Thèse

Présentée à l'Université de Lille, Ecole doctorale sciences pour l'ingénieur

pour obtenir le grade de

Docteur

Spécialité : Electronique, microélectronique, nanoélectronique et micro-ondes

par

Ángel de Dios González Santos

Circuits de traitement de signal numérique en temps continu ultra-faible consommation en technologie 28nm FDSOI pour applications audio

Ultra-low power Continuous Time Digital Signal Processing Circuits in 28nm FDSOI technology for audio applications

Soutenu publiquement le 29 septembre 2020 devant le jury composé de

Monsieur Dominique Dallet, Bordeaux INP, **Rapporteur**
Madame Marie-Minerve Louërat, CNRS, **Rapporteuse**
Monsieur Gilles Jacquemod, Université Côte d'Azur, **Président du Jury**
Monsieur Andrei Vladimerescu, UC Berkeley – ISEP, **Examinateur**
Monsieur Andreas Kaiser, Université de Lille, **Directeur de thèse**
Monsieur Antoine Frappé, Yncréa Haut-de-France – IEMN, **Encadrant**
Monsieur Philippe Cathelin, STMicroelectronics, **Encadrant, Invité**
Monsieur Benoit Larras, Yncréa Haut-de-France – IEMN, **Invité**

This project is co-financed by the European Union with the European Regional Development Fund (FEDER)

"Sometimes you look in a field and you see a cow and you think it is a better cow that the one you've got on your own field. It's a fact. Right? And it never really works out that way. Probably the same cow it's as good as your own cow"

-Sir Alex Ferguson

"There is a driving force more powerful than steam, electricity, and nuclear power: the will"

-Albert Einstein

Abstract

The focus of this work is the study and development of a feature extraction system using Continuous-Time Digital Signal Processing (CT DSP) techniques, to mitigate the drawbacks of existing implementations based on traditional analog and digital solutions of always-on monitoring sensors for the Internet of Things (IoT).

The target is to extract the spectral content of an audio signal using a novel architecture based on a cascade of configurable CT DSP Finite Impulse Response (FIR) filters. An efficient cascade scheme is enabled by the proposed glitch elimination and delta encoding techniques. Additionally, this work introduces a CT function to estimate the instantaneous power within selected frequency bands to build an output spectrogram. The proposed 12-band system has been validated using behavioral simulations.

The key element for the implementation of this system is the digital delay element. A new delay element has been designed and fabricated in 28nm FDSOI technology and achieves a record tuning range from 30 ns to 97 µs with a power consumption of 15 fJ/event. By extrapolating this result, the system would have an overall peak power consumption of 2.85 µW when processing typical female speech, while consuming approximately 100 nW when no events are generated. Thus, the average system power consumption outperforms state-of-the-art feature extraction circuits.

Résumé

L'objectif de ce travail c'est l'étude et développement d'un système d'extraction des caractéristiques en utilisant techniques de traitement de signal en temps continu, afin de mitiger les inconvénients des implémentations existantes basées en techniques analogiques et numériques conventionnelles, d'un système toujours en veille pour l'Internet des Objets.

La cible est l'extraction du contenu spectral d'un signal audio en utilisant une nouvelle architecture basée en une cascade configurable de filtres à réponse impulsionnelle finie en temps continu. Un schéma efficace pour cascader des filtres est obtenu grâce aux techniques proposées pour l'élimination des glitches et du codage delta. Par ailleurs, ce travail introduit une fonction en temps continu pour estimer la puissance instantanée dans des bandes de fréquences sélectionnées et construire un spectrogramme à la sortie. Le système proposé à 12-bandes fréquentielles a été validée par des simulations comportementales.

L'élément-clé pour l'implémentation de ce système est un élément de délai numérique. Un nouvel élément de retard a été conçu et fabriqué en technologie 28 nm FDSOI et atteint une plage de délai record entre 30 ns et 97 µs avec une consommation de puissance de 15fJ/événement. En extrapolant ce résultat, le système proposé atteint une consommation approximée de 2.85 µW lors du traitement d'un signal vocal généré par une femme, tandis que la consommation statique est autour de 100 nW dans les périodes où il n'y a pas d'activité. Donc, la performance en matière de consommation moyenne d'énergie de ce système surpasse celle des implémentations dans l'état de l'art.

Acknowledgements

Above all I want to thank the Holy Trinity, God Almighty The Father “The I Am That I Am” (Exodus 3:14), God the Son our Lord Jesus Christ and God the Holy Spirit. I will have never started nor finished this project without powerful divine intervention aligning cosmos forces at the right time and the right place. This work with its merits and its shortcomings belongs to God.

I thank M. Andreas Kaiser my PhD. Director for his exhaustive commitment in this project, his time, his patience and the constant thrill to improve this work.

To my Industrial assessor Philippe Cathelin, I have really enjoyed working 3 meters away from him for the last year, thank you so much for the patience, the support and the invaluable design knowledge and experience that helped to bring this work to fruition.

To Antoine Frappe for hist commitment and for open me the doors to this project.

To Benoit Larras for his time and commitment.

To the members of the jury Marie-Minerve Louërat, Dominique Dallet, Gilles Jacquemod and Andrei Vladimerescu, thank you so much for your time and for all your valuable remarks and comments that have contributed to increase the quality of this work.

To all the ST designers Raphael Paulin, Robin Wilson, Mathilde Sie, Patrick Scheer, Andreia Cathelin, Clement Charbuillet that passed me invaluable design knowledge. A special mention to Nicolas Pelloux and Jasmina Antonijevic, who made me feel like just another colleague.

To my now friend Jeff Nowakoswki for his encouragement and support during measurements.

