thermometer-to-Binary Decoder –** Delivers the fine measurement of the TDC. It is responsible for the acquisition of the output code from the delay-line and converting it into a final binary word. As it was said before, it has a high symbiotic relationship with the Control Logic, being responsible for the accurate functioning of the control block.

- **Registers –** Last block of the proposed TDC architecture. Like the name suggest this clock is responsible to save the result of the measurement performed and to deliver it to a posterior circuit. Additionally, this circuit prevents data lost, due to its regenerative topology.

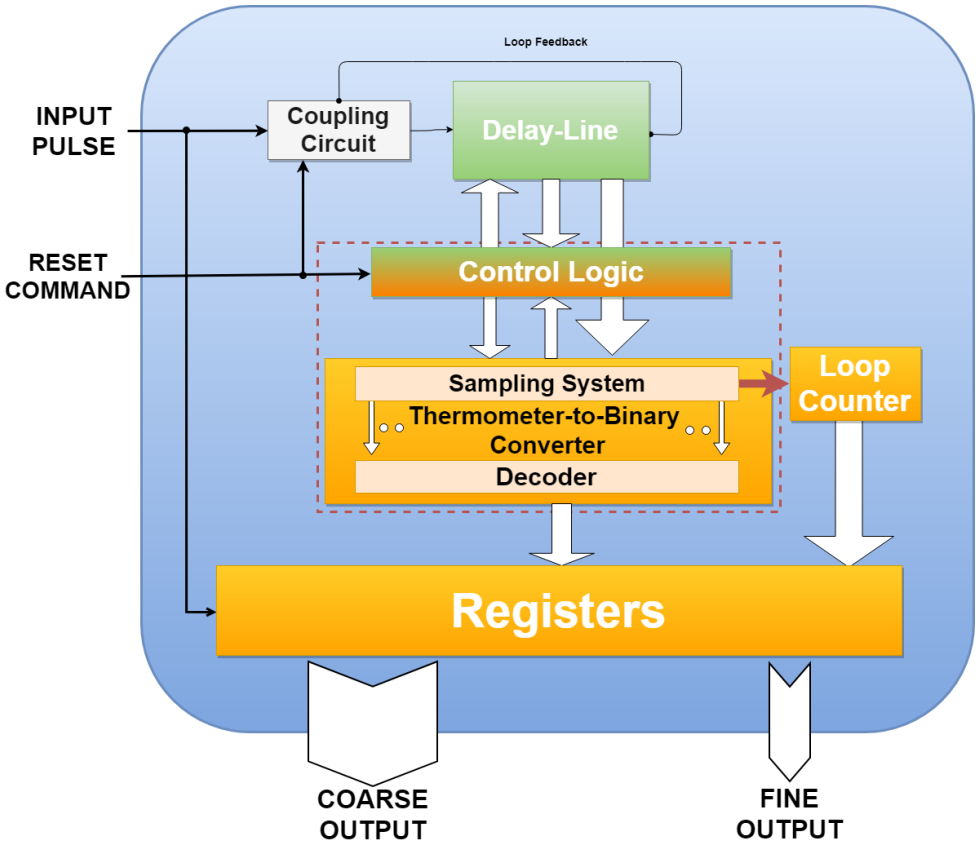


Figure 3.1 - System Overview Architecture

3.2 Coupling Circuit

This circuit, looking to the system's description, represents the "Gateway" of the system receiving the input pulse from an outside source or from the end of the chain and (re)introducing it to the TDC delay-line. Additionally, it is also responsible for the reset of all the delay-line, cleaning and preparing it to a new measurement.

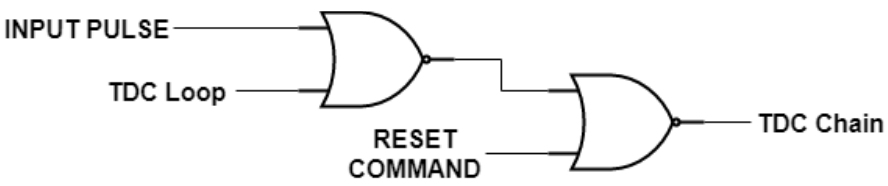


Figure 3.2 – Coupling Circuit

The circuit is based on two NOR gates, connected in series, as shown in Figure 3.2. The measurement pulse is introduced in the beginning of the delay-line if there is no reset command active.

The input is not directly introduced into the delay-line, mainly because it may not have “enough strength” to be introduced without suffer any voltage drop, which could lead to data loss. Therefore, to avoid that, the pulse is regenerated. Another advantage of this measure is that it excludes any possible noise that the input source could bring.

Although this circuit is the best and simplest way to insert a pulse into the delay-line, and also to clean it, the fact that this circuit is powered by NOR gates and the delay-line in based on other different gates, as it is going to be explored in section 3.3.1, it could lead to a linearity problem.

3.3 **Delay-Line**

This section will describe each step to carefully design the Time-to-Digital Converter delay-line, developed in this thesis. As was referred to before as the “System Heart”, all the system’s performance depends mainly on the behaviour of this block, so an extra effort was made to achieve the best performance behaviour possible. The main requirement used in this system’s design was “Keep it simple”, so a new improved technique will be proposed based on the TDC techniques that were studied so far.

Firstly, the intention is that the architecture should achieve high resolution, hence automatically sub-gate resolution technique will be chosen. Considering the study that was made in the previous chapter, three techniques were considered, i.e. Parallel scaled delay-line, Vernier delay-line and Pulse-Shrinking delay-line.

Parallel scaled delay-line presents the simplest way to get high resolution performances. It is simply based on conventional balanced buffers in series with capacitors. However, considering the high area consumption, due to the use of big capacitors, high time offset, and the difficulties encountered in proposing a balanced layout, this technique is not the best option when compared to the other two options.

Vernier on the other hand is widely used mainly because of its simple concept, and high-resolution capability. Similar to the previous one, Vernier is based on simple buffers, but in this case, they are distributed in a 2-parallel delay line, where each independent delay-line possesses different element sizes leading to different propagation times. With a considerable improvement in the linear response, due to a much lower offset, Vernier really takes the lead comparing it to the previous one. Nevertheless, it needs a dual delay line, meaning that an increased control effort needs to be applied in order to make proper use of the technology. Another aspect is that the buffers need to be perfectly balanced, so the travelling pulse will not suffer a variation, and if PVT variations are considered, that will be quite difficult. That is why sometimes in order to make sure that the pulse does not affect the measurement a clean pulse is created every single loop.

That is where Pulse-shrinking has the advantage, because unlike other techniques that rely on perfectly balanced delay elements, pulse-shrinking TDC makes use of the unbalanced delay elements, in a controlled way, to perform extremely high resolutions. The natural process variations that are in fact unwanted effects in other techniques, here they can be used as an advantage and be finally profitable. It is not suggested that the variation effects do not disturb the pulse-shrinking working procedure, because like in any other system, they will. However, if they are unavoidable, perhaps the system could possible take advantage of it.

Pulse-shrinking could also be seen as an improved version of a basic delay-line based in a Vernier concept.It uses the concept of a basic delay-line, meaning it has the advantage of a simple architecture with almost the same structure taking it through to the sub-gate resolution. Thus it uses much less area with a simpler control than the Vernier ones, and instead of using the time difference between the propagation time of two different elements, it uses the time difference but in this case between the rising and falling times of the same delay element, thus preventing the possible mismatch between the two delay-lines used in the Vernier TDCs method.

3.3.1 **Pulse-Shrinking Topology**

Based on the evaluation presented so far, it has been decided that in this thesis a Looped Pulse-shrinking TDC will be developed, as it fulfils all the requirements that are desired, namely very high resolution, low area and power consumption, simple design and implementation. Additionally, the proposed technique will try to fill a gap existing in the traditional pulse-shrinking TDC techniques.

3.3.1.1 Delay Element Sizing

Pulse-shrinking TDCs are based on buffers delay-lines designs [47,48] , and buffers are built with two inverters in series. So, that is the first step, a looped delay-line will be proposed, where the delay elements will be inverters instead of the traditional buffers. This will permit a decrease in the area and power consumption, and an increase in the maximum resolution that this topology achieves.

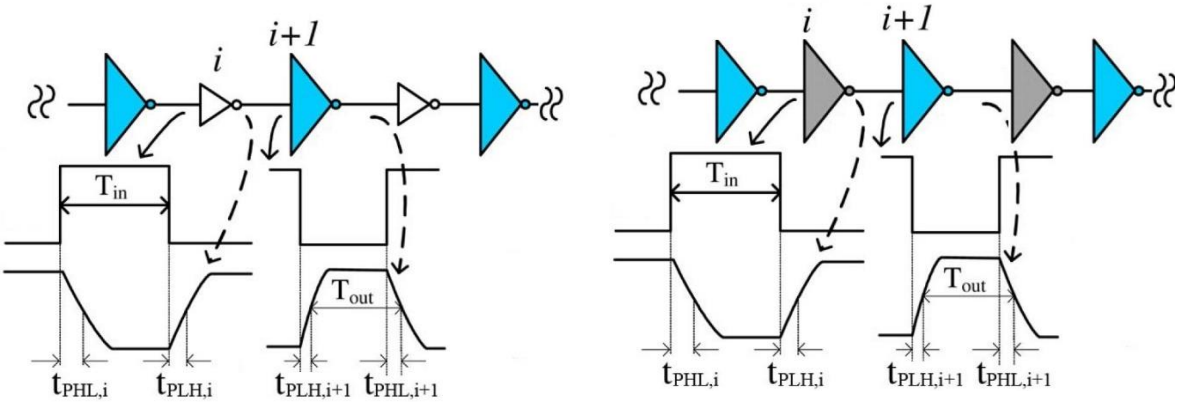


Figure 3.3 – Common Pulse-Shrinking Delay-line (a), Proposed Pulse-Shrinking Delay-line (b)

In order to have a better understanding of the basic concept behind the proposed architecture, let's take another look at equation (2.29), rewriting it in a simpler and understandable way,

$$\Delta\tau = (t_{PLH,i} - t_{PHL,i}) - (t_{PLH,i+1} - t_{PHL,i+1}) \tag{3.1}$$

The last equation translates the time-resolution behind the traditional delay-line, Figure 3.3(a). As referred before, it is the sum of the cumulative effects that both inverters in series produce, therefore the effect produced by one buffer.

Hence, what is proposed in this thesis is to divide the equation into 2 parcels, meaning that every single inverter will shrink the travelling pulse, doubling at least the resolution of the delay-line and decreasing the conversion time take to get the result. Following this idea and merging it with Figure 3.3(b), the following equations translate the working principle behind this concept,

$$\Delta\tau_i = t_{PLH,i} - t_{PHL,i} \tag{3.2}$$

$$\Delta\tau_{i+1} = t_{PHL,i+1} - t_{PLH,i+1} \tag{3.3}$$

Taking a closer look at both equations, they look symmetrical. However, for symmetrical rising and falling delays or for asymmetrical but equal inverts, the pulse width will not change; the shrink will be null. Considering the first equation (3.2), matching the *i,* delay elements in Figure 3.3(b), it is concluded that to produce a shrink at the input travelling pulse, it must be ascertained that $\tau_{PLH,i} < \tau_{PHL,i}$. In the second equation, which matches the *i+1* delay element, is exactly the opposite. To produce a shrinking element the aim will be $\tau_{PLH,i+1} > \tau_{PHL,i+1}$. Although, in both cases the closer these parameters are, the higher the time resolution obtained, always limited by PVT variations.

The following test bench will be used to optimize as much as possible the last parameters. Still, to make the dimensioning process easier, the last two equations will be introduced in the test, directly searching the parameters which present the best performance. This circuit is built within the concept of the real charge with which the inverters will work in the real delay-line.

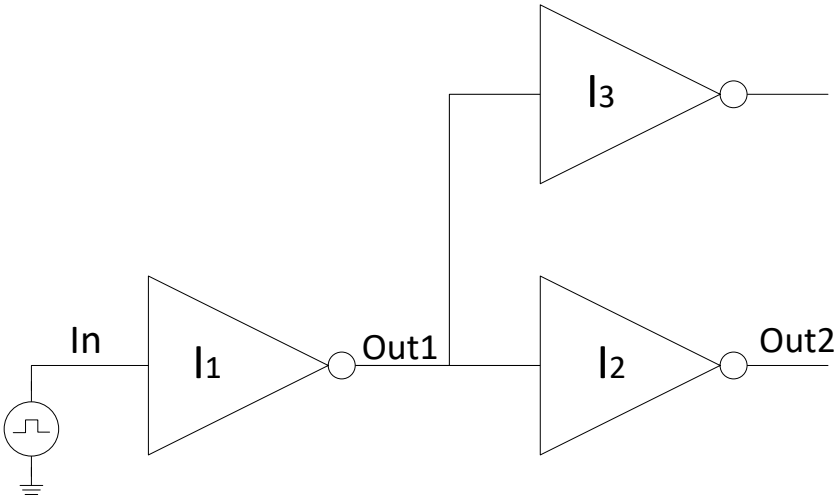
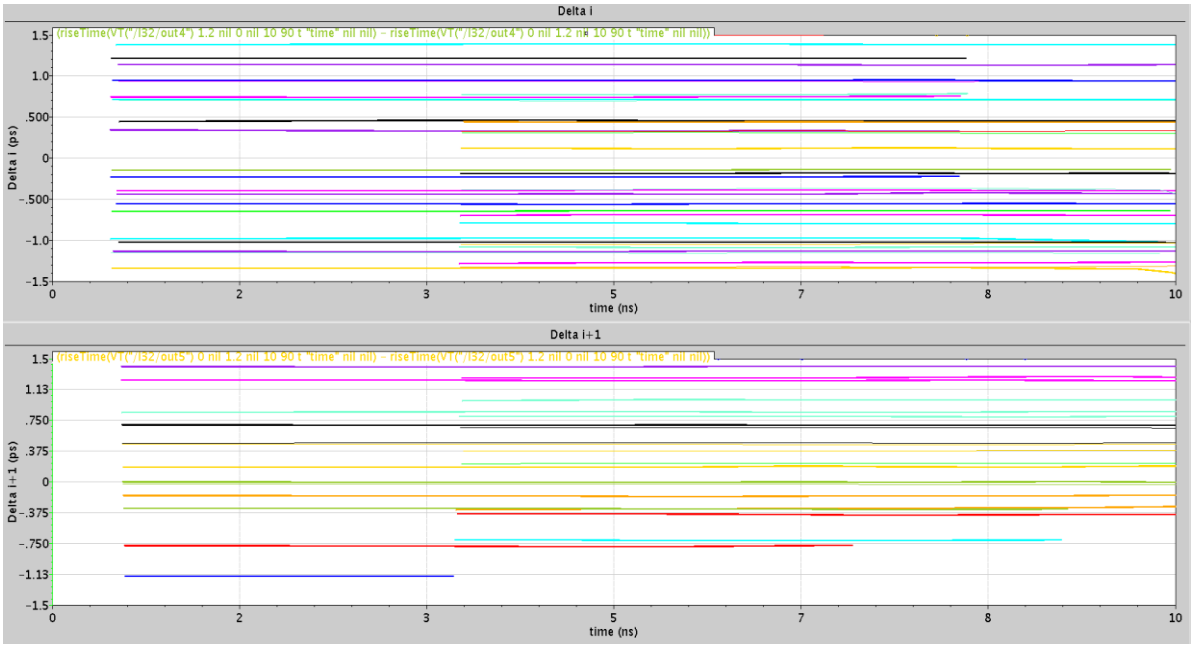


Figure 3.4 - Inverter Sizing Test Bench

The presented test bench consists of 3 inverts, instances I0:I3, representing one portion of the real TDC delay-line where the test pulse, generated via a voltage source named V0, travels through the line represented by the inverts I1 and I2. The inverter I3 will only be used by the sampling circuit, as it is going to be shown later, having no implication in the measurement other than in the load introduced into the output node of the inverter I1. However, and for the intended design, this inverter will have exactly the same dimensions as the inverter I2, following the delay-line concept proposed in Figure 3.3(b). There is another reason for the use of the inverter *I3* rather than directly connecting the acquisition circuit to the delay-line, as with this the delay-line is isolated from the rest of the circuit, preventing any possible disturbance introduced by the preceding circuits. Additionally, it will improve the linearity of the delay-line because the input charge in each node will be the same, equal to the delay elements themselves, which reduces the input charge, when compared to the sampling elements, thus improving the resolution.

Notice that for both cases the dimensioning will be performed entirely in the inverter I1. According to the analysis that was performed in the inverter in chapter 2, the dimensioning will start with a swipe around the widths of the 4 transistors intending to look at all the results and choosing the one that fulfils the desired requirement. Although, for safety measures the transistors cannot have width parameters lower than 700n, a guarantee of 2 connectors at each side of the transistor will prevent its failure, if for some reason one of them does not connect properly.



(a) (b)

Figure 3.5 - Pulse Shrinking Inverter i (a), Pulse Shrinking Inverter i+1 (b)

The graphics lastly presented, obtained from the swipe test, translate a selection from the achieved pulse shrink that each i, and i+1 inverter, Figure 3.5(a) and (b) obtain respectively, with each transistor size combination.

Taking a first look at the graphics, it seems likely that it could easily achieve resolutions closer to 200fs. However, crossing the measures between both graphics invalidates the previous statement. That is because certain parameters combination could generate a high resolution at the *i* transistors. But looking at its precedents, the resolution achieved is much lower. The main reason for this behaviour is the mutual influence that the two types of inverters have on each other.

Therefore, and taking into account the required relationships between rise and fall times, for the technology used the maximum resolution achievable will be 0.82ps, with the following transistor sizes:

Table 3.1 - Delay-Line Transistor sizes

	i	*i+1*
$\mathbf{W_p/L_p}$	2,4u/120n	2,4u/120n
$\mathbf{W_n/L_n}$	700n/120n	740n/120n
	m	

3.3.1.2 Delay-line Loop Sizing

Lastly but not least, it is still necessary to define the length of the delay-line. The length of the delay-line will define the maximum time measurement that it will be able to handle. This length will be restricted both by its requirements, and by the propagation delay of each element. However, due to the final binary code representation and to maximize the dynamic-range system performance, it should also be guaranteed that the number of delay-elements obtained is a 2^n multiple.