To Ioanna Kriekouki, I will have never finished this project without you, the MVP award goes to you, thank you for the encouragement and for helping me in recovering my passion and swagger for electronics. Thank you for all the discussions and for helping me to see things in life from your rational perspective. I wish you the absolute best.

To Annemarie thank you so much for the support.

To my dear friends and former office colleagues in Lille Dipal Ghosh and Alexis Clemente, thank you so much for the magnificent working environment we created, and the free-speech zone where we had awesome discussions. To Dipal my admiration for always keeping his coolness, I’ll cherish all those restaurants we visited together. Alexis, you are very good, one day you will be the best. Also, greetings to Fabien “El Patron” VanLaehre.

To Henry Happy for his help and to Nora Benbahlouli for her support and kindness.

To Bruno Stefanelli for his help and Jean Marc Capron for giving me the opportunity of teaching.

To those women seated next to me in a train trip, that helped me with the orthography corrections of an important document I wrote in French

To all my friends in Saint Claude, Anne, Greg, Claire, Anna, Victor, Camille, Hamdi, Eveline, Sébastien, Virginie, Emmanuelle, Florine and Mathilde. I lived awesome moments with you. But particularly Claire and Anna I would have never discover that I love riding roller coasters without you.

Greetings also to the army (past and present) of PhD students at ST particularly to Rodrigo Iga and Francesco Renzini, my best wishes to you.

My Mexican Grenoble friends, sister from another mother Karina, Sahara, Adela and the legend from the day he was born in Alcalá de Henares, Spain the one and only Ricardo Gómez Gómez.

Finally to all my friends in Mexico, and my family, specially my grandfather Ramiro Santos the integrated circuit designed in this work was named Gallo in part to honor him, to my grandmothers Lili Vara all the way to heaven, and Alicia González for all her support, to my aunt Carmen Santos for her support and finally to my mother Violeta Santos who has always showed me immense support and who planted on me the interest for science.

To my aunt Marilú González de Elizondo:

Salmo 34: 15:19

15Los ojos de Jehová están sobre los justos,
Y atentos sus oídos al clamor de ellos.
16 La ira de Jehová contra los que hacen mal,
Para cortar de la tierra la memoria de ellos.
17 Claman los justos, y Jehová oye,
Y los libra de todas sus angustias.
18 Cercano está Jehová a los quebrantados de corazón;
Y salva a los contritos de espíritu.
19 Muchas son las aflicciones del justo,
Pero de todas ellas le librará Jehová.

Contents

List of Figures

List of Tables

List of abbreviations

ADC	Analog to Digital converter
ASR	Automatic speech recognition
BPF	Band-pass filter
CA	Continuous amplitude
CAGR	Compound annual growth rate
CMOS	Complementary Metal Oxide Semiconductor
CT	Continuous-time
CT-DSP	Continuous-time digital signal processing
DA	Discrete amplitude
DAC	Digital to Analog converter
DIBL	Drain induced barrier lowering
DSP	Digital signal processing
DT	Discrete time
EEG	Electroencephalogram
ENOB	Effective number of bits
FD-SOI	Fully depleted silicon-on-insulator
FFT	Fast Fourier Transform
FIR	Finite impulse response
IIR	Infinite impulse response
IoT	Internet of Things
LCS	Level crossing sampling
LPF	Low-pass filter
LSB	Least significant bit
LVT	Low voltage threshold
NBA	National basketball association
NMOS	N-channel Metal Oxide Semiconductor
NTPS	Number of tokens produced per second
PD-SOI	Partially depleted silicon-on-insulator

PMOS	P-channel Metal Oxide Semiconductor
PVT	Process, voltage and temperature
RDF	Random dopant fluctuations
RVT	Regular voltage threshold
SER	Signal to error ratio
SFDR	Spurious-free dynamic range
SNDR	Signal to noise and distortion ratio
SNR	Signal to noise ratio
SR	Set-Reset
UD	Up-down
ULP	Ultra-low power
US	United States of America
UTBB	Ultra-thin body and box
VAD	Voice activity detection
VFC	Voltage to frequency converter
VFS	Full-scale voltage
VR	Variable resolution
ZCD	Zero-crossing detectors

Chapter 1
Introduction

1.1 Background (IoT)

The Internet of Things (IoT) is a communication paradigm that will allow every day's objects to communicate and interact with each other and with the users. Always-on wireless sensors networks will be deployed to gather environmental data, the combination of which will make available an enormous amount of information in a very wide range of applications from medical aids to smart farming.

Figure 1 from [1] shows the different markets/sectors of IoT, also presented in some cases are the number of existing IoT devices and the compound annual growth rate (CAGR) which is a reliable economic measure to quantify the growth of an industry and/or market sector. At the current growth rate, the number of IoT devices for certain applications like smart homes, smart cities and manufacturing can be expected to double over the next five years.

Furthermore, IoT drives the convergence of verticals, which is the process through which different industries increase their bonds and cooperation with one another (for instance, manufacturing and intelligent retail, or smart energy management with smart cities and smart buildings). Overall, the new information gathered thanks to the IoT, will increase the productivity and efficiency across each economic sector, improving our interaction with the environment and our life quality.

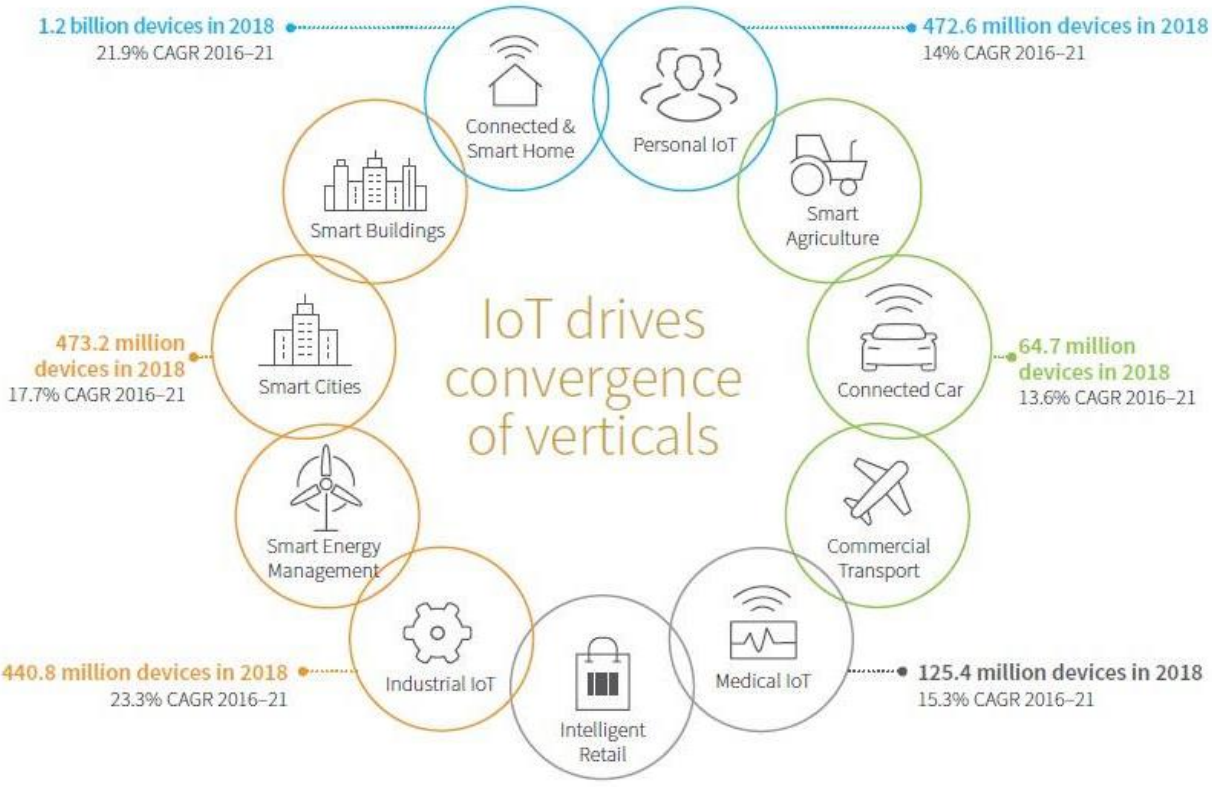


FIGURE 1 DIFFERENT IOT APPLICATIONS, IN SOME CASES THE CAGR OF THE SECTOR IS INDICATED [1]

Figure 2 from [2] shows the market size and the expected growth of the IoT market in North America in billion U.S. dollars. The Figures indicate the contribution of each application sector to the overall market share. By 2022 the market size is expected to be above half a trillion dollars. The main sectors driving the growth of the IoT market are consumer electronics, transportation, and health care.

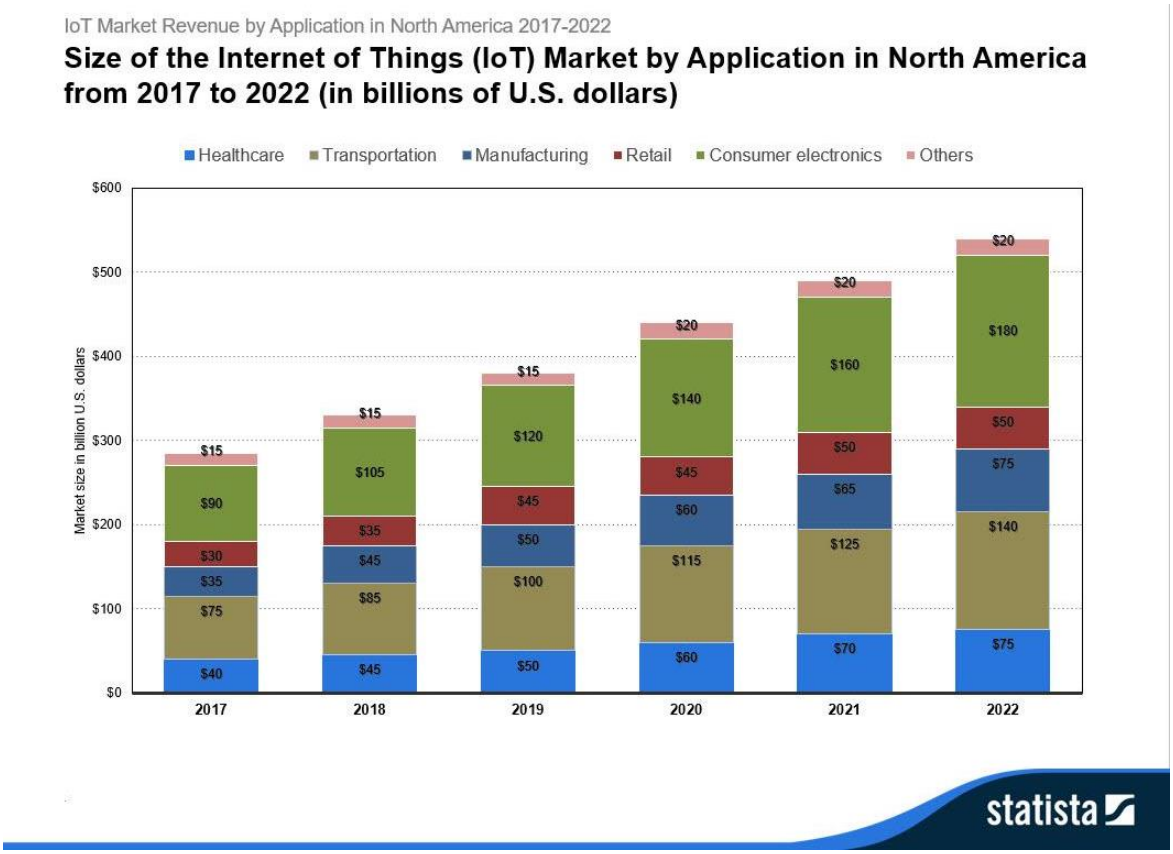


FIGURE 2 EXPECTED GROWTH OF THE IOT MARKET BY APPLICATION IN NORTH AMERICA [2]

1.2 Motivation for this work

IoT demands efficient detection systems, with high configurability capable to operate in wide spectrum of backgrounds, with high levels of integration to operate among other systems, with ultra-low-power consumption which is particularly critical for battery powered systems and with low production cost.