With the transistor sizes already defined and looking to the waveform at the input and output of the inverter, waveform *outi*, *outi+1* respectively from Figure 3.6, and according to that which referred to in chapter 2 it is easy to obtain the propagation-delay of each element, which in this case will be 22,8ps.

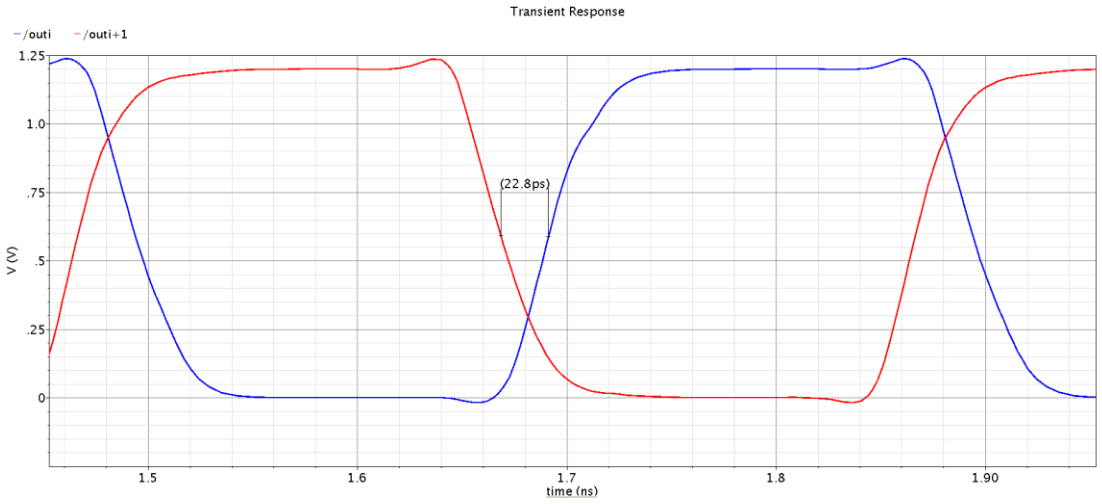


Figure 3.6 - Transient Response, Propagation-delay

According to the system requirements, the desired Dynamic range would be at least *2ns*, so with the obtained propagation delay, it would need 87,72→88 delay elements. However, as it should be a multiple of 2^n, the Pulse-shrinking Looped delay-line will have 128 elements, leading to a maximum Dynamic range of 2,918ns.

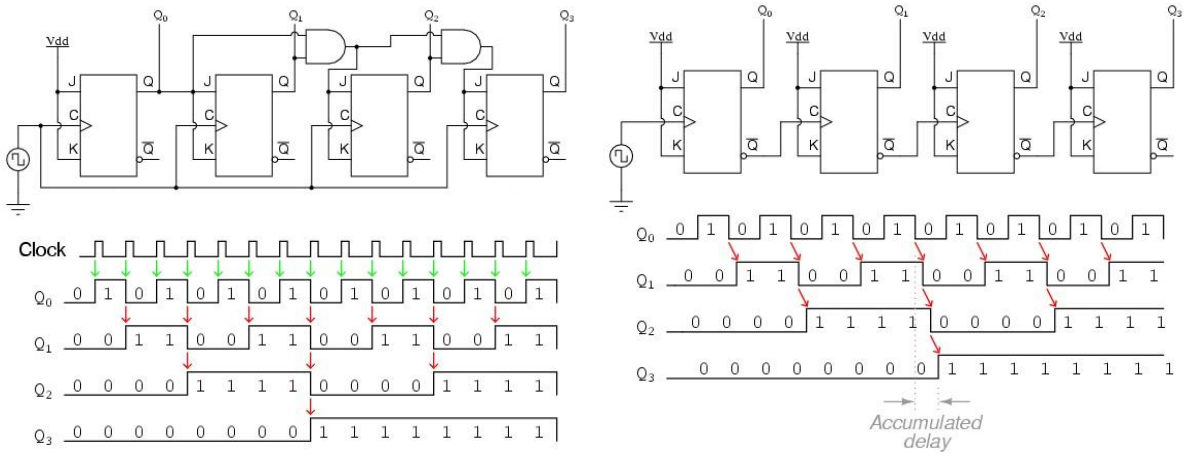
3.4 **Loop Counter**

According to the loop concept, there is a final external stage installed after the delay line that closes the circle. This element is the loop counter enabling the count of the number of loops that the pulse has made into the delay-line. According to the study realized in the previous chapter, there are 3 main reasons to consider in the choice of the loop technique, area, power reduction, and non-linearity due to PVT variations. [2]

3.4.1 **Counter Topologies**

A counter can be constructed following one of two topologies: asynchronous or synchronous, being distinguished by their assertion with the input clock. In both cases the basic component in their constitution will be the Flip-flop.

In asynchronous counters, also known as ripple counters, the clock only triggers the first flip-flop in the "chain". Each successive flip-flop derives its own clock input from its direct antecedent. Therefore, the output bit of each flip-flop will appear slightly delayed, reducing the operation speed due to the delay accumulation through the flip-flop chain. This will be problematic in high speed circuits with a high number of bits, because it could lead to errors if for some reason the counter is sampled before assuming the new stage. [49,50]



(a) (b)

Figure 3.7 - Synchronous Counter (a), Asynchronous Counter (b) [50,51]

Synchronous counters on the other hand change all the output bits at the same time, that is because all the flip-flops are connected to the same input clock pulse.

The most important advantage of synchronous counters is that there is no cumulative time delay because all flip-flops are triggered in parallel. Consequently, its operation frequency will be slightly higher than the asynchronous counter, avoiding possible counting errors.

However, compared to an asynchronous type it takes up a much larger area, it needs some extra components to work; slightly more power and presents higher layout difficulties, because it must be

guaranteed that all the flip-flops present will have much the same propagation time, even with PVT variations, and the clock arrives to all flip-flops at the same time, otherwise it will have the same problems as the asynchronous ones. That is why these systems sometimes use something called a clock tree, to guarantee that the clock signal arrives at the same time to the input of all flip-flops, leading to a higher area and power consumption. [49,51]

3.4.2 **Asynchronous Counter Technique**

With the study presented, it was decided to develop an Asynchronous Counter. It has all the requirements: simplicity, (always remember "Keep it simple") low area and power consumption and simple design/implementation.

Its bottleneck is obviously the maximum speed that the circuit could operate without eventual code errors. Although looking at the TDC loop architecture, each clock pulse occurs approximately every *2,918ns*, leading to maximum operation frequency of *0,3427 GHz*. Therefore, the operation speed achieved by this counter, will be more than enough to fulfil the requirements.

3.4.2.1 Counter Number of Bits

The first thing that needs to be defined before building the counter will be the number of bits that it will possess, which determines the maximum number of loop cycles. As it was discussed before the TDC delay line was dimensioned to hold a *2ns* pulse, however, due to restrictions it will hold a maximum time of *2,918ns,* $\Delta t_{\max}$. Considering equation (2.19) and that in each delay stage the resolution will be approximately *0,82ps,* the maximum number of loops in the chain will be given by,

$$N_{count} = \frac{\Delta t_{\max}}{M * T_{LSB}} = \frac{2{,}918n}{128 \times 0{,}82p} = 27{,}80 \Rightarrow 28 \qquad (4)$$

where $\Delta t_{\max}$ is the maximum hold time in the line, M is the number of elements in the chain and T_{LSB} is the TDC resolution. Therefore, a *5-bit* counter will be built for this purpose. The following Figure 3.8 presents the proposed architecture for the ripple counter, based in type D flip-flops.

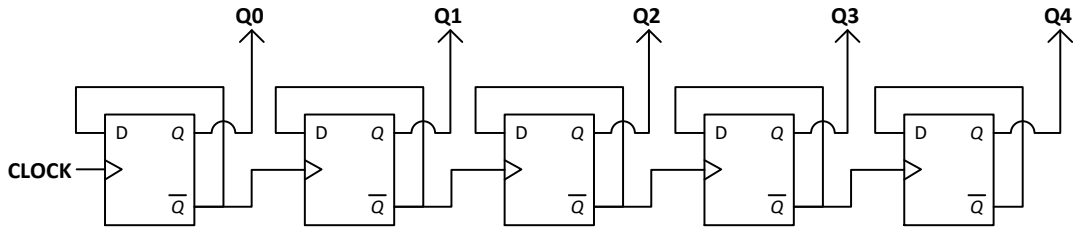


Figure 3.8 - 5-Bit Ripple Counter Proposed Architecture

3.4.2.2 Flip-Flop Design

There are many subclasses of Flip-flops, these classifications are mainly based on the behaviour of the clock and their input and output construction. However, for the sake of simplicity and based on what is intended for this circuit, only two types of flip-flops will be considered: Static and Dynamic flip-flops.

Starting with the Standard ones, Static flip-flops, represent a group of flip-flops that can preserve their stored value, even if the clock signal stops. [52] Their structure is prepared in order to self-regenerate, maintaining well defined logic levels at the decision points.

In contrast, Dynamic Flip-flops lose their stored logic information if the clock signal does not trigger the flip-flop for a while. That is because the stored logic states in this type of flip-flops is highly dependent on the intrinsic and parasitic device capacitances distributed in the nodes of the circuit, and due to the leakage currents those nodes are discharged, resulting in invalid data. Therefore, these kinds of devices are focused on high-speed operations and low-power and area consumption circuits, and possess a very simple architecture.

Analysing the proposed topologies in [52,53,54], True Single-Phase Latch (TSPC) will be the best option, it has a simple design, high speed performance, full dynamic and uses only 9 transistors to provide a well-defined flip-flop output, which means that it possesses one of the lowest areas. In addition, comparing to other topology's, it only uses 1 clock signal, removing the need for additional elements to generate a clock signal [52], like, e.g. Semi-dynamic Flip-Flop (SDFF), Hybrid latch Flip-Flop (HLFF). The following Figure 3.9 contains the proposed architecture for the TSPC topology.

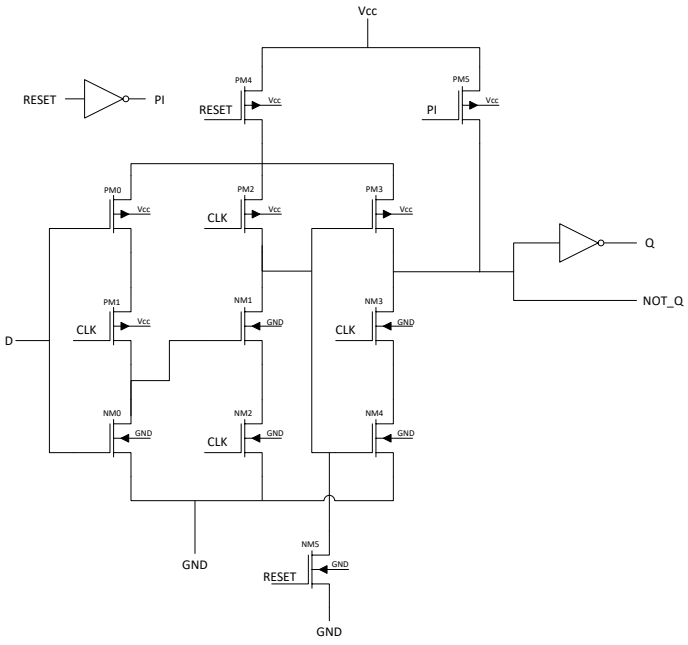
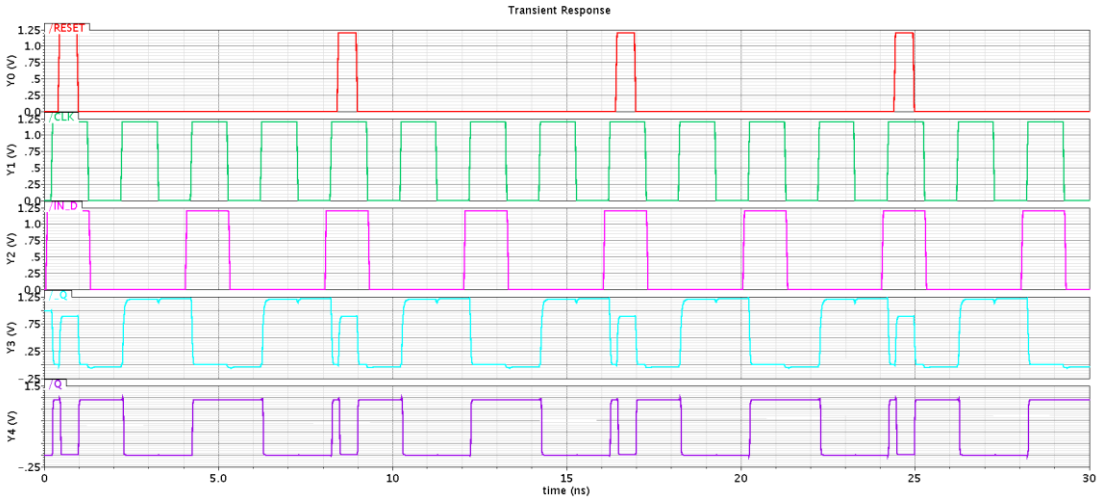


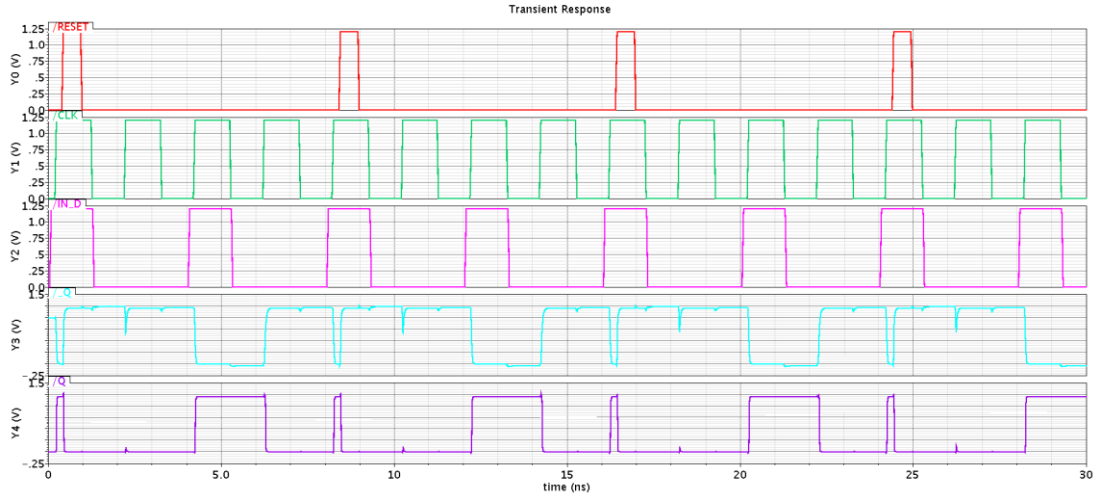
Figure 3.9 - Proposed TSPC architecture

The proposed architecture follows the basic concept presented in [52,55], although it was complemented with some extra components to provide a non-inverted output, Q, and a Reset option, RESET, provided by an outside source command. The non-inverter output was provided by the insertion of the Inverter

I2, connected to the existing inverted flip-flop output, NOT_Q. The reset function works based on the inverter I1 plus the transistors NM5, PM4 and PM5. The Reset Command arrives at the inverter I1, which generates an inverted output that is called PI that forces the inverted flip-flop output node, NOT_Q to assume a high logic level, $\overline{Q}=1$ making the non-inverted output discharge at a low logic level, $Q=0$. Although, if the clock input is in a high state, $CLK=1$, and the node X is also in a high state, causing NM4 to be ON, after the reset command disappears the NOT_Q output will return to zero and consequently the Q will go to a high logic state without a new clock rising edge. So, connected to that node the transistor NM5 has been inserted, thus forcing it to discharge, preventing the Output NOT_Q from becoming zero Figure 3.10 (a). Lastly, the transistor PM4 works as a complement to the previous ones cutting the current to the flip-flop, when the Reset signal arrives, facilitating the process and preventing short circuits.



(a)



(b)

Figure 3.10 - Transient Response Flip-Flop, with NM5 (a), without NM5 (b)

Finally, connecting the flip-flops adopting the proposed technique presented in Figure 3.6 the following transient response was obtained,

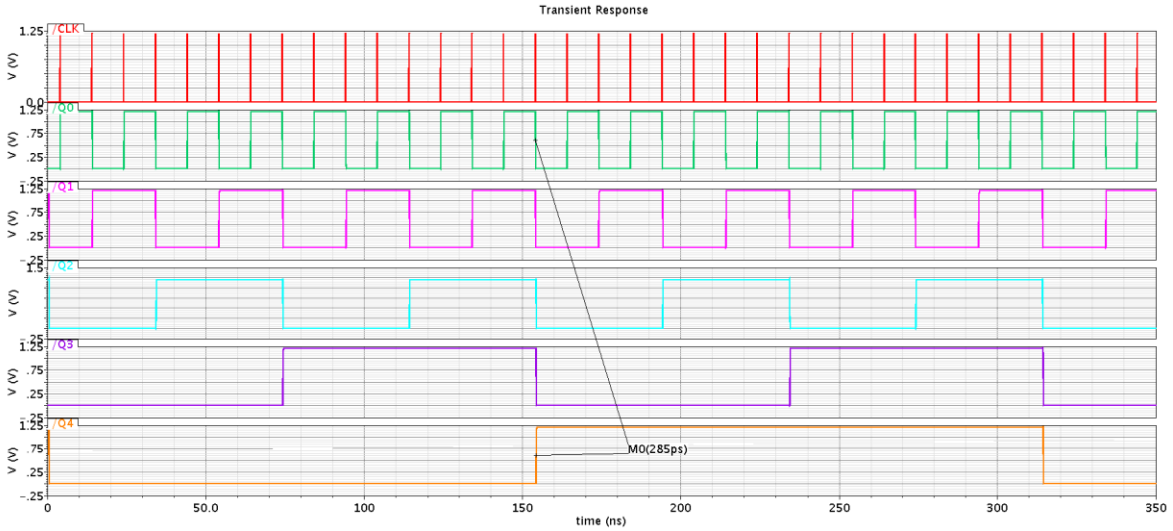


Figure 3.11 - Counter Transient Response

As it is shown, each time that a rising clock edge arrives, at the first flip-flop input, the binary output code increases one-by-one as expected. Additionally, the time taken to trigger the last flip-flop in the counter is 285ps (leading to a maximum operating frequency of 3,5087 Ghz). As it was referred before, compared to the delay-line period, 2,918 ns, 342,7 MHz, the counter projected fulfils all the requirements to perform this task.

3.5 **Control Logic**

As the System Brain, this circuit will be responsible to guarantee the proper work flow of every part of the TDC architecture, in such a way that at the end of the measurement, the converter outputs provide a proper and “error-free” result. Looking at it this way, it may seem like a very complex circuit, although, having simplicity as one of the main objectives of this work, it will end up the other way around.

This circuit will be mainly responsible for 2 tasks; cleaning the delay line thus preparing it to receive a new pulse for measurement and keeping track of the location of the travelling pulse in the delay-line. Additionally, it also helps with the prevention of bubble errors in the Thermometer-to-Binary Decoder as it will be shown later.

Starting with the first, the control circuit will receive an input pulse that will shunt the delay line, with a transmission gate to ground, cleaning it for the next input pulse. Additionally, it will reset all the sampling elements in the system. This command pulse will be known as “Master Reset” because in addition to resetting the whole all the TDC system, it can be used at any-time, and for any necessary reason, even during a measuring process, obviously losing any related data.

The TDC delay-line is built in a looped structure concept, therefore it must be guaranteed that the pulse correct position is kept in track so that at the end of the present measurement the last recorded position could be added to the number of total loops. Following this proposition, the second task of the control system will be to clean the sampling elements that the travelling pulse has already left behind, making sure that it is already sampled by the two following units. As it is an inverter delay line there are two

different conditions, to guarantee that the pulse has completely passed through the element. The reason why it uses the following two sampling elements instead of just one to clean the present unit, is to guarantee that the sampling elements have enough time to settle into a well-defined high state due to the low delay that each element in the TDC line possesses and also preventing possible PVT variations.

Combining the previous propositions and translating both into logical equations, one for each case, the following equations are obtained,

$$\overline{Y} = \overline{A.B.\overline{C} + D} \tag{3.5}$$

$$\overline{Y} = \overline{A.B.C + D} \tag{3.6}$$

Where, *A* and *B* refer to the two following stages, C and $\overline{C}$ refer to input clock signal acquired from the delay-line and D is the Master Reset of the circuit. When the travelling pulse at the delay line is translated into a high state, a falling edge at its output means that it has completely left the element, therefore equation (3.5) represents all these cases in the element chain. When the travelling pulse is translated into a low state, the opposite occurs, a rising edge at the output represents that the pulse has completely left the element; therefore equation (3.6) will represent all these occurrences in the delay-line.

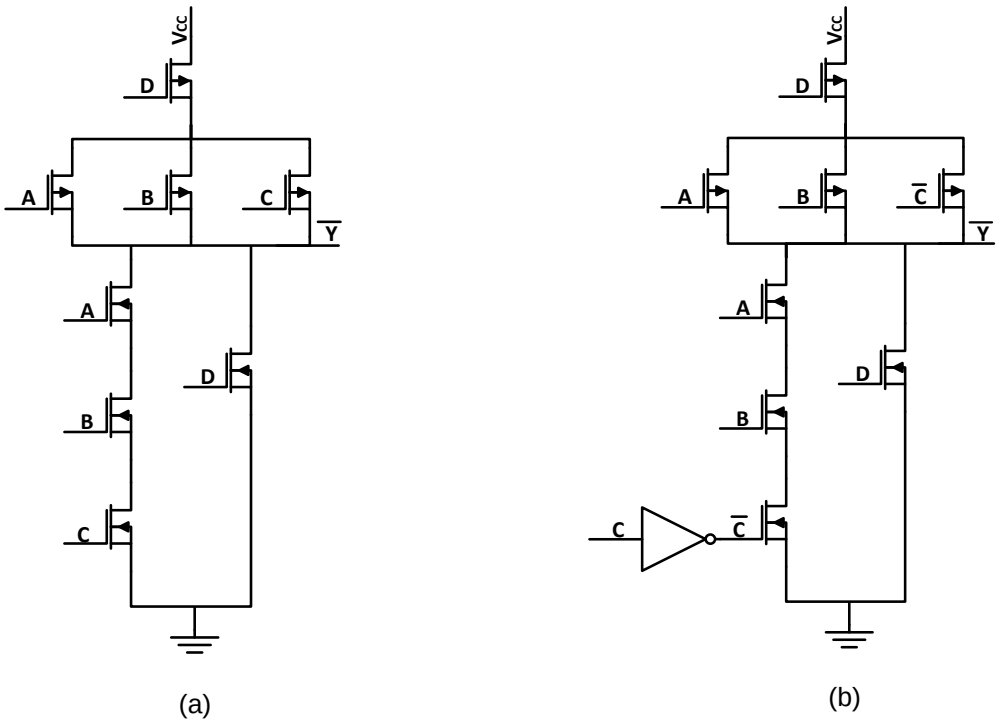


Figure 3.12 - Control Logic Circuit Equation 3.5 (a), Control Logic Circuit Equation 3.6 (b)

The last Figure 3.12 shows the proposed architecture of both circuits. The only difference between them is the additional inverter due to the inverted input from equation (3.6). The following Figure 3.13 contains the Transient response for the control system.

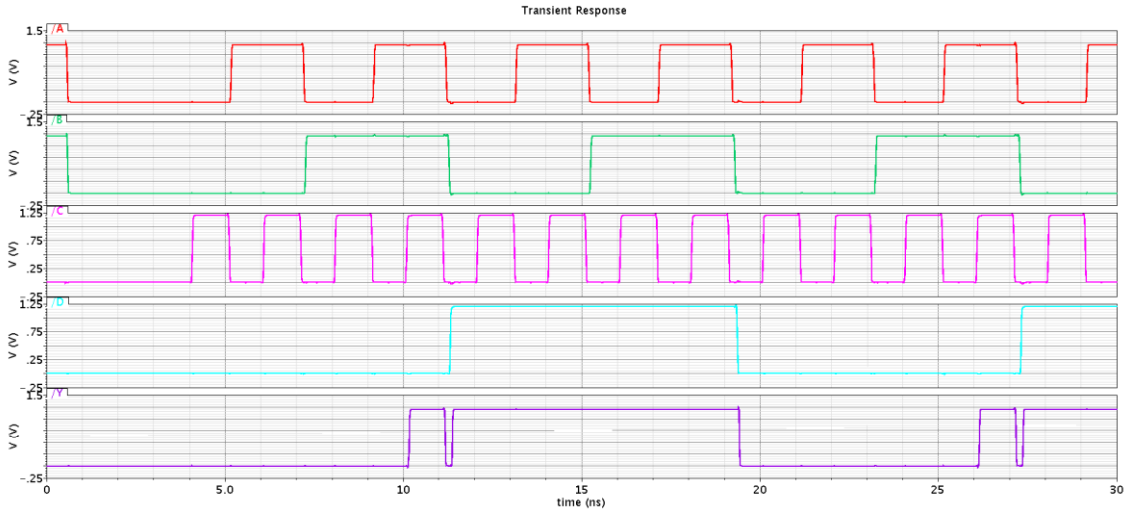


Figure 3.13 - Transient Response for the Control Circuit

Looking at the circuit response, it can be seen that the only time that the output is in a high state is when it gets a Master Reset command, or when it already has the pulse in the two next stages and has left the present stage.

3.6 Thermometer-to-Binary Decoder

Thermometer-to-Binary Decoder (TC-to-BC) is responsible for the sampling of the TDC delay-line and delivery of an “error-free”, fine resolution code. In chapter 2.6, some of the main decoder architectures were presented. From all the proposed techniques, there were two architectures that stand out especially for their high-speed conversion, low area and power consumption, Multiplexer based decoder and Fat-tree Decoder.

Multiplexer decoder would be the best option comparing it directly with the Fat-tree design. It presents a lower area and power consumption, and its layout is less time consuming and therefore simpler to build [41,43,46]. However, considering that the final circuit is a TDC, the chosen decoding architecture will be the Fat-tree design, mainly due to one crucial factor, i.e. the input charge to the circuit. Comparing both topologies, Figure 2.27 and Figure 2.29, fat-tree design has a more balanced/lower input charge, exactly the same structure and charge through all the input nodes, and the delay from the inputs to the outputs are uniform [44] . It may seem unimportant, since the decoder is not connected to the TDC line. But it is, even more due to the use of dynamic flip-flops.

3.6.1 Sampling System

The First thing to be done in the decoding will be acquiring the output sample from the TDC delay-line. To do so, an inverter in series with a complementary circuit, based on the structure of a flip-flop, will be used. The inverter, introduced at section 3.3.1.1 (named as I3, Figure 3.3), provides the “input clock” trigger to the sampling element and establishes the same input charge for all the elements in the delay-line because it uses the same transistor sizes, and compared to the sampling elements it has a lower charge. Additionally, it protects the circuit from any disturbance that may be felt in the precedent elements.

Relatively to the sampling elements, their structure is based in a flip-flop structure, as mentioned before. Therefore, and for the same reasons that were introduced in the counter structure, plus the fact that it has a very low hold time, TSPC flip-flop will be the chosen technique.

Referring back, it has been mentioned that a balanced low input charge has a huge impact in the flip-flops structure. Taking the multiplexer decoder as an example, it has a highly different input charge in some of the nodes. So, higher input charge, leads to lower operation speed due to the longer setup time, hence, the flip-flops will need higher drive capability to complement it, by increasing the transistors widths. With the increase in the size of the transistors, the clock input driver capability will also have to increase. Consequently, the isolation transistors (I3 from Figure 3.3) will increase their size, which will lead to higher load capacitance in each node of the TDC delay chain, reducing the maximum achievable resolution. Additionally, the TDC needs an equal structure throughout the entire system to promote a linearity response, leading to an increased area and power consumption. So, the lower and more balanced the input charge is, the better the system will perform.

Since it is an inverter-based chain, a negative edge triggered flip-flop is also needed to sample the elements where the pulse position is determined by a low state. Hence, based on the Positive edge already presented in Figure 3.9, the proposed falling-edge TSPC flip-flop is shown in Figure 3.14. Comparing both structures, including the reset design, it is concluded that the only things that had changed were the nodes and the transistors (NM0 and PM4) directly related with the clock input signal. Additionally, looking at the transistors PM1 and PM2, they switched positions. The clock transistor was initially at the top of the structure, due to the charge coupling problem that the node suffers, which in the worst case could prevent the transistor from driving.

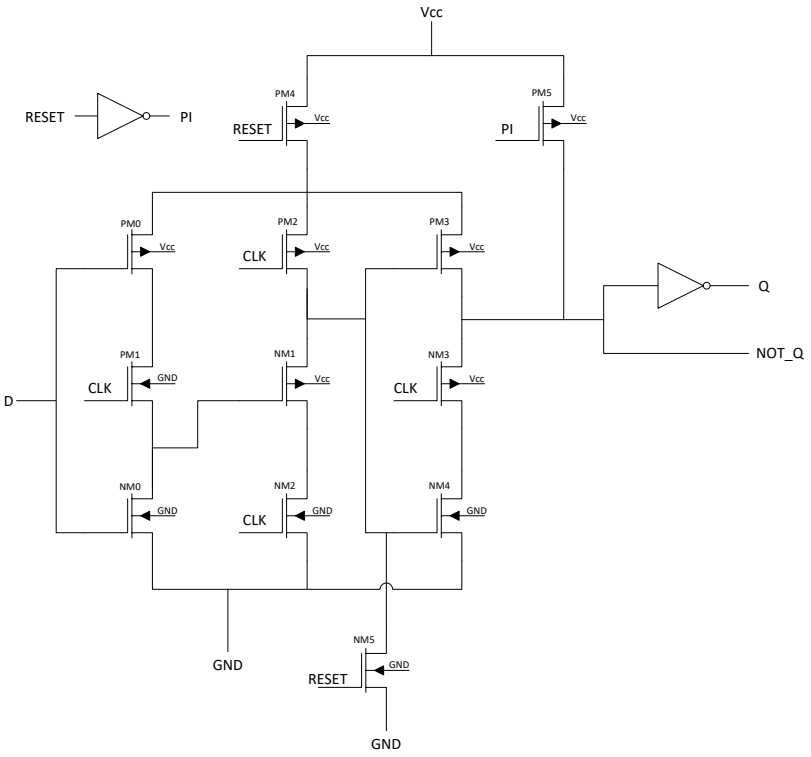


Figure 3.14 - Proposed Falling-Edge TSPC Flip-Flop

The acquisition circuit function will be to sample a high state output each time that the clock's input is triggered and resetting it each time that there is a command at the reset input from the control logic. Thus, indicates that the pulse already passed the element in the delay-line, thus keeping track of the pulse position in the chain.

Therefore, to achieve a high state from the flip-flop non-inverted output, Q, each time the clock is triggered our input data, D, will be connected to Vcc presenting a high logic level. But if the D input is connected to ground, low logic level, it could have the same high-logic level with the inverted flip-flop output. Hence, the final inverter could be excluded from the circuit. It is still necessary to rearrange reset method used, since it needs to force the node to ground rather than to Vcc.

The next steps are based on the analysis of the functional concept behind the rising edge flip-flop.

Since D input is always to be connected to ground, PM0 will be always be closed and NM0 will always be open, therefore from the moment that CLK once turns to zero, NM1 will possess a high logic level at its gate independently of future clock transitions, meaning that it will always be closed.

Since it is a rising edge flip-flop it is important that the gate at PM3 becomes connected to ground, via the transistor NM2, in order to become active pushing the inverted output to VCC. Therefore, PM2 seems to have no use in this circuit due to NM3. However, it can be used to reset the output from the new acquisition circuit, connecting its gate to the output from the Reset circuit, *PI*.

Taking a similar analysis in the falling-edge flip-flop, adjusting it to its reality, the two-new acquisition circuits will be presented in the following Figure 3.15.

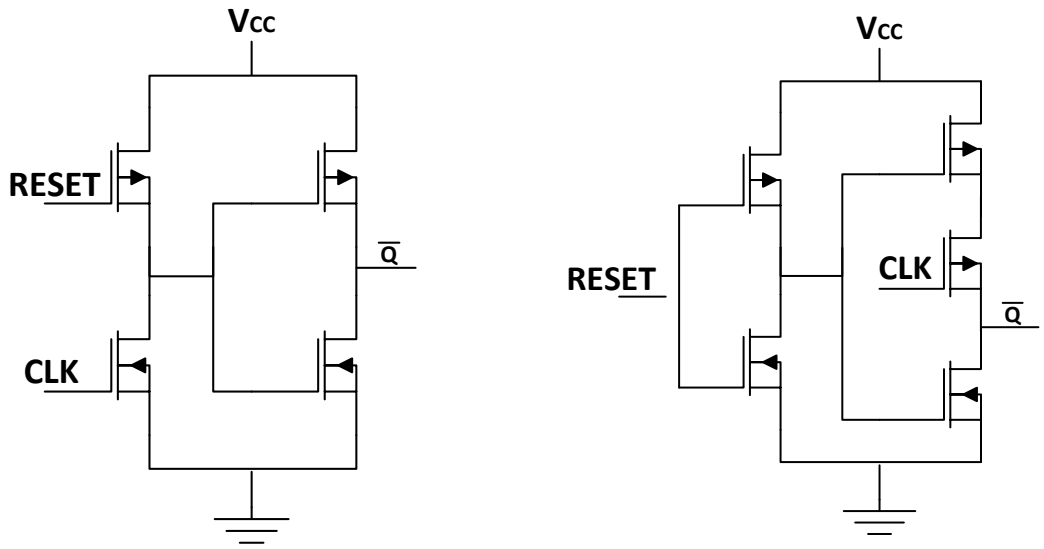


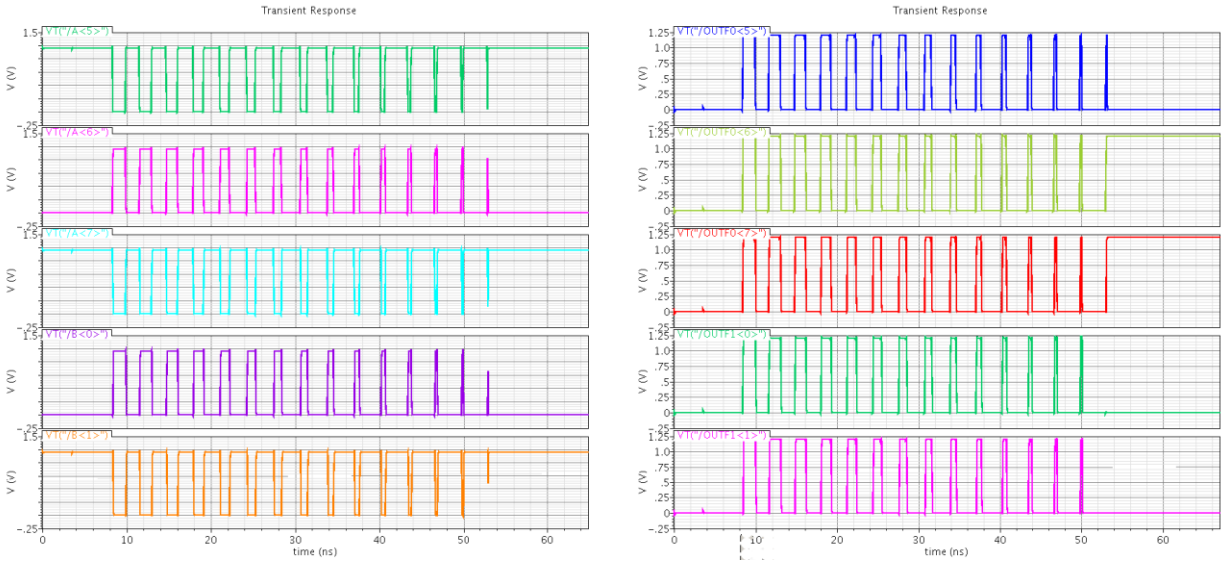
Figure 3.15 - Proposed Rising-Edge Acquisition Circuit (a), Falling-Edge Acquisition Circuit (b)

3.6.2 **One-of-N-Code**

The “First final step” to obtain the intended output binary code, section 3.6, will be to obtain the final position of the pulse in the delay-chain that has been previously sampled. This step is performed via a structure named one-of-N-Code that grabs in all of the outputs from the acquisition circuit, and selecting from those the one that achieves a higher position in sequence thermometer code of 1s. Additionally this circuit prevents the possible occurring of bubble errors that could cause serious errors at the output binary number.