To comply with the demands posed by the implementation of IoT systems several aspects come into play: Circuit design techniques to improve configurability and overall performance of existing systems; novel advanced technologies that offer a larger integration, and improved performance in speed and power consumption; the need for applications in new domains like medical IoT or smart farming that drives innovation by posing new problems, for which existing implementation don't offer an optimal solution. For instance in order to finish the NBA season despite the coronavirus pandemic, the league has instaured the use of monitoring rings for the players and coaching staff [3]. These titanium made rings monitor the heart rate, temperature, sleep and motion patterns, of the individuals wearing it, detecting the symptoms of a coronavirus infection at least 72 hours in advance with a 90% accuracy.

Techniques for data acquisition through analog-to-digital conversion are well studied. There are many implementations for different applications as described by [4].

However, as new applications emerge conventional techniques where the analog input is converted into a stream of uniformly time spaced digital values, may not be best suited to answer the ultra-low-power specifications of IoT applications as further described in the next section. As described by [4] the race of the analog-to-digital conversion has been driven by the need of increasing the signal to noise ratio of signals, however for IoT applications the most critical parameter is power consumption.

In traditional sensors implementations, the device does not embed significant processing units and thus its role is limited to sensing and transmitting the information. The largest contributor to the overall power consumption is thus the transmission of raw data to the remote unit in charge of recognition and signal processing [5]. To drastically reduce the consumption of a battery-powered sensor (which means increasing its lifetime), a new generation of intelligent sensors capable of performing recognition tasks, whatever the background is, is needed. Additional energy reduction is obtained by activating the different processing units when strictly required and by transmitting data only in the presence of relevant information. Thus an important issue is to find the best balance between local and remote processing.

Therefore, the motivation of this work is to explore novel design techniques that may be suitable to implement a new generation of sensors, highly configurable, capable of performing a significant part of the signal processing tasks, while exploiting the benefits offered by advanced technologies. Furthermore, demonstrate the feasibility of an implementation for a given application offering a performance that meets the configurability, low cost and low power consumption required by IoT applications.

1.3 Feature extraction

Sensor data acquisition systems typically use an analog-to-digital converter (ADC) at a constant sampling rate, scaled to the signal bandwidth, to digitize the incoming waveforms for subsequent processing. This approach may however not have optimal efficiency for detection and recognition applications for two reasons. The first one is that many signals (particularly those involve in the monitoring of natural phenomena) have a physical bandwidth which is much higher than the actual information rate. Take the example of speech recognition systems, although audible sounds have a frequency up to 20kHz (one new data each 25µs in a sampled system), the “useful information” rate is less than 10 Hz (the average duration of a syllable is 150 ms) [6]. Thus instead of implementing a system that digitize the audio signal generating and transmitting 40 thousand samples per second to a complex processing unit, we can envision a system that transmits only 10 samples per second. The second reason is that, in many applications, the information is very sparse in time, meaning long time periods without any useful information. Therefore, a complete signal reconstruction may not be required. Instead, specific information called features can be sufficient for further processing [7].

Figure 3 shows the architecture of a feature extraction system. The analog waveform is first processed using a feature enhancing filter that can be implemented in both the analog and digital domains. This block applies a signal processing that increases the difference between signals containing specific characteristics. The next step is to sampled the signal at knowledge/information rate and finally send it to a classification/detection algorithm. The overall performance of a feature extraction systems is evaluated in terms of hit rate, which is the percentage of correct decisions that the system makes for a given signal in a certain scenario.

It is to note that in the case where the feature enhancing filter is implemented in the digital domain, two samplings occur, the first at Nyquist-rate in the feature enhancing block and the second one at information rate to send the information to the feature extraction classifier. Additionally the feature enhancing filter is adapted to the complexity of the classification/detection algorithm.

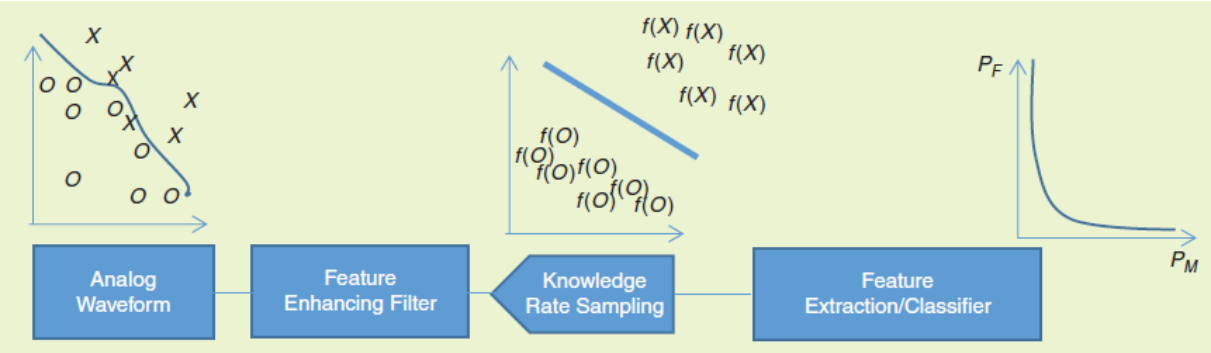


FIGURE 3 ARCHITECTURE OF A FEATURE EXTRACTION SYSTEM [7]

An implementation of a feature extraction system is presented in [8]. The system is a scalable electroencephalogram (EEG) acquisition system which predicts seizure events of people suffering from epilepsy. The task is to differentiate between normal neural spikes employed to move a limb and abnormal spikes that will trigger a seizure episode. To perform this task, it is required to extract the spectral content analysis of neural spikes, in effect spikes that will trigger a seizure episode and normal neural spikes have a different spectral content.