Recalling the declaration in the control logic, section 3.5, it describes that, when the pulse enters into the next two delay elements and completely vanishes from the present one, the present sampling element is reset. Therefore, besides keeping track of the pulse position, it also somehow prevents any possibility of getting bubble errors.

So, during the measurement of the input pulse, it will get smaller and smaller due to shrinking, hence at some point the input “clock pulse” will not provide enough hold time to trigger the next sampling element. Consequently, the control logic and the measurement will stop, and the code returned from the sampling elements will at most have two outputs with a high logic level.



(a) (b)

Figure 3.16 - Partial TDC transient response, "clock inputs" (a), sampling elements output (b)

Figure 3.16 (a), represents the pulse travelling through the delay-line, showing both flacks (rising and falling). As it is shown, the pulse width shrinks over time through the line and at a certain time, in this example at the element 100, it vanishes. In Figure 3.16(b), the respective sampling outputs are shown, and as it was described in the control logic, section 3.5, each stage receives a reset command when the following two elements received the pulse as it passed its respective element completely. At the end of the measurement, the pulse will not reach the next two sampling elements, therefore it leaves, as it has been previously explained, two outputs with a high logic level.

This could be problematic, because if these outputs stand between two transition bits, it will result in a measurement error that can get worse if it happens in the most significant bits.

Since the control logic helps in the prevention of bubble errors, a complex circuit to correct them will not be necessary. Therefore, for the One-of-N-code a simple circuit that selects the correct sampling output and prevents the conversion errors will be more than enough.

Following the analysis done in the start-of-the-art, concerning Thermometer-to-Binary Decoder, especially for the Fat tree design [30, 31, 32, 34], the following Figure 3.17 shows the proposed tecnnique design for the One-of-N-Code circuit.

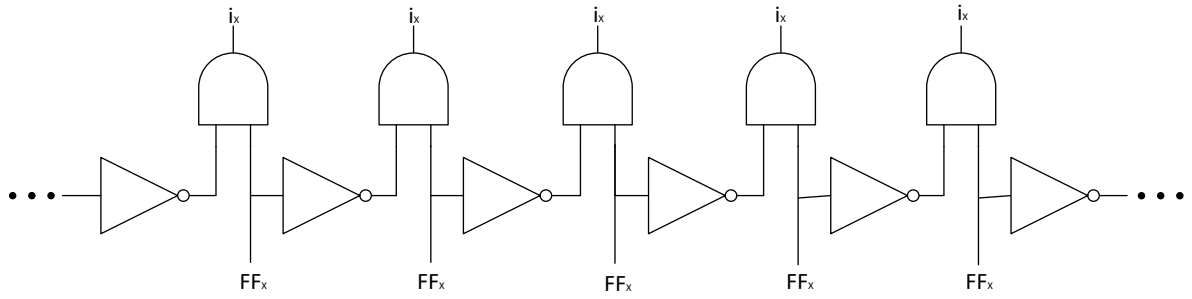


Figure 3.17 - Partial Proposed One-of-N-Code

The circuit is implemented by 2-input AND gate where one of the inputs is the inverted. As the delay-line has 127 elements this circuit will have 127 AND elements, one for each output of the sampling system. So, the outputs are determined by the below formula,

$$OUT_i = In_i.\overline{In_{i+1}}, \quad i = 1,2,3,...2^n - 2 \tag{3.7}$$

where OUT_i is the output at the position i and In_i, In_{i+1} are the inputs from the position I, and $i+1$ respectively.

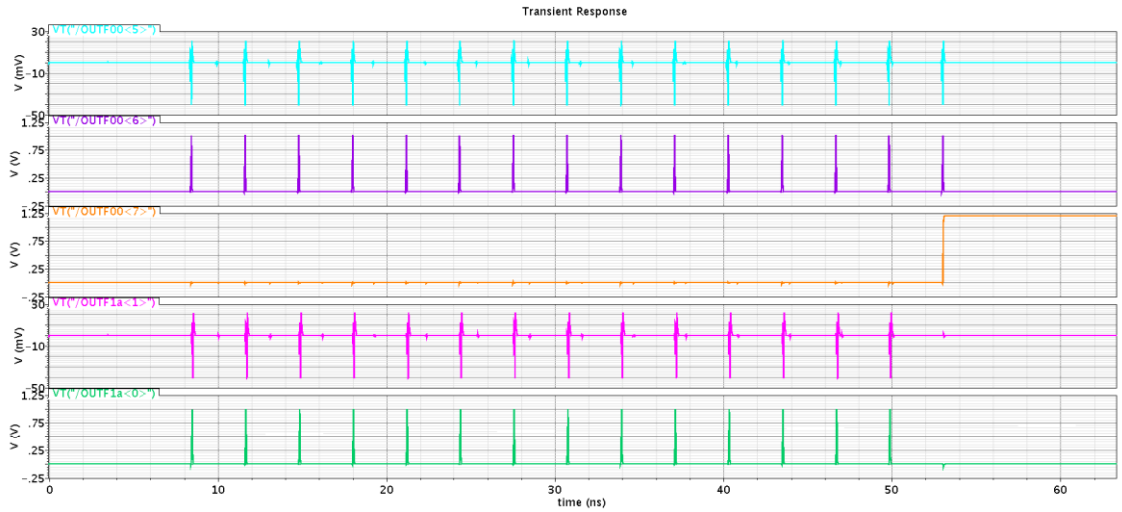


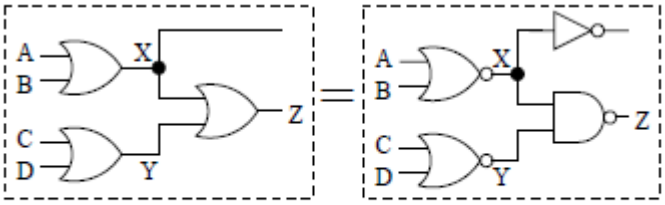
Figure 3.18 - Transient Response Out-of-N Code Circuit

The wave forms presented in Figure 3.18 resulted from the outputs of the sampling elements presented at Figure 3.14(b). It selects from the high outputs the one that is furthest away in the delay-line, pushing the other one-of-n outputs to zero.

3.6.3 **Fat-Tree Circuit Design**

The Final step in the decoder process will be the conversion of the one-of-n code into a binary code, using a fat-tree system design. The Fat-Tree design, section 2.6.4, is based on simple CMOS OR gates built in a tree structure that enables it to consume less power, eliminating static power consumption, and area. [41] Also based on this design and on the use of CMOS gates, it achieves high speed performances, and it does not require either clock signal, sense amplifiers or pull-up resistors, making it a circuit much more tolerant to noise. However, the layout of the fat tree is more difficult when compared with other topologies. Moreover, it is a 3D structure, challenging the layout on a two-dimensional chip.

The most usual fat tree uses Or Gates as its design, as it has been mentioned before. However, for a more efficient implementation in CMOS, using the DeMorgan's theorem it is possible to replace those gates with NOR and NAND gates.



$$Z = \overline{XY} = \overline{\overline{(A+B)} \bullet \overline{(C+D)}} = (A+B)+(C+D)$$

Figure 3.19 - Fat tree logic Implementation [44]

The only thing left to do before building the structure itself will be to define its size. However, this is automatically done because this will follow the same size as the delay-line and the one-of-N-Code, therefore 127 inputs result in a 7-bit fat-tree, Figure 3.20.

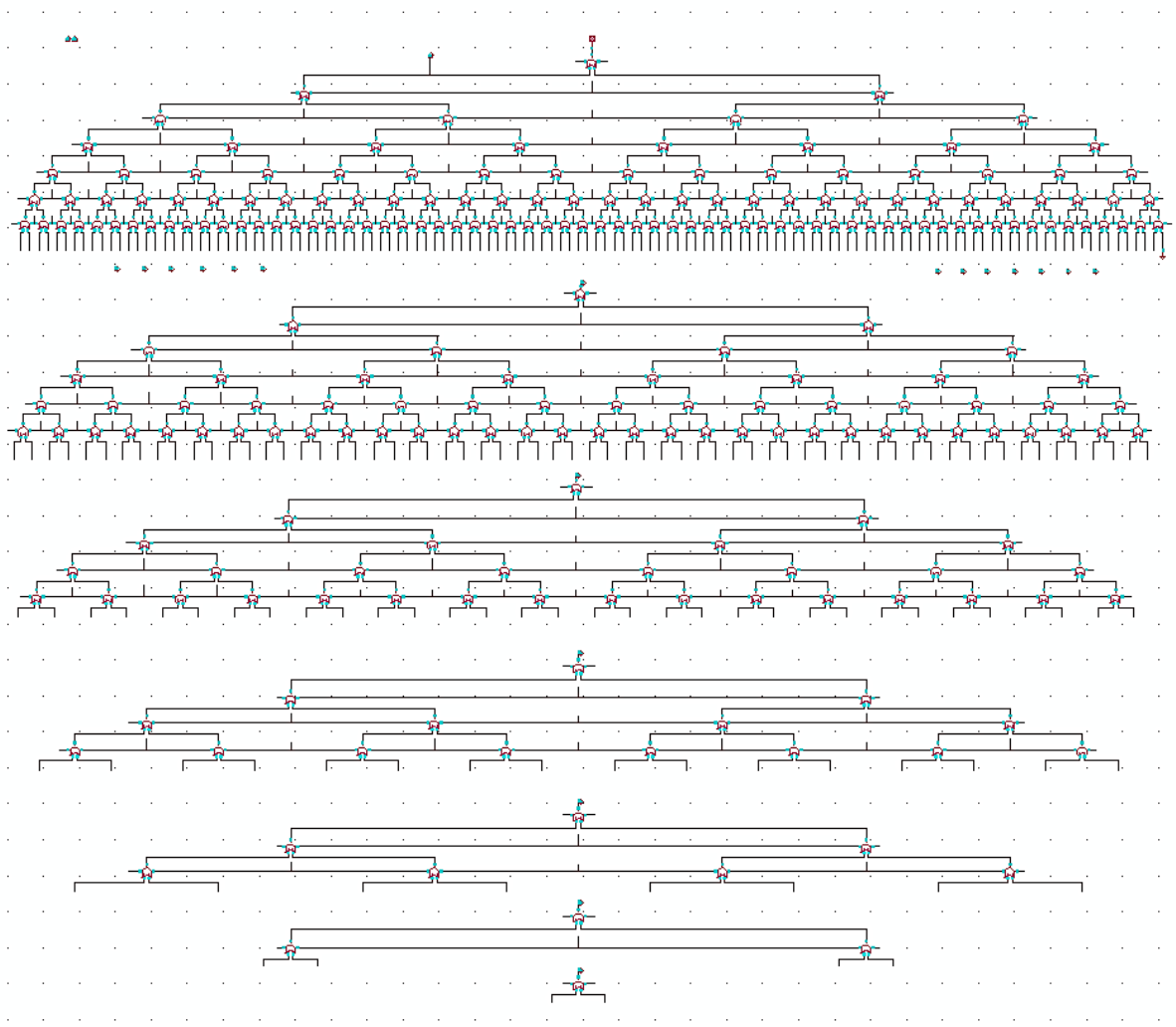


Figure 3.20 - Proposed 7-bit Fat-Tree

3.7 **Registers Circuit**

The Registers are the last block in the proposed architecture for the TDC. Like it was introduced before, they are responsible of saving the output measurement result from the Loop counter and Thermometer-to-binary Decoder. Additionally, it is necessary that the circuit is regenerative in order to prevent data loss in case that the circuit stop making measurements for a while.

The proposed topology used to perform the registers are simple and well-known SR flip-flops connected in a master-slave architecture triggered by the Input. The proposed circuit is shown in the following Figure 3.21.

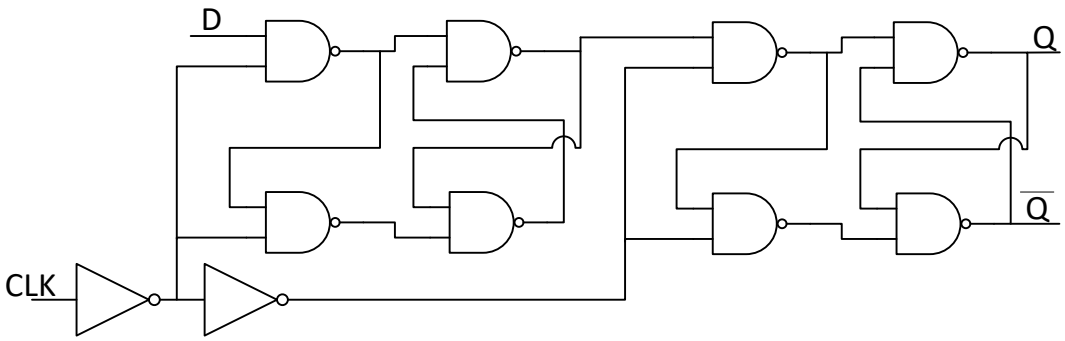


Figure 3.21 - Proposed Master-Slave SR Flip-Flop

The new input data rising edge triggers the Master SR flip-flop to obtain the last measuring result from the Fat-tree and Ripple counter, but the Slave flip-flop data does not change, maintaining its last result. When the input data has its falling edge the Slave SR flip-flop stored the information of the Master SR, goes to the Slave Flip-flop, but the first one also maintains its data. This operation principle will help to prevent any data loss but mainly prevent any possible metastable state in the flip-flop.

3.8 **Time-to-Digital Converter Architecture Summary**

The proposed design for the Time-to-Digital Converter, was based mainly on three key words, i.e. high-resolution, simplicity and low consumption. Therefore, the global structure is only composed of what is strictly necessary for the system to work, i.e. the pulse acquisition circuit, the delay-line, control logic, loop counter and thermometer-code decoder.

The Delay-line structure is based on a Pulse-shrinking loop technique, which is an improvement on the simplest delay-line concept, section 2.2.3, with a mixture of the Vernier concept, making it possible to achieve sub-gate resolutions, with the minimum area and power consumption. Additionally, it has been taken one step forward, the usual Shrinking techniques use delay-line based on buffer elements. Here the proposed architecture uses inverter-based delay-lines, allowing it to at least double the resolution. The only drawback is the additional control and sampling concept that need to be created, in order to make the system work, since now it is working with inverted pulses that are propagating through the inverter delay-line.

As it is a TDC looped structure, a counter topology was proposed to count the number of times that the pulse completed the line. The proposed architecture is a well-known asynchronous counter, i.e the ripple counter, with a 5-bit design returning the course TDC measurement. This counter has a simple design and consumes less area and power when compared with the synchronous ones, and since the delay time that the last bit, after the clock triggers, takes to change, it is not a problem. The asynchronous concept is more than enough to achieve the intended task. However, to decrease even further the area and power used by the counter, dynamic flip-flops were used.

The next structure in the system is the control circuit, that as it has already been shown, it has a very simple concept and circuit, fulfilling the intended thesis objectives. It is responsible for cleaning the circuit to receive a new measurement and keeping track of the location of the travelling pulse in the line.

Additionally, it also helps with the prevention of bubble errors in the Thermometer-to-Binary Decoder. As it is an inverter-based delay-line, two circuits need to be built, in order to control both rising and falling pulses.

Last but not least, but not the least the Thermometer-to-Binary decoder is responsible for the fine TDC measurement which was performed with a fat-tree concept, with a 7-bit design. The Fat-tree concept is a two-step process to get a final Binary word. The first step performed by a One-of-N code simply chooses from the thermometer code that has already been acquired via a sampling circuit based on a TSPC flip-flop concept, the high logic level that goes further forward in the delay line chain. The second step is performed by the designed fat-tree, and takes the output from the previous step and converts it into the final word, using CMOS gates built in a tree structure.

Finally, there is the Register circuit that is mainly SR Flip-Flops with a master-slave topology, which are responsible for the “register” of the output result from the performed measurement. Preventing any possible data lost.

With the 7-bit Fat-tree decoder and the 5-bit ripple counter the proposed TDC will have a 12bit architecture, leading to 4096 different time codes. However, and as it was referred in section 3.3.1.2 the maximum time that the TDC will be able to measure without overflow is 2,918ns. Therefore, the effective number of bits in the TDC will be 11,7.