A bank of filters is used to generate a feature vector that is latter employed to classify the signals accordingly.

Feature extractions systems are currently implemented in both the analog and the digital domains. Digital implementations occupy a smaller silicon area [9], have high configurability [10], and may easily be implemented in advanced technology nodes. Analog implementations have a lower power consumption [11], [12] but lack configurability, occupy a large silicon area and are not easily integrated into advanced technology nodes.

Since, each approach has advantages and drawbacks that are complementary an alternative approach capable of combing the best attributes of both analog and

digital systems is needed. A possible alternative is continuous-time digital signal processing systems (CT-DSP) and will be presented in the next chapter.

1.4 Thesis contributions

The framework of this work is the implementation of a CT-DSP system applied for the feature extraction of audio signals. Thus, the main contributions of this work are the following:

- A study of the implementation requirements of an audio feature recognition application [13].
- A novel architecture inspired by the human cochlea to generate spectrograms of an audio signal using CT-DSP.
- A new technique based on glitch elimination and delta encoding to enable the cascade of CT-DSP finite impulse response (FIR) filters.
- A new CT-DSP block that calculates the instantaneous power of a signal and integrates it over a time window to estimate the corresponding energy, needed to create spectrograms [Patent filed].
- A delay element in 28nm FD-SOI technology, exhibiting a record tuning range between 30ns – 97µs and a dynamic power consumption below 15fJ for a VDD voltage of 0.7 V [14].

As a summary, this work gives a significant step forward towards the implementation of CT-DSP systems for audio feature extraction applications. System-level performance estimations based on the measured delay values show the system could be able to process typical speech with an event-drive power consumption below 2µW in range of existing analog implementations and a silicon area inferior to 1.5mm² in range of existing digital implementations.

The principles presented can be used for any other detection or recognition application that demands the extraction of spectral content information.

1.5 Thesis Layout

The rest of the thesis is organized as follows. *Chapter 2* introduces the theoretical concepts required for understanding CT-DSP systems, such as signal quantization, aliasing, and sampling. For this purpose, a comparison between continuous-time (CT) and discrete-time (DT) systems is presented. Next, the state-of-the-art of existing CT systems is detailed and advantages and challenges are identified. This analysis concludes that the implementation of CT-DSP systems is particularly interesting for applications in which signal activity is sparse in time, with a varying spectral content. The research directions to meet the challenges of implementing and improving the performance of a CT-DSP system are traced.

In *Chapter 3* we show that the recognition and detection of audio signals could benefit from a CT-DSP implementation. The basic characteristics of audio signals and

the functioning of human hearing are presented. Audio recognition applications require a unit to extract the spectral content (spectrogram) of the signal. For low-complexity recognition tasks, like voice activity detection (VAD), the unit that generates spectrograms consumes most of the system resources. Thus, the state-of-the-art of existing solutions to generate spectrograms, such as Fast-Fourier transforms (FFT), analog filter banks and digital filter banks, is presented. Since the advantages and disadvantages of each approach are complementary, a new approach encompassing the positive aspects and overcoming the drawbacks of existing implementations is proposed.

Chapter 4 introduces the proposed system to generate spectrograms of an audio signal using CT-DSP. The system requirements and the architectural choices made to size the system are presented. Novel approaches to overcome the challenges of CT systems, such as the cascade of CT FIR filters and the estimation of the instantaneous power of a signal are detailed. Delay lines are a key element of CT-DSP systems, representing approximately 70% of the total system power consumption and silicon area for the envisaged application. The design of ultra-low-power and low area digital delay elements is therefore required. The requirements on the digital delay element are detailed and an estimation of the overall system performance for the targeted delay line requirements is derived. Additionally a novel alternative approach that combines cascaded low-pass filters and energy estimation circuits to create equivalent and band-pass filter functions is presented.

Chapter 5 describes the design of a novel digital delay element in 28nm CMOS FD-SOI technology by STMicroelectronics, compatible with the system-level constraints detailed in Chapter 4. This has been possible thanks to the specific properties of the FD-SOI technology. A detailed description of implemented solutions to improve the performances of the digital delay element are presented. Finally, the design of an integrated circuit containing individual delay structures for single delay characterization and a delay line with several hundreds of cascaded delay elements is presented.

Chapter 6 presents the measurements setup and results of the integrated circuit. Results are consistent with simulations. System performance estimations based on the measured delay performance prove that a CT implementation of a filter bank to generate spectrograms is an alternative to existing approaches, which offers performance in line with state-of-the-art implementations when processing a human voice and benefiting from the event driven nature of the solution to further reduce the overall power consumption.

Chapter 7 concludes this work by highlighting the main aspects of this research work and presents perspectives and possible directions for future work.

Chapter 2
Continuous-Time Digital Signal Processing Systems

2.1 Description

This chapter starts with discussing the fundamental differences between CT and DT analog to digital conversion and their impact on the system. Next, the principles, advantages, and challenges of CT-DSP will be presented. The state of the art will be reviewed to identify the approaches that were taken, and the reasoning behind the different architectural choices. We will conclude this chapter by identifying the critical points that need to be addressed to improve the performance of CT-DSP systems.

2.2 Introduction and context

Shown in Figure 4 are the four different types of electronics systems, which correspond to the different types of signals with which they operate. Signals can either have a continuous amplitude (CA) or discrete amplitude (DA) and they can be either continuous in time (CT) or discrete in time (DT). Thus, on the bottom left, conventional digital signal processing systems are discrete in time and discrete in amplitude (DT DA). Conventional analog systems, which are continuous in time and continuous in amplitude (CT CA). Systems discrete in time and continuous in amplitude (DT CA), whose best-known example is switched-capacitor circuits. Finally, systems continuous in time and discrete in amplitude (CT DA) which seek to combine the attributes of both DT DA and CT CA are the focus of this chapter [15], [16].

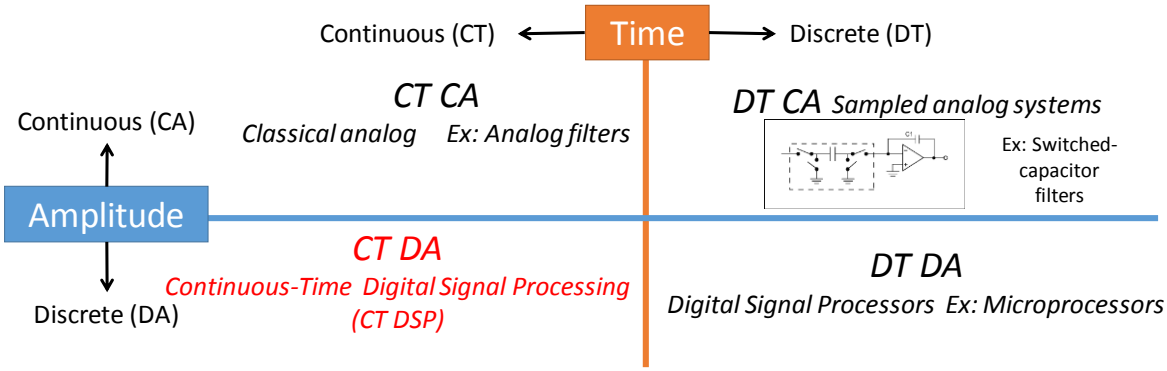


FIGURE 4 QUADRANT SHOWING THE DIFFERENT TYPES OF ELECTRONIC SIGNAL PROCESSING

On top of this classification, each category has its own specificities. Conventional digital design is automated allowing the design of more complex systems, with higher configurability and flexibility than conventional analog systems. Digital circuits are also less affected by component noise, component tolerances and device matching than analog circuits. However digital signal processing suffers from the effects of aliasing and quantization noise. In DT-DSP the power consumption due to clock distribution does not scale with the signal activity, unless some important changes are made on an architectural level, and which can be a penalty in periods of low or no activity. Continuous-time digital signal processing (CT-DSP) is a technique in which signals are continuous in time but discrete in amplitude, allowing the system to advantageously combine the attributes of digital and analog systems. In a CT DSP system samples are only generated when there is activity and consumption adapts dynamically to the content of the signal. Due to the absence of a sampling clock, no signal aliasing occurs,

thus increasing the signal-to-noise ratio [17]. The digital nature of the processing provides configurability and robustness [18].

It has been demonstrated that CT-DSP systems have a power consumption that scales automatically to the input signal activity and keeps the advantages of digital signal processing like configurability and integration with advanced node technologies. In the following section, the principles behind analog to digital conversion (ADC) will be discussed, as well as the main differences between DT and CT systems [16].

2.3 CT vs DT ADC conversion

Analog-to-digital converters (ADCs) are essential building blocks in modern electronic systems. They translate analog quantities, which are characteristic of most phenomena (temperature, pressure sound, optical measures) in the “real world” to digital language, used in signal-processing, computing, data transmission and control systems functions. In DT DA systems, the analog-to-digital conversion is done in three steps: time discretization (sampling), amplitude quantization and digital coding when in CT DA only quantization and coding are required. In the scope of this work we will focus only on amplitude quantization and time discretization aspects.

2.3.1 Quantization

The quantization consists in converting an analog signal into a finite sequence of discrete states/values (quantization steps). The distance between two successive quantization steps is the Least-Significant-Bit (LSB) value, which is the smallest value the quantifier can detect.

The selected step is coded into a digital word. The ADC resolution is the number of bits (N) needed to code the output (and thus the number of steps is 2^{N}). The full-scale voltage (VFS) which is the maximum signal an ADC can handle, the ADC resolution N and LSB are linked with equation (1)

$$\mathrm{LSB} = {}^{\mathrm{VFS}}/_{2^{\mathrm{N}}} \tag{1}$$

The quantization of a signal adds an error, which is equal to the difference between the input signal and the quantized signal. Figure 5 shows the quantization error of an ideal N-bit ADC. The maximum quantization error corresponds to ½ LSB.