3.8.1 **Measurement Result**

To achieve the result is only necessary to sum up the results from the Loop counter with the result obtained from the Thermometer-to-Binary Decoder Circuit.

$$\Delta t = (128.N + B_{Fat-Tree}).T_{LSB} \quad (8)$$

Where N is the number of loops that the pulse made in the delay-line, $B_{Fat-Tree}$ is the result from the thermometer-to-binary decoder and T_{LSB} is the resolution of each element in the delay-line.

4. TDC PERFORMANCE RESULTS

The present chapter will evaluate the performance of the proposed Inverter based Pulse-Shrinking TDC technique. A Full dynamic range measurement will be performed for the different corners, in order to retrieve the output code from the different input times. This will enable the calculation of the circuit INL and DNL respectively.

Additionally, it will be proposed a circuit Layout performed in the 0,13µm UMC process technology, followed by its extraction and comparison with the test-bench results.

All the measurements will be performed using the Hspice simulator.

4.1 Process Corners

The Process Corners will study the influence of the PVT variations in the performance of the circuit. There are 4 process corners, Figure 4.1, although in this thesis it will only be performed the extreme cases and its respective temperature and voltage variations.

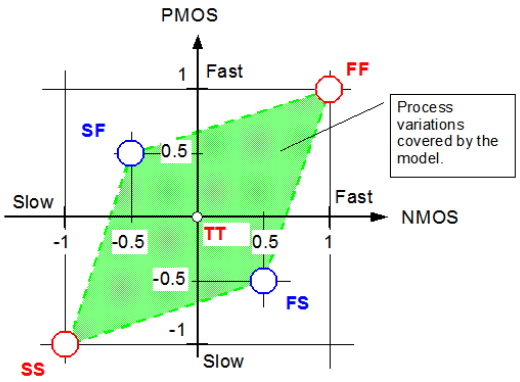


Figure 4.1 – Process Corners

Thus, the following Table 4.1 presents the process corners used in the performed measurements.

Table 4.1 - Tested Corners

Corner	Temperature (ºC)	Voltage Supply (V)
TT (typical)	50	1,2
SS	-40	1,08
		1,32
	120	1,08
		1,32
FF	-40	1,08
		1,32
	120	1,08
		1,32

4.1.1 **TDC Transfer Function**

In order to obtain both measurements a full dynamic range input time sweep was performed, obtaining the correspondent output binary codes. Additionally, the corresponding output transfer function was obtained. The following Figure 4.2, shows the obtained transfer function for each corner tested.

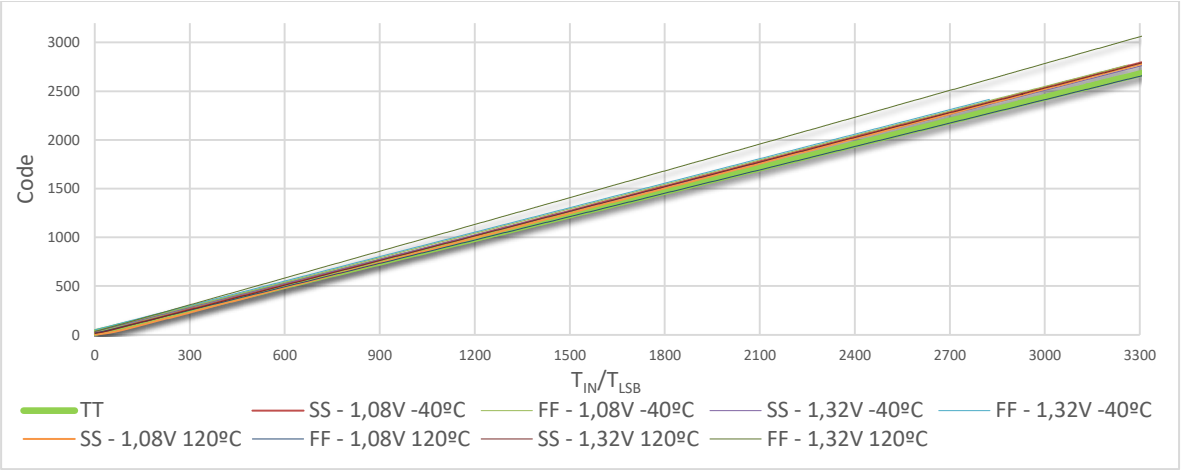


Figure 4.2 - TDC Transfer Function

The Time-to-Digital Converters just like the conventional ADC are characterized by a uniform staircase curve. If there are no errors or missing codes in the converter output codes the length of each step will be equal to the resolution of the converter itself, in this case T_{LSB}. This means that ideally, if a line was drawn between each stair boundary it will end up with a straight line.

As the graphic shows, each of the tested corners presents the intended increasing linear response. However, due to its dynamic-range and huge number of samples taken, the staircase characteristic is not perceptible. Therefore, it will not be possible to conclude that the circuit does not have any missing or error code in its output function. To test all the non-linear imperfections and all the deviations of the TDC characteristic from its expected shape that lead to a non-linear distortion, the DNL and INL of the converter will be calculated.

Additionally, as suggested each corner transfer function presents a different slope. As already mentioned above, this is due to the effect that each of the tested corner has on the delay-line time-resolution, as the equations (2.12), (2.13) and (2.14) show. In addition, the delay of each element will also change and with this the dynamic range will suffer variations.

4.1.2 **DNL Measurement**

DNL (Differential nonlinearity) describes the deviation between the two output values corresponding to adjacent input quantities. It is an important specification for measuring error of a converter, translating its accuracy.

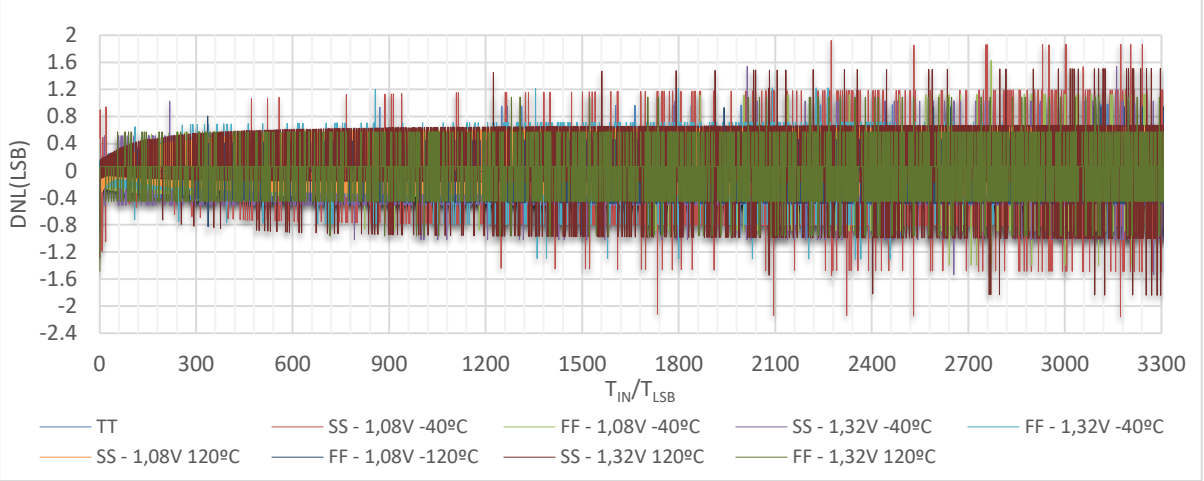


Figure 4.3 – DNL TDC Performance

The present graphic in Figure 4.3 represents the DNL evolution through all the dynamic range of the converter. The results shown in each corner were normalized to one T_{LSB}. The following Table 4.2 resumes the maximum and minimum values achieve by the DNL, for each of the analyzed corner:

Table 4.2 - DNL Corner Results

Corner	T (ºC)	Vcc (V)	DNL (LSB)	
			MIN	MAX
TT	50	1,2	-0,98	0,97
SS	-40	1,08	-2.16	1,92
		1,32	-1,53	1,54
	120	1,08	-2,21	1,97
		1,32	-1,84	1,50
FF	-40	1,08	-1,40	1,62
		1,32	-1,30	1,22
	120	1,08	-0,96	0,94
		1,32	-1,48	1,1

In a macroscopic analysis DNL is the description of existent errors or missing codes in the converter dynamic performance. So, and since it is in normalized, each time that DNL is different from zero means that sample presents an error in the output code. Looking at the Figure 4.3 the first conclusion is that the proposed TDC could present multiple missing codes among all the dynamic range, hence it is not performing well. Moreover, as suggested in Table 4.2 suggest, those missing codes may rise up to 2,21 LSB. However, it will be demonstrated that this assumption cannot be made.

4.1.3 **INL Measurement**

INL (Integral nonlinearity), is a measure of the deviation between the ideal output value and the actual measured output value for a certain input code. This measurement is performed after gain errors have been compensated. To obtain it, a best fit line of the obtained transfer function is used against the transfer function itself.

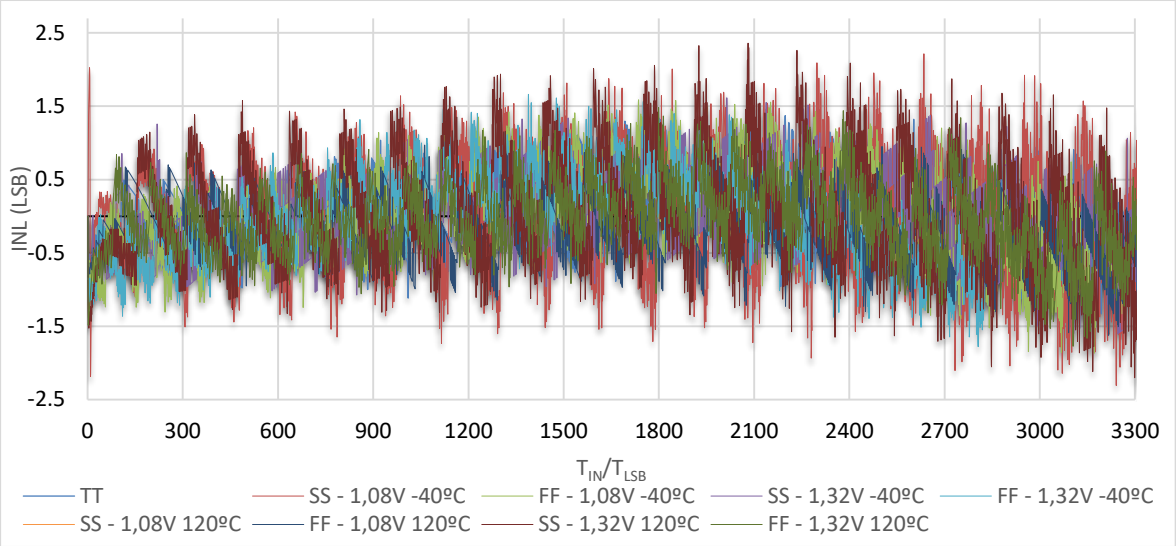


Figure 4.4 - INL TDC Performance

The present graphic in Figure 4.4 represents the INL evolution through all the dynamic range of the converter. The results shown in each corner were normalized to one T_{LSB}. The following Table 4.3 resumes the maximum and minimum values achieve by the INL, for each of the analyzed corner:

Table 4.3 - INL Corner Results

Corner	T (ºC)	Vcc (V)	INL (LSB)	
			MIN	MAX
TT	50	1,2	-1,52	1,27
SS	-40	1,08	-2,30	2,21
		1,32	-1,60	1,61
	120	1,08	-1,87	1,55
		1,32	-2,1	2,3
FF	-40	1,08	-1,86	1,58
		1,32	-1,77	1,54
	120	1,08	-1,58	0,97
		1,32	-1,77	1,48

In a macroscopic view INL represents the description of the bending of a converter characteristic, describing the deviation of each step position from the ideal curve of the converter. As the name suggests “integral”, the INL sums all the non-linear events that lead to the deviation of the obtained fit line from the obtained results.

After the analysis of all the presented results for all of the tested corners, it can be concluded that in fact the TDC transfer function follows the intended linear curve, with some deviations in the slope due to the process corners. However, when we look to the INL and especially to the DNL results it suggest that the converter presents missing codes. This may lead to concluding that the TDC is not performing so well, which in this case is incorrect. The origin of this incoherence lies in the way that the tests and results were obtained, as it going to be shown further beyond.

4.2 **LAYOUT**

The circuit Layout will be presented in this section, starting by the basic component blocks, followed by their Assembly. In this design the main component will be the delay-line and its 128 inverters. The goal of the proposed Layout will be to obtain a simple, compact and modular design.

Seven metal levels were used to implement all the circuit connections between the different components. It was decided that all the PMOS transistors will use 2 fingers. The main objective of transistor fingering is to reduce the influence of process variations, due to the symmetrical variations that each transistor finger will suffer. Fingering also reduces the S/D junctions area, capacitance and gate resistance. Additionally, as a positive side effect it helps to create a more compact circuit.

4.2.1 **Basic Block Components Assembly**

The Layout Assembly is critical in a Time-to-Digital Converter Design, mainly in its delay-line where the actual measurement occurs. Any asymmetry in the layout can cause non-linearity behaviour in the TDC output response, which translates in the increase of the DNL and INL. Therefore, it is important to assure that the elements have the same structure, and are equality spaced to assure that their output nodes have the same parasitic resistances and capacitances associated. In that case it can be concluded that if the elements were disposed in a straight line, the critical point would be the feedback in the loop structure.

The perfect solution to avoid this would be to dispose the elements like a circle. However, it is impossible to build a layout based on a circle structure. So, the best structure to minimize the layout asymmetries would be to dispose the delay elements in a square based structure and place all the remaining components (sampling units, control units and Thermometer decoder) in multiples levels above the delay-line itself maintaining the square structure. In each level it must be guaranteed that all the components have the same width in order to perfectly assembly them, making the connections between them and the adjacent levels easier.

A square structure will guarantee that all elements are placed with the exact same distance from each other, with the exact same connections and metal levels, leading to the exact same parasites in the component nodes, making it a very homogenous circuit. Additionally, this will guarantee that the proposed layout will be approximately symmetrical from every point of view.

4.2.1.1 Delay Line

Composed by inverters connected in series, this block is the heart of the system and it is where all the measurements are performed. A portion of the proposed delay-line is proposed in the following Figure 4.5. As it shows each delay element is equally spaced from each other, approximately 9,2µm from each other. At first glance, it may look like the distance between the delay elements are exaggerated and they could be closer. Although this would be possible if the sampling elements, and the control logic had the same size of the delay elements, if the delay elements where connected closer to each other, the distance between them and the respective sampling element would increase through the delay-line. Thus, despite having harder connections between the different delay elements and their respective sampling elements, this could increase the INL and DNL responses of the system, due the increasing parasitic effects and the unequal line sizes.

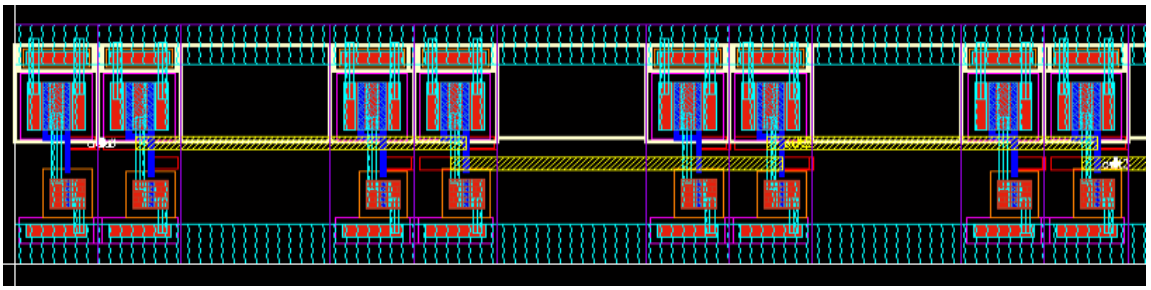


Figure 4.5 - Proposed Delay-Line Layout

4.2.1.2 Loop Counter

Responsible for the count of the number of loops that the pulse did to the delay-line and based in an Asynchronous architecture, also known as Ripple Counter, the proposed Circuit is built in TSPC flip-flops connected in series. The proposed Circuit layout is proposed in the following Figure 4.6. The main concern about this circuit is to assemble the different elements as close as possible in order to get low connections lengths, avoiding the increase of the flip-flops delays, and parasitic effects that will change the operation frequency, setup and hold times. Accordingly, in the final circuit assembly the counter input will be introduced as close as possible to the last sampling element that triggers the counter.

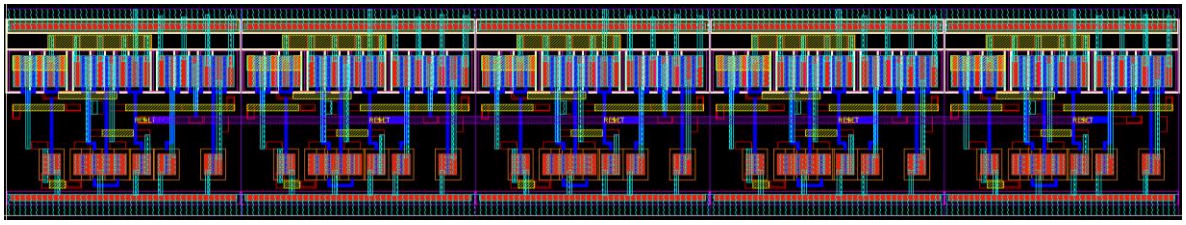


Figure 4.6 - Proposed 5-Bit Ripple Counter

4.2.1.3 Sampling Elements and Control Circuit

In order to make the assembly easier and to guarantee that the circuits have the exact same connections lengths, it was decided to build the sampling elements and the control elements as a unique block, divided in 2 levels. The first level has the sampling elements, and the second level the control elements. The result is shown in the following Figure 4.7. Looking at the proposed layout and taking into account

what was referred before about the delay-line it is clear now why the delay elements were disposed with such distance from each other. It can be concluded that the minimum distance between the delay elements does not depend on the size of the delay elements but on the size of the sampling elements that are directly connected to them. Additionally, with this measurement and same as referred before, it can be assured that the connections length between the delay elements and the respective sampling elements are exactly the same. The same will occur between the sampling elements and the control circuit directly connected to them.

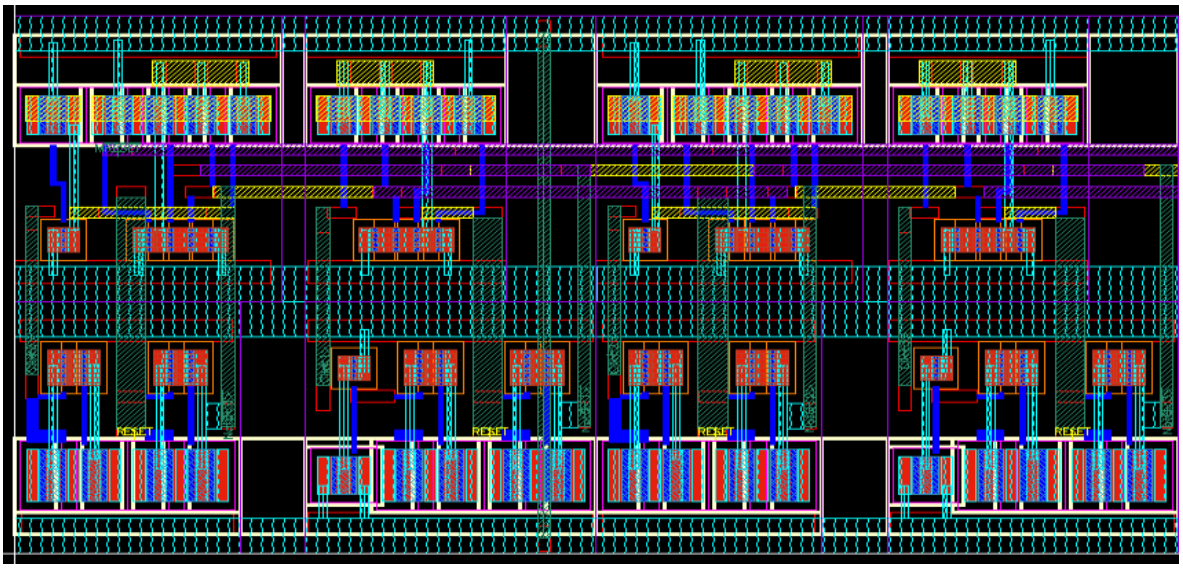


Figure 4.7 - Proposed Sampling and Control Layout

4.2.1.4 Thermometer-to- Binary Decoder

4.2.1.4.1 One-of-N-Code

As it was referred earlier, this circuit is the first final step to obtain the binary word from the thermometer code sampled from the delay-line. Based on inverters and NAND gates it was a very simple and modular structure that is directly connected to the outputs of the sampling elements. So, to make the system more compact and the connections easier, the gate were placed directly above the respective sampling units. The following Figure 4.8 shows the proposed one-of-N code layout.

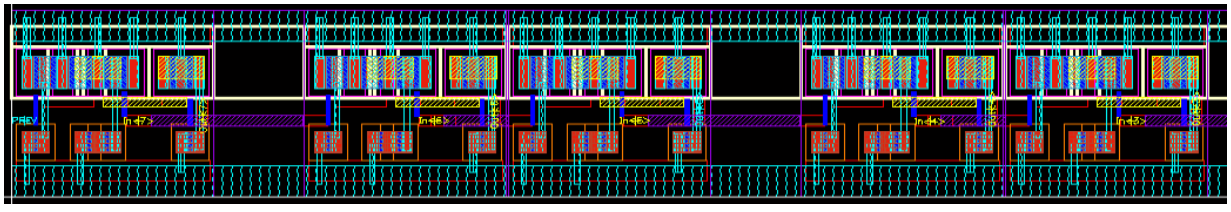


Figure 4.8 - Proposed One-of-N-Code Layout

4.2.1.4.2 Fat-Tree Circuit Layout

The final step to obtain the final fine binary word is the fat-tree circuit. It converts the one-of-N-code obtained by the previous circuit into a 7-bit binary code. Although this circuit is based on simple NOR and NAND logic Gates, its design is very complex due to the numerous gates and connections used to

perform the decoding, as can be seen in the Figure 3.20. The first step in order to make the tree layout and place it like a square structure, the first step is to identify the multiple levels of the tree their elements. Then, in an attempt to make it “symmetrical”, each element, is placed between the outputs of the previous level, where possible. This is in other to guarantee that the connections used have exactly the same length. The following Figure 4.9 shows a portion of the concept used in the build of the fat-tree layout

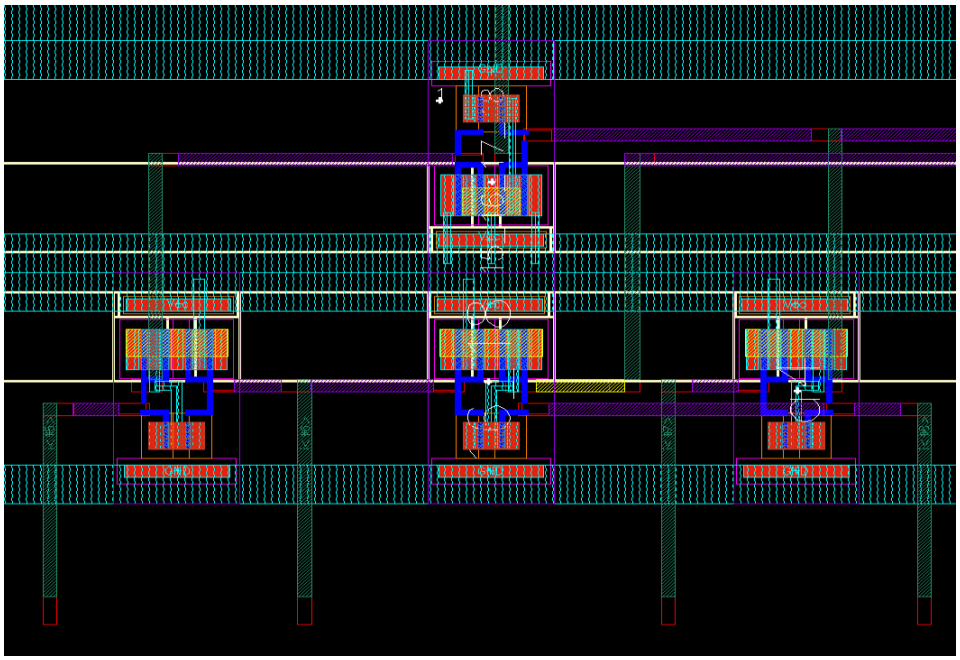


Figure 4.9 - Fat-Tree Sample Layout

4.2.1.5 Registers

In order to save the final results from the loop counter and the fat-tree, a circuit edge-triggered latch was used, preventing the data from getting lost mainly due to leakage currents. The aforementioned register is based on a master slave flip-flop design. Additionally, they are placed near of the outputs they need to register, in order to reduce the length of the connections as much as possible. The following Figure 4.10 shows the proposed layout for the register.

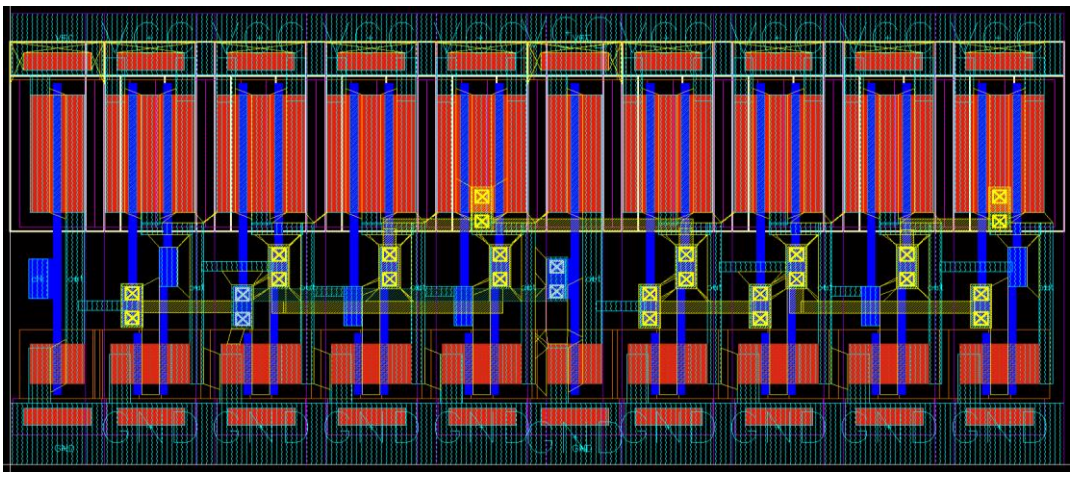


Figure 4.10 - Proposed Register Layout

4.2.2 Time-to-Digital Converter Assembly

After describing each part that constitutes the proposed Time to Digital Converter, this chapter will show the final layout circuit layout design linking every part presented before. Based on a 128 pulse-shrinking delay-elements, the proposed layout shown in the following Figure 4.11, exhibits the squared architecture, as it was referred before. The proposed layout is divided in 6 main blocks: the fine registers and fat-tree (yellow box), the one-of-n code (white box), the sampling and control logic (red box), the pulse-shrinking delay-line (green box), the loop-counter and the course registers (orange box).

Giving that the circuit has multiple levels, a power grid technique was used to power the system, alternating VDD and GND power lines. The power grid was introduced with 2 metal layers, metal 7 and 8 disposed horizontally and vertically respectively. Additionaly, and since the middle of the square was empty, decoupling capacitors (blue box) were connected between VDD and GND. In addition to fill up the empty area, the introduction of these capacitors filters any disturbance in the power lines.

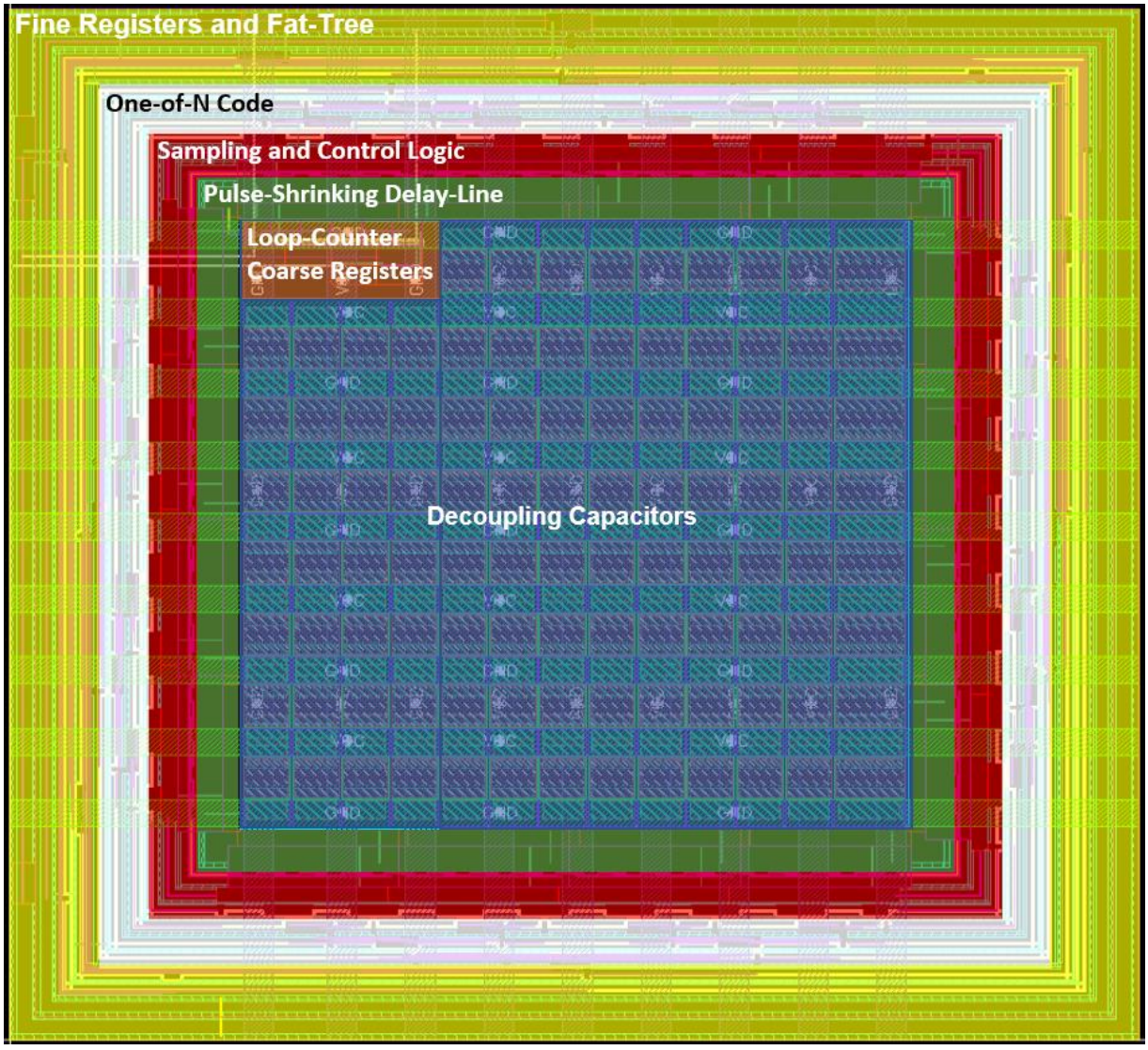


Figure 4.11 - TDC Full Layout

The global Time-to-Digital Converter Structure occupies an area of 0,14768mm^2.

4.3 Extraction Results

The present chapter will evaluate the performance of the proposed TDC layout architecture. As previously done, a full-range dynamic-range measurement is going to be performed to obtain the TDC transfer function curve. From the characteristics curve the DNL and INL will be obtained. Simultaneously, these results will be compared with the ones obtained for the TT corner (typical case), to verify the influence of the parasitic components in the performance of the proposed TDC.

Finally, the power consumption and the Figure of merit (FOM) of the circuit will be obtained.

4.3.1 TDC Transfer Function

This section describes the behaviour of the extracted circuit in a full dynamic range measurement, comparing it to the TT corner. Same as before a growing input pulse was introduced in the TDC input gate to obtain the respective output measurement binary code. The following Figure 4.12 describes the evolution of the output code through different time input pulses.

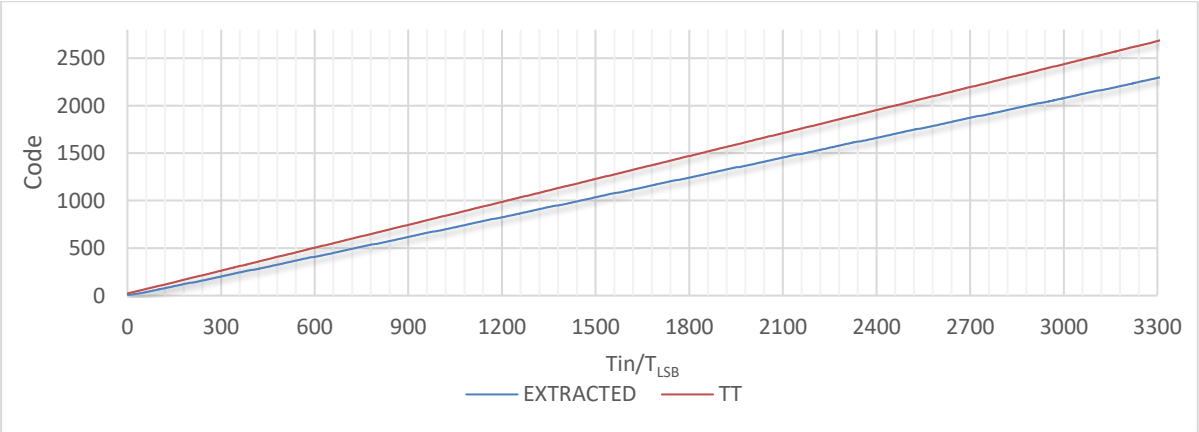


Figure 4.12 - TDC Transfer Function - Extracted vs TT Corner

As the present graphics illustrates the extracted circuit follows the same uniform staircase case as the TT corner. However, and as aforementioned, due to the parasitic capacitances, the obtained extracted circuit presents a lower time-resolution, achieving a maximum resolution of 1.2ps. For the same input time pulse, the achieved output code is lower. Therefore, the presented extracted transfer function curve (blue line) has a lower slope when compared to the TT corner.

4.3.2 DNL and INL Measurement

Through the transfer function previously acquired, the DNL and INL performance of the extracted TDC circuit was calculated. The Following Figure 4.13 and Figure 4.14 present the DNL and INL results, respectively, versus the TT Corner.

As it was referred before each time the DNL result is different from zero, it means that the output code presents a deviation from the intended output. The results suggest that the extracted circuit also presents errors in the codes throughout all dynamic range. The reason why that is going to be explored in chapter 4.4.