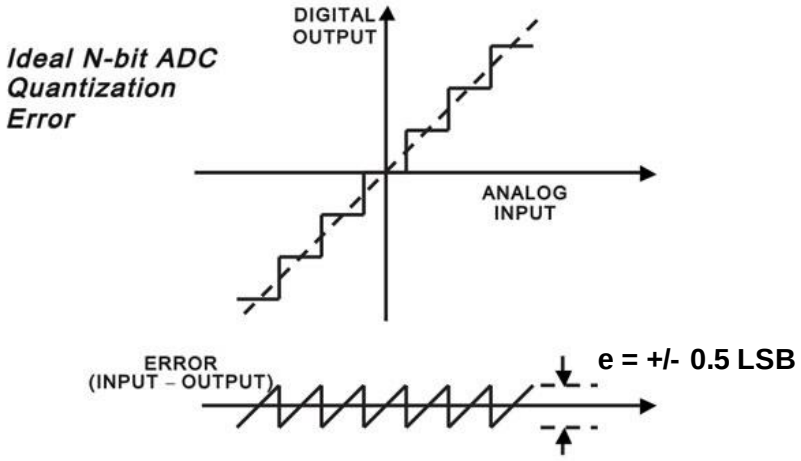


FIGURE 5 IDEAL N-BIT ADC QUANTIZATION ERROR

Figure 6 shows the discrete-time quantization of a sine wave, the time domain waveforms of the input sine wave (a), the quantized signal (c), the quantization error (b) and their corresponding frequency spectrums. In this case, the quantization error waveform can be divided into a sawtooth-like waveform e_{SAW} and a bell-shaped waveform e_{BELL}. The sawtooth-like error occurs during fast portions of the input when quantification levels are traversed quickly. The bell-like error occurs during slow portions of the input, for a sinusoidal input, at its peaks. Figure 6 shows that the quantized signal is considered as the sum of the original signal and an error signal. Thus, the frequency spectrum has two components: the first one corresponds to the sinusoidal wave and the second one corresponds to the error signal.

The slow varying bell shape error contributes to low-frequency distortion spurs, whereas the fast-varying saw-tooth error generates high frequency distortion spurs. The highest distortion spur occurs at approximately the saw tooth frequency f_{SAW}. If we consider the sawtooth part of the error waveform, the minimum duration is the time it takes a sinusoidal input of the form $x = A\sin(2\pi f_{IN} t)$, to cross the interval equal to the LSB [19]. Thus, the sawtooth frequency f_{SAW} is approximately equal to:

$$f_{SAW} = \frac{1}{T_{SAW,min}} = \frac{2\pi A f_{IN}}{LSB} \tag{2}$$

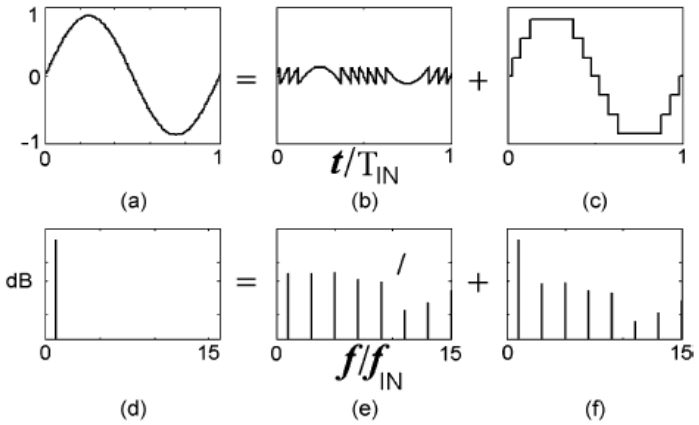


FIGURE 6 EXAMPLE WAVEFORMS OF A 3-B ADC QUANTIFICATION. TIME DOMAIN WAVEFORMS ARE NORMALIZED TO THE INPUT PERIOD, OF THE (A) INPUT SINUSOID, (B) QUANTIZATION ERROR, (C) QUANTIZED INPUT AND THEIR CORRESPONDING FREQUENCY SPECTRUMS, IN (D), (E), AND (F), RESPECTIVELY [20]

As the resolution is increased, the quantization level or LSB is reduced, and thus the spectral components of the spur distortion are reduced in amplitude and shifted away from the frequency spectral components of the input signal; the results in [21], show that the amplitude of the spur decreases by 9dB for each additional bit of resolution. Table 1 shows the spurious-free dynamic range (SFDR) which is the strength ratio of the fundamental signal to the strongest spurious signal, for different ADC resolutions, the harmonic that is the strongest spurious is also indicated.

TABLE 1 LARGEST HARMONIC AND SFDR FOR DIFFERENT BIT RESOLUTIONS

Number of Bits	Larger Harmonic	SFDR
4	H47	-35 dBc
5	H99	-45 dBc
6	H195	-51 dBc
8	H795	-67 dBc
10	H3201	-84 dBc
12	H12850	-100 dBc

2.3.2 Sampling

For DT-ADCs, samples are generated at a constant rate dictated by the Nyquist theorem. The Nyquist theorem imposes that to properly reconstruct a signal, the sample rate must be least twice the highest frequency component present in the signal. Figure 7 shows a signal with a limited bandwidth in the time domain and its frequency spectrum; also shown is the sampled signal and its frequency spectrum. Sampling in time corresponds to replication in the frequency domain. If a signal bandwidth $f_B \geq \frac{f_S}{2}$ where f_s is the sampling frequency, then the replica spectrum will overlap the useful spectrum causing loss of information. This mechanism is called aliasing.

Aliasing has two impacts:

- Need of anti-alias filter before the ADC, in the receiver if the spectrum of the analog filter is not limited.
- Overlap of the ADC quantization noise with the useful signal. Reducing the signal to noise ratio, in the [-Fs/2, Fs/2] band.