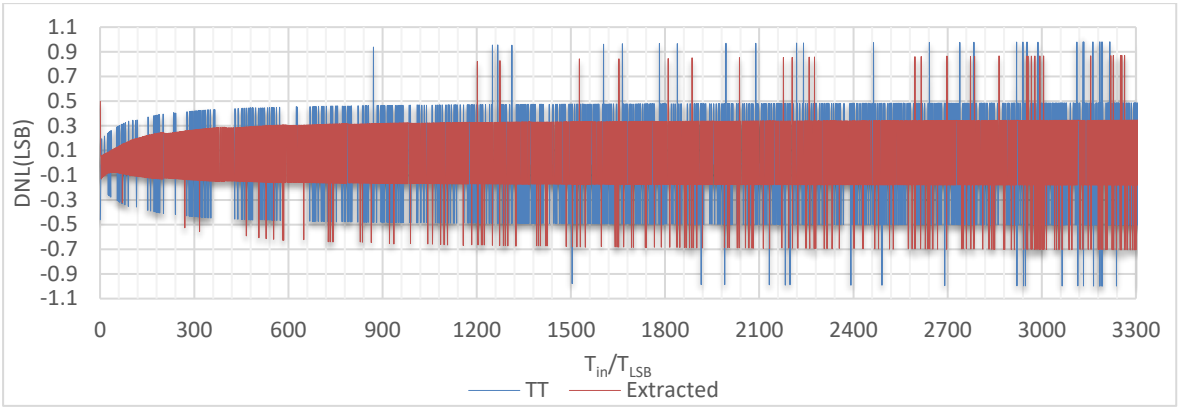


Figure 4.13 - Extracted vs TT DNL performance

However, if the previous facts are discarded and results considered “normal”, the extracted results when compared with the TT corner exhibits a lower DNL. This seems contradictory, because if the parasitic elements of the circuit were added the DNL answer should have achieved higher values. But what really happened, was that in fact the DNL decrease, which is positive, but the consistence of the occurrence of errors increased. This effect is translated in the INL response as the Figure 4.14 will show. This is expectable due to the decrease of the time resolution, and the way that the test was performed, section 4.4.1. The decrease of the time-resolution is also related to the fact of the DNL is tendentially more negative.

From the difference between the transfer function of extracted circuit presented before (Figure 4.12) and its best fit line the INL response of the system was obtained. The following Figure 4.14 compares the extracted circuit INL with the TT simulated corner.

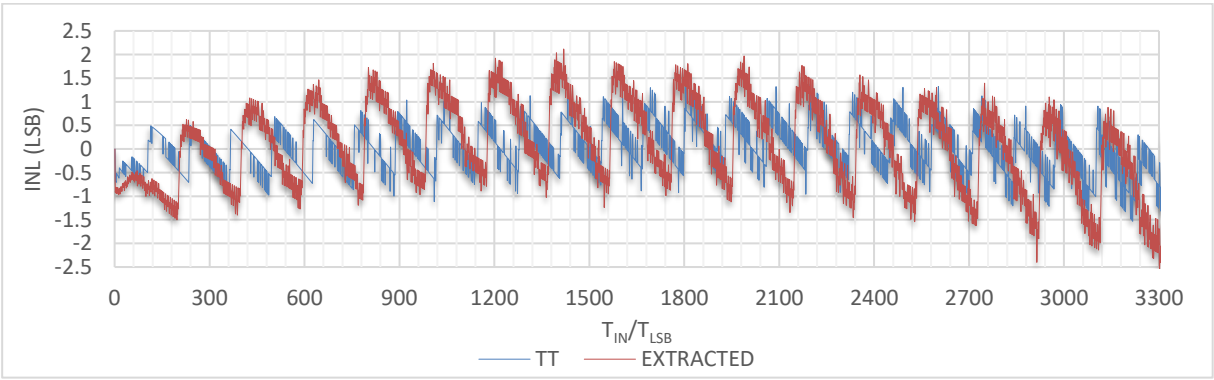


Figure 4.14 – Extracted Circuit vs TT Corner INL Results

The presented graphical results show that the extracted INL follows the same curve as the TT corner, but with higher INL values. As discussed before in the DNL results, even though the absolute values decreased, their consistence increased and knowing that the INL draws the accumulation of the normalized deviations it was expectable that the resultant INL for the extracted circuit increased too.

As the graphic suggests, both results present a “jump” in the INL value, periodically. These events happen each time that the pulse does a loop to the delay-chain. The following Figure 4.15 shows the delay-line “closing loop” section.

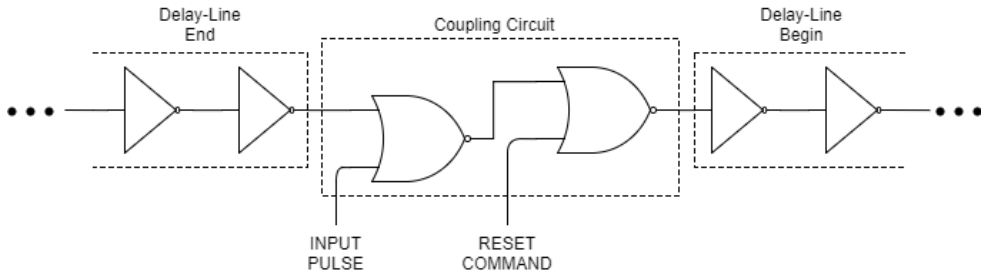


Figure 4.15 - TDC delay-line Loop

As the circuit diagram shows, and as referred before, the beginning and the ending of the delay-line are linked by the Coupling Circuit, in section 3.2, closing the delay-line loop. The coupling loop, based on NOR gates, presents bigger transition times than the inverters. However, the coupling circuit has ideally been projected in such a way that leads to a null pulse-shrinking. The time difference needed to obtain the signal of the output is bigger in the inverters. Adding the fact that the sampling elements only exist in each inverter output, it leads to a failure of the output binary code. Hence, it leads to a non-linearity in the TDC delay-line.

The following Table 4.4 resumes the obtained results from the DNL and INL, for the extracted circuit and TT corner analysis.

Table 4.4 - DNL and INL results comparison

Result	**DNL (LSB)**		**INL (LSB)**	
	MIN	**MAX**	**MIN**	**MAX**
TT	-0,98	0,97	-1,52	1,27
Extracted	-0,91	0,87	-2,4	2,11

4.3.3 **Power Consumption**

To obtain the consumed power by the TDC proposed circuit, some measurements test were performed, from low to high measurement times in the dynamic range. The following Figure 4.16 shows the obtained results for each test.

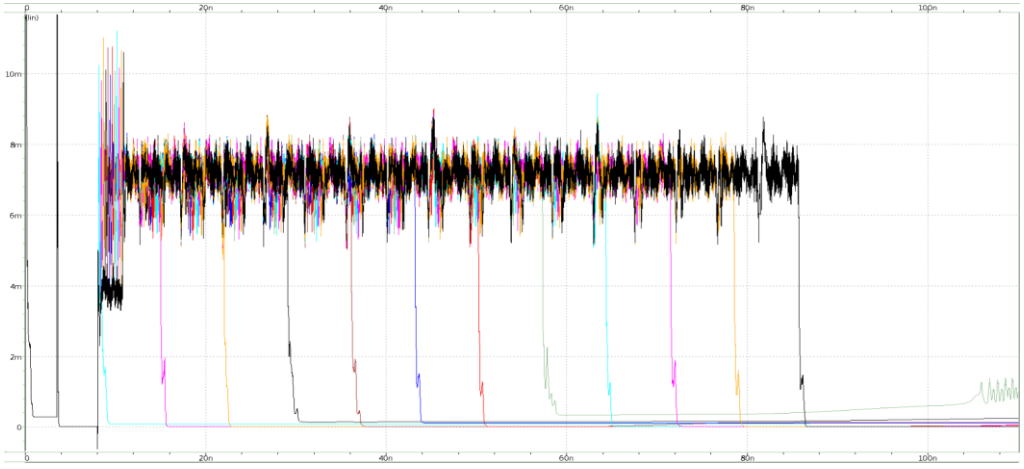


Figure 4.16 - TDC Power Consumption

As the last Figure 4.16 describes, at the starting point, the TDC circuit has a higher power consumption, this is due to the initialization of the circuit, meaning delay-line, counter and Fat-tree reset. Here the power consumption has peaks of 12mW.

As it can be seen, during the pulse measurement as we can see the power consumption is constant independently of the length of the measurement pulse. From there, it is possible to conclude that the power consumption during the measurement itself is around 7,5mW.

4.3.4 **Figure-of-Merit**

To compare the performance of ADCs of different architectures or design specifications the following Figure-of-Merit, quantifies the overall performance of an ADC by evaluating the amount of the power consumption of the ADC per conversion step. The same can be imported to the Time-to-Digital Converters to compare them.

$$FOM = \frac{P}{2^{ENOB} \cdot f_s} \text{ [J/conv]} \tag{4.1}$$

where P is the power consumption during the measurement operation, ENOB is the effective number of bits and f_s is the sampling frequency, that in the TDC case will be the frequency of each measurement.

4.3.4.1 Effective Number of Bits

To obtain the FOM of the circuit, first the effective number of bits (ENOB) needs to be calculated. The proposed TDC architecture uses 12 bits to deliver the output binary word. This means that the converter should be able to deliver 2^{12}=4096 different time codes, that with a time resolution of 0,82ps results in a maximum measurement time pulse of 3,35ns. However, section 3.3.1.2 shows that the maximum time that the delay-line can measure to prevent it from overflow is 2,918ns, in typical conditions. Hence and considering the corners variations in all the measurements performed the maximum number of samples tested was 3300. Therefore, the ENOB of the Proposed TDC is going to be,

$$ENOB = \log_2(3300) = 11,701 bits \tag{4.2}$$

4.3.4.2 Sampling Frequency

The Sampling Frequency (f_s) of a conventional ADC translates the necessary time needed to perform each measurement. Although, in the proposed TDC and in the most of the know techniques, the time needed to measure a time pulse depends of the length of the pulse itself and in the resolution achieve by the TDC.

In order to evaluate the FOM of the proposed TDC it was consider the longest conversion time, which happens when the maximum time is measured. The necessary time to perform the measurement was 81ns. Therefore, the Sampling Frequency of the Proposed TDC is going to be,

$$f_s = \frac{1}{T_{mes}} = \frac{1}{81n} = 12,35 MHz \tag{4.3}$$

where T_{mes} is the measurement time.

With both factors now evaluated the FOM for the proposed TDC is going to be,

$$FOM = \frac{P}{2^{ENOB}.f_s} = \frac{7,5m}{2^{11,7} * 12,35M} = 91,2\ fJ\ /\ conv \tag{4.4}$$

4.4 TDC Performance Summary

In section 4.1, the global performance of the proposed of the TDC inverter-based Pulse-shrinking architecture was evaluated. Starting by the Corners, the influence of the PVT variations was analysed, for the extremes cases.

To test the Linearity of the System and extract the DNL (differential non-linearity) and INL (integral non-linearity), a full dynamic-range time measurement was performed, extracting the resultant output binary output codes from the Fine and Course measurements. With the obtained results, the respective transfer functions where plotted, section 4.1.1. Each of the tested corners presents the intended increasing linear response, however, due to the huge number of samples taken, the staircase characteristic is not perceptible. Due to the different time-resolutions that each corner achieves, the different transfer functions presenting different slopes.

From the obtained transfer functions, the DNL and INL results were obtained, section 4.1.2, and section 4.1.3 respectively. From the DNL analysis, it was concluded that the TT corner achieved a variation between -0,98 and 0,97 LSB, and the worst process corner, SS corner with a Vcc=1,08V and T=120ºC achieved a DNL variation between -2,21 and 1,97 LSB. In the INL analysis, the TT corner achieved a variation between -1,52 and 1,27 LSB, and the worst process corner, SS corner with a Vcc=1,08V and T=-40ºC achieved a variation between -2,30 and 2,21 LSB.

In section 4.2, the full circuit-layout was proposed, with the objective of obtaining a simple, compact and modular design. Seven metal levels were used to implement all the circuit connections. The Time-to-Digital Converter Layout Design is critical, as asymmetry can cause a non-linearity behaviour in the TDC output response, translated in the DNL and INL. To assure that each element has the same structure, and are equality spaced to get the same parasitic resistances and capacitances associated, a multi-level square-based structure was presented. The obtained structure occupies of 0,14768µm^2.

After designing the layout for the proposed TDC system, the extracted circuit was obtained, and with that all the intrinsic parasitic components associated. Same as before, multiple measures were performed to measure the influence of the parasitic components in the projected TDC. The results from these measures were presented in section 4.3, and compared with TT Corner.

The full dynamic range measurement, at section 4.3.1, presented the obtained transfer function from the extracted circuit. The obtained curve follows the same curve type as the TT corner. However, due to parasitic capacitances, with a lower time-resolution results in a curve with a lower slope when compared to the TT corner. The maximum resolution achieve by the extracted circuit was 1,2ps.

From the transfer function previously acquired, the DNL and INL performance of the extracted TDC

circuit was calculated. The obtained results were presented in section 4.3.2. From the DNL analysis it was concluded that the extracted circuit achieved a variation between -0,91 and 0,87 LSB. The INL analysis presents a variation between -2,4 and 2,11 LSB.

Relatively to the power consumption the section 4.3.3 shows that the power consumption is independent from the length of the intended measure time, consuming 7,5mW.

The FOM obtained for the proposed TDC architecture is 91,2 fJ/conv

The Following Table 4.5 summarizes the obtain results for the Proposed TDC Architecture comparing them with other state-of-the art TDCs.

Table 4.5 - Performance Comparison with prior arts

Ref.	[18]	[25]	[56]	[57]	[58]	**This work**
Tech [nm]	90	180	130	65	90	**130**
T_{LSB} [ps]	-	1.8	6,98	7	1.25	**0.82**
N. Bits	9	9	11	7	9	**11.7**
INL [LSB]	1.5	8.7	1.5	3.3	2	**2.4**
DNL [LSB]	1.2	1.2	0.8	<1	0.8	**0.91**
Power [mW]	0.014	3.4	0.328	1.7	3	**7,5**
Area [mm²]	1	0.07	0.28	0.28	0.04	**0.147**
FOM [pJ/conv]	0.098	0.75	0.40	0.28	0.96	**0.091**

4.4.1 **DNL And INL Results Analysis**

In Sections 4.1.2, 4.1.3 and 4.3.2, the DNL and INL performance analysis for the extreme corners cases and for the extracted circuit was performed, respectively. As referred before, both measures are fully dependent on the transfer function obtained from the full dynamic range test.

From the analysis, especially from the DNL (Figure 4.3 for the Corners and Figure 4.13 for the extracted circuit), it suggests that the obtained circuit has multiple errors and missing codes through all its dynamic range, that could ascend up to 2,21LSB in the worst case. This data may lead to the wrong conclusion that the circuit is not performing well and that is not reliable.

The main source for this problem resides in the method used to obtain the transfer function of the TDC. As explained before, an array of Time pulses covering all the TDC dynamic-range was introduced in the TDC input. After the measurement is finished, the correspondent Output binary word was extracted and saved. The pulse length for each element in the array was obtained through the following equation:

$$T_{in} = T_{LSB}.N \ , \ N=1,2,3\ldots. \tag{4.5}$$

Plotting the obtained output code versus each correspondent input time, T_{in}, normalized to one $T_{LSB,}$ will result in the TDC transfer function. The following Figure 4.17 presents an ideal staircase curve for a

Time-to-Digital Converter. The expected transfer function is an uniform staircase, where each step correspond to one output code and the length of each step is the time resolution, T_{LSB}, of the TDC itself.

The DNL results can be obtained applying the following equation to the transfer function,

$$DNL_i = \frac{T_{OUT_{i+1}} - T_{OUT_i}}{T_{LSB}} - 1 \tag{4.6}$$

So, looking at the last equation and merging its information with the one presented in the graphic, it is possible to conclude that in order to have a perfect DNL (equal to zero), each of the T_{in} pulses obtained by the equation (4.5), should be between the limits of each stair steps in the transfer function. Ideally, each T_{in} should be exactly in the middle of the staircase step.

However, in the real circuit, time-resolutions errors\approximations, PVT variations, parasitic elements and mainly the offset in the circuit, may alter the length and the position of each step in the transfer function. These variations are described by the enlargement or shrinkage of each step, which will result in the translation of the steps through the Tin/Tref axis.

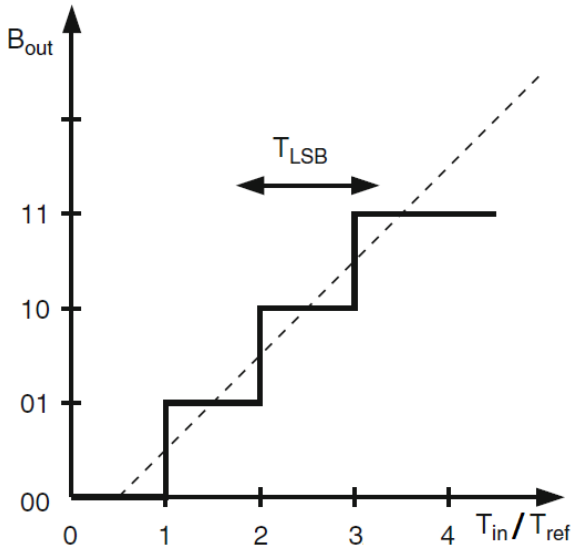


Figure 4.17 - Ideal TDC Transfer Function

The equation (4.5) does not take those variations into account, meaning that is not possible to know the position where each sample is going to be performed. Therefore, the result that was shown to have a possible error could have been a victim of the measurement itself.

Additionally, there are two additional effects that may contribute to this result. The first one, already mentioned in section 4.3.2, is due to time transitions differences between the coupling elements and the delay-line. The second one, and common to all the TDC topologies is the cumulative effect of all non-linearities through the delay-line (section 2.2.5).

Gathering all these facts, it is possible to conclude that the measurement process used for the TDC still needs additional research and development in order to avoid possible measurement errors, that may lead to false conclusions about the performance of the proposed TDC system.

5. CONCLUSIONS AND FUTURE WORK

This thesis presents the development of one of the major known time-mode circuits, the Time-to-Digital Converter. This section presents the closing remarks, and the future work directions for continuing the development this circuit increasing its performance and erase its errors.

5.1 Conclusions

The rapid scaling of technology has lead to a fine increase in the time-resolution, in opposition to voltage/current-mode circuits resolutions. Therefore, time-mode circuits where information is represented by the time difference between two events instead of the nodal voltages or branch currents of electric networks will offer a viable alternative to reduce scaling-induced performance degradation.

Since TDC's are highly digital structures, the key performance points of the converter, such as power consumption, speed and silicon area, are going to improve with the scaling of CMOS technology.

One of the main goals in the development of the TDC was to obtain a simple, and modular structure which, by simply adding or replacing some of the blocks in its constitution, it leads to a new circuit functionality. Following this assumption, the architecture for the Time-to-Digital Converter System, was based on one of the simplest high-resolution TDC structures, which could achieve sub-gate resolution, the Pulse-shrinking technique. Built behind a looped concept, the proposed TDC replaced the traditional buffer shrinking delay-line elements by simple inverters delay-elements, leading to a huge increase in the time resolution. As a result, a 128-inverter delay-chain was obtained, able to measure time pulses up to 2,918ns. Being a loop structure, a 5-bit ripple counter was added to the system in order to count the number of completed loops that the time measurement pulse did to the system (course measurement result). To complete the measurement and obtain the fine result, it is still necessary to sample the delay-line, obtaining from there the right place where the pulse has vanished. For that, Sampling-elements were added to the system. The proposed sampling-elements were based in TSPC Flip-Flops, specifically modified for this intended function, leading to a decrease in the area consumption, when compared to the original flip-flops. Nonetheless, the sampling-elements output is not a binary work, but a Thermometer-Code. To perform the conversion from of a Thermometer-Code into a final binary word, two structures were used: firstly, a One-of-N code, which simply chooses from the thermometer code, the high logic level that goes further forward in the delay line chain and finally, a 7-bit fat tree that takes the output from the previous step and converts it into the final word, using CMOS gates built in a tree structure.

With the 7-bit Fat-tree decoder and the 5-bit ripple counter the proposed TDC will have a 12bit architecture, leading to 4096 different time codes. However, and as it was referred in section 3.3.1.2 the maximum time that the TDC will be able to measure without overflow is 2,918ns. Therefore, the effective number of bits in the TDC will be 11,7.

In section 4.1.2, and section 4.1.3, the DNL and INL, corner analysis was performed, respectively.

The DNL analysis shows that the TT corner has a maximum variation between -0,98 and 0,97 LSB, and in the worst corner, SS corner with a Vcc=1,08V and T=120ºC a variation between -2,21 and 1,97 LSB. The INL analysis presents the TT corner with a variation of -1,52 to 1,27 LSB, and the worst process corner, SS corner with a Vcc=1,08V and T=-40ºC, a variation between -2,30 and 2,21 LSB.

A full circuit-layout for the TDC Architecture was proposed, based on UMC 0,13µm technology. A compact and modular design based on a squared structure was achieved. However, the main objective for this kind of structure was to assure that all elements were subjected to the same conditions and the global circuit has a symmetrical design. This will prevent non-linearities due to parasitic elements in the circuit. The obtained structure occupies 0,14768mm^2.

The post-layout simulation (extracted circuit simulated) performed against the results of the typical case (TT corner), with a 1,2V power supply, and a temperature of 50ºC, revealed a maximum achievable resolution of 1.2ps, a DNL variation of -0,91 to 0,87 LSB and a INL variation of -2,4 to 2,11 LSB. For the Power consumption the circuit revealed a stable consumption of 7,5mW. Finally, the obtained FOM for this architecture was 91,2 fJ/conv.

The DNL and INL results presented in both corners and post-layout simulations suggest multiple errors and missing codes through all its dynamic range. This data leads to the false conclusion that the circuit is not performing well and that is not reliable. As it was demonstrated in section 4.4.1, the cause for this problem resides mainly in the method used to perform and obtain the TDC transfer function, from which these results were obtained. The measurement process used for the TDC stills needs additional research and development in order to avoid possible measurement errors that may lead to false conclusions about the performance of the proposed TDC system.

5.2 **Future Work**

Future work suggestions are based on the drawbacks of the proposed circuit. As the results suggest, there is mainly one big drawback in the proposed architecture, the Offset Time. To, solve this problem, there are two possible solutions. The first one would be to sum the offset time to the required measurement pulse, and then subtract that same time in the result output code. This will null the offset of the TDC, making it possible to measure time intervals from zero ps. However, the disadvantage of this measure will be the decrease of the dynamic-range of the TDC, as a part of the delay-line will always be occupied by the offset pulse.

The second alteration would be the implementation of a Time-Amplifier as presented in section 2.4.4. This technique will also null the offset time, and additionally, will highly increase the time-resolution of the TDC. Like the previous measure, the time amplifier has the disadvantage of decreasing the dynamic range of the circuit. Besides, this is an analog based circuit, thus, some of the advantages of being a full digital circuit will be lost.

One other major improvement would be related with the TDC measurement tests. As the section 4.4.1 suggests the method used to obtain the Transfer function from which the DNL and INL results where resultant, needs additional research and development. The results were negatively influenced by the method used, so there is a need to develop this method to avoid errors that lead to false conclusions about the behavior of the circuit. Firstly, the number of samples described by equation (4.5), should be highly increased in order to correctly identify each of the steps and its transitions between adjacent levels. From there, DNL and INL equations should be developed to null the offset effect and take the correct result from the linear region of the TDC.

References

[1] V. Ravinuthula, "Time-Mode Circuits for Analog Computation," Ph.D. Thesis, Department of Philosofy, University of Florida, Florida, United States of America, 2006.

[2] S. Henzler, *Time-to-Digital Converters*, 1st ed., Dr. Kiyoo Itoh *et al.*, Eds. Netherlands: Springer Netherlands, 2010.

[3] F. Yuan, *CMOS Time-Mode Circuits and Systems: Fundamentals and Applications*, 1st ed., Krzysztof Iniewski, Ed. Vancouver, British Columbia, Canada: CRC Press, 2015.

[4] S. Henzler *et al.*, "A Local Passive Time Interpolation Concept for Variation-Tolerant High-Resolution Time-to-Digital Conversion," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 7, pp. 1666-1676, 2008.

[5] Y.-B. Kim, "Challenges for Nanoscale MOSFETs and Emerging Nanoelectronics," *Transactions on Electrical and Electronic Materials*, vol. 11, no. 3, pp. 93-105, June 2010.

[6] R. Staszewski *et al.*, "All-digital TX frequency synthesizer and discrete-time receiver for Bluetooth radio in 130-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 12, pp. 2278-2291, 2004.

[7] J.D.A. v.d. Broek, "Design and implementation of an Analog-to-Time-to-Digital converter," Master Thesis, Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, Box 217, 7500 AE Enschede The Netherlands, Netherlands, 2012.

[8] M. S. Kim, "0.18um CMOS Low Power ADPLL with a Novel Local Passive Interpolation Time-to-Digital Converter Based on Tri-State Inverter," Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts, United States of America, Master Thesis 2012.

[9] P. Effendrik, "Time-to-Digital Converter (TDC) for WiMAX ADPLL in State-of-The-Art 40-nm CMOS," Master Tehesis, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands, 2011.

[10] B. K.S.e. al., "A 100-ps Time-Resolution CMOS Time-to-Digital Converter," *IEEE Journal of Solid-Circuits*, vol. 39, no. 11, pp. 1839-1852, 2004.

[11] N. H.a.J.G. M. Santos, "An 8bit logarithmic AD converter Using cross-coupled inverters and a time-to-digital converter," in *2016 12th Conference on Ph.D. Research in Microelectronics and Electronics*, Lisbon, 2016, pp. 1-4.

[12] N. H.a.J.G. M. Santos, "Logarithmic AD Converter With Selectable," *IEEE Transactions on Circuits*

and Systems II: Express Briefs, vol. 63, no. 3, pp. 234-238, 2016.

[13] N. H.a.J.G. M. Santos, "Logarithmic AD conversion using latched comparators and a time-to-digital converter," in *21st IEEE International Conference on Electronics Circuits and Systems (ICECS)*, Marseille, 2014, pp. 319-322.

[14] J. Kalisz, "Review of methods for time interval measurements with picosecond resolution," *Metrologia*, vol. 41, no. 1, pp. 17-32, December 2004.

[15] F. A.P.C. P. Napolitano, "A Novel Charge-Pump Based Time-to-Digital Converter," in *International Instrumentation and Measurement*, Singapure, 2009, pp. 1473-1478.

[16] N. H.E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design : A system Perspective*, 2nd ed., M. C. Varley, Ed. Menlo Park, California, USA: Pearson Education India, 1988.

[17] I. Bombay. (2009, December) National Programme on Technology Enhanced Learning. [Online]. http://nptel.ac.in/courses/117101058/downloads/Lec-16.pdf

[18] S. Naraghi, "Time-Based Analog to Digital Converters," PhD Thesis, Department of Electrical Engineering, The University of Michigan, Michigan, United States of America, 2009.

[19] R. D.a.T.K.B.a.X.G.a.E.A.M. Klumperink, "Optimized Stage Ratio of Tapered CMOS Inverters for Minimum Power and Mismatch Jitter Product," in *2010 23rd International Conference on VLSI Design*, Bangalore, India, 2010, pp. 152-157.

[20] A. A.A.a.S. Samadian, "Phase noise in inverter-based differential CMOS ring oscillators," in *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference*, San Jose, CA, USA, 2005, pp. 457-460.

[21] A. A. Abidi, "Phase Noise and Jitter in CMOS Ring Oscillators," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 8, pp. 1803-1816, 2006.

[22] P. Dudek, S. Szczepanski, and J. V.H. Hatfield, "A High-Resolution CMOS Time-to-Digital Converter Utilizing a Vernier Delay Line," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 2, pp. 240-247, February 2000.

[23] G. S. Jovanovic and M. K. Stojcev, "Vernier's Delay Line Time–to–Digital Converter," *Scientific Publications of the State University of Novi Pazar*, vol. 1, no. 1, pp. 11-20, May 2009.

[24] J. Yu, F. F. Dai, and R. C. Jaeger, "A 12-Bit Vernier Ring Time-to-Digital Converter in 0.13um CMOS Technology," in *2009 Symposium on VLSI Circuits*, vol. 45, Kyoto, Japan, 2010, pp. 232-233.

[25] T. Iizuka, T. Koga, T. Nakura, and K. Asada, "A Fine-Resolution Pulse-Shrinking Time-to-Digital Converter with Completion Detection utilizing Built-In Offset Pulse," in *IEEE Asian Solid-State Circuits Conference*, Toyama, Jaoan, 2016, pp. 313-316.

[26] K. Karadamoglou *et al.*, "An 11-bit high-resolution and adjustable-range CMOS time-to-digital converter for space science instruments," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 1, pp. 214-222, 2004.

[27] S.F. Lin and C.W. Lin, "Design of Time-to-Digital Converter for Precision Control," in *International Conference of Electrical, Automation and Mechanical Engineering*, Taiwan, 2015, pp. 408-410.

[28] P. Chen, S.-I. Liu, and J. Wu, "A CMOS Pulse-Shrinking Delay Element For Time Interval Measurement," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 9, pp. 954-958, 2000.

[29] C. C. Chen, S. H. Lin, and C. S. Hwang, "An Area-Efficient CMOS Time-to-Digital Converter Based on a Pulse-Shrinking Scheme," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 3, pp. 163-167, 2014.

[30] C.-C. Chen, W. Chang, and P. Chen, "A precise cyclic CMOS time-to-digital converter with low thermal sensitivity," in *IEEE Symposium Conference Record Nuclear Science 2004*, vol. 3, Rome, Italy, 2014, pp. 1364-1367.

[31] J. P. Jansson, A. Mantyniemi, and J. Kostamovaara, "A CMOS time-to-digital converter with better than 10 ps single-shot precision," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 6, pp. 1286-1296, 2006.

[32] L. M.J.X.a.H.M. C. Li, "An asynchronous delay line TDC for ADPLL in 0.13um CMOS," in *2015 IEEE 11th International Conference on ASIC (ASICON)*, Chengdu, China, 2015, pp. 1-4.

[33] R. B. Staszewski, S. Vemulapalli, P. Vallur, J. Wallberg, and P. T. Balsara, "1.3 V 20 ps time-to-digital converter for frequency synthesis in 90-nm CMOS," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 3, pp. 220-224, March 2006.

[34] T. E. Rahkonen and J. T. Kostamovaara, "The use of stabilized CMOS delay lines for the digitization of short time intervals," *IEEE J. Solid-State Circuits*, vol. 28, no. 8, pp. 887-894, 1993.

[35] A. Liscidini, L. Vercesi, and R. Castello, "Time to digital converter based on a 2-dimensions Vernier architecture," in *2009 IEEE Custom Integrated Circuits Conference*, Rome,Italy, 2009, pp. 45-48.

[36] J. Angevare, K. Blutman, A. Zjajo, and N. v.d. Meijs, "A CMOS 0.23pj Freeze Vernier Time-To-Digital Converter," in *2013 NORCHIP*, Vilnius, Lithuania, 2013, pp. 1-4.

[37] S. Tisa, A. Lotito, A. Giudice, and F. Zappa, "Monolithic time-to-digital converter with 20ps resolution," in *ESSCIRC 2004 - 29th European Solid-State Circuits Conference (IEEE Cat. No.03EX705)*, Estoril, Portugal, 2003, pp. 465-468.

[38] D. P, "Low Power Thermometer Code to Binary Code Encoder Based Flash ADC," Master Thesis, Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, Rourkela, Odisha, India, 2014.

[39] Wikipedia. (2017, June) Wikipedia. [Online]. https://en.wikipedia.org/wiki/Unary_coding

[40] E. Sall and M. Vesterbacka, "Thermometer-to-binary decoders for flash analog-to-digital converters," in *2007 18th European Conference on Circuit Theory and Design*, Seville, Spain , 2007, pp. 240-243.

[41] A. V. Kale, P. Palsodkar, and P. K. Dakhole, "Comparative Analysis of 6 Bit Thermometer-to-Binary Decoders for Flash Analog-to-Digital Converter," in *International Conference on Communication Systems and Network Technologies*, Rajkot, India , 2012, pp. 543-546.

[42] M. H. Fino, M. Fakhfakh, and E. Tlelo-Cuautle, *Performance Optimization Techniques in Analog, Mixed-Signal, and Radio-Frequency Circuit Design*, 1st ed., Austin deMarco, Ed. Hershey, Pennsylvania, United States of America: IGI Global, 2014.

[43] V. H. Bui, S. Beak, S. Choi, J. Seon, and T. T. Jeong, "Thermometer-to-binary encoder with bubble error correction (BEC) circuit for Flash Analog-to-Digital Converter (FADC)," in *International Conference on Communications and Electronics*, Nha Trang, Vietnam , 2010, pp. 102-106.

[44] D. Lee, J. Yoo, K. Choi, and J. Ghaznavi, "Fat tree encoder design for ultra-high speed flash A/D converters," in *The 2002 45th Midwest Symposium on Circuits and Systems*, vol. 2, Tulsa, OK, USA, March 2003, pp. 87-90.

[45] M. V.a.K.O.A. E. Sall, "A study of digital decoders in flash analog-to-digital converters," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)*, vol. 1, Vancouver, BC, Canada, 2004, pp. 129-132.

[46] K. R.R.a.M.M. G. L. Madhumati, "Comparison of 5-bit Thermometer-to-Binary Decoders in 1.8V, 0.18µm CMOS Technology for Flash ADCs," in *2009 International Conference on Signal Processing Systems*, Singapore, 2009, pp. 516-520.

[47] Y. J. Park and F. Yuan, "0.25–4 ns 185 MS/s 4-bit pulse-shrinking time-to-digital converter in 130 nm CMOS using a 2-step conversion scheme," in *2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Fort Collins, USA, 2015, pp. 1-4.

[48] Y. Liu *et al.*, "Multi-stage pulse shrinking time-to-digital converter for time interval measurements," in *2007 European Microwave Integrated Circuit Conference*, Munich, Germany, 2007, pp. 267-270.

[49] Y. Gao. (2011, Jan.) The educational encyclopedia. [Online]. http://educypedia.karadimov.info/library/dns_module3_p1.pdf

[50] T. R. Kuphaldt. All About Circuits. [Online]. https://www.allaboutcircuits.com/textbook/digital/chpt-11/asynchronous-counters/

[51] T. R. Kuphaldt. All About Circuits. [Online]. https://www.allaboutcircuits.com/textbook/digital/chpt-11/synchronous-counters/

[52] S. T. Oskuii, "Comparative study on low-power high-performance flip-flops," Master Thesis, Departement of Electrical Engineering, Linköping University, Linköping, Sweden, 2003.

[53] V. Kursun and E.G. Friedman, "Node voltage dependent subthreshold leakage current characteristics of dynamic circuits," in *International Symposium on Signals, Circuits and Systems. Proceedings, SCS 2003*, San Jose, CA, USA, 2004, pp. 104-109.

[54] M. Pelleti, T. K. Murthy, and K. Neelima, "Low-Power and Low-Area Dual Dynamic Node Hybrid Flip-Flop Featuring Efficient Embedded Logic for Low Power Cmos VLSI Circuits Using 120nm Technology," *International Journal of Emerging Technology and Advanced Engineering*, vol. 4, no. 8, pp. 471-476, August 2014.

[55] S.-H. Yang, Y. You, and K.-R. Cho, "A New Dynamic D-Flip-Flop Aiming at Glitch and Charge Sharing Free," *IEICE TRANS. ELECTRON*, vol. 86, no. 3, pp. 496-505, March 2003.

[56] Y. K.a.T.W. Kim, "An 11 b 7 ps Resolution Two-Step Time-to-Digital Converter With 3-D Vernier Space," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 8, pp. 2326-2336, August 2014.

[57] A. L.a.R.C. L. Vercesi, "Two-Dimensions Vernier Time-to-Digital Converter," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 8, pp. 1504-1512, August 2010.

[58] M. L.a.A.A. Abidi, "A 9 b, 1.25 ps Resolution Coarse–Fine Time-to-Digital Converter in 90 nm CMOS that Amplifies a Time Residue," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 769-777, April 2008.