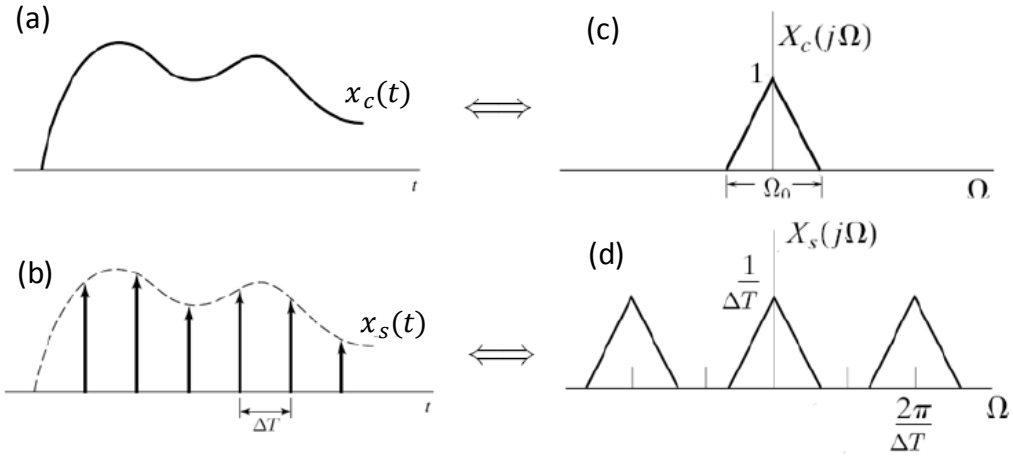


FIGURE 7 EXAMPLE WAVEFORMS OF A SAMPLED SIGNAL. TIME DOMAIN WAVEFORMS, OF THE (A) INPUT SIGNAL CONTINUOUS IN TIME, (B) SAMPLED INPUT SIGNAL AND THEIR CORRESPONDING FREQUENCY SPECTRUMS, IN (C), AND (D).

2.3.3 Comparison between CT and DT ADC conversion

Figure 8 shows the comparison between CT and DT ADC conversion, the main difference between them is that for DT ADC conversion both quantization and sampling are required, while for CT ADC conversion only quantization is required. Since sampling is not required in CT, aliasing does not occurs and quantized signals have a cleaner spectrum with a larger SNDR for the same ADC resolution.

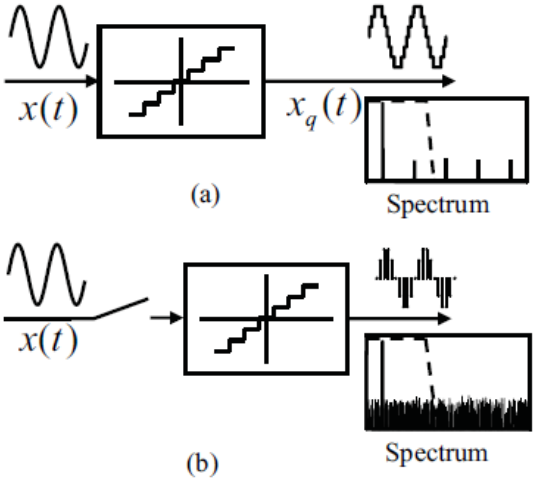


FIGURE 8 INPUT WAVEFORM, OUTPUT WAVEFORM, AND OUTPUT SPECTRUM WITH (A) QUANTIZATION ONLY, (B) SAMPLING AND QUANTIZATION [18]

2.4 CT-ADC State of the art

2.4.1 Level crossing Analog-to-Digital converter

Level crossing quantification is a method used to digitize an analog signal in continuous time, where a new sample is only generated when the analog input crosses a quantization threshold, resulting in a sample generation rate which depends on the analog input [22]. A common implementation of a CT ADC using a level crossing sampling (LCS) scheme is shown in Figure 9 that is implemented using asynchronous delta modulators. Two comparators compare the input signal to two signals V_{HIGH} and V_{LOW}, generated by a feedback Digital-to-analog-converter (DAC). The DAC can generate 2^N quantization levels, where *N* is the ADC resolution. V_{HIGH} and V_{LOW}, are the consecutive levels surrounding the current input.

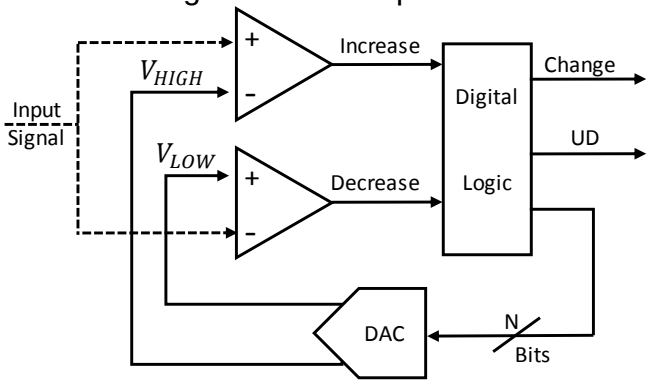


FIGURE 9 LEVEL CROSSING SAMPLING SCHEME

When the input signal increases or decreases its value beyond the comparator's thresholds, the corresponding comparator changes its output from 0 to 1, this change is detected by a control logic circuit that generates a pair of two binary signals CHANGE and UD. These signals are fed to the DAC to update the V_{HIGH} and V_{LOW}, signals. The input signal is again within the range given by V_{HIGH} and V_{LOW} The change signal is a pulse, whose values changes to 1 each time that a comparison level has been crossed; UD is a signal whose value goes to 1 if V_{HIGH} has been crossed, and 0 if V_{LOW} has been crossed [18], [22]–[24]. To properly track the input, the time required to update V_{HIGH} and V_{LOW} must be inferior to T_{gran} which is the maximum speed at which the input signal can cross consecutive comparison levels, this value is given by

$$T_{gran} = \frac{1}{2^N \pi f_{INPUT}} \tag{3}$$

Where f_{INPUT} is the maximum input frequency and *N* is the resolution of the ADC. T_{gran} is also used to dimension the delay taps needed to implement the CT DSP FIR filters. [22].

The bandwidth of the ADC is defined as the maximum input frequency for which the system is able of tracking the input signal. In CT this value depends on the ADC resolution.

As in DT ADC the in-band signal-to-noise-and-distortion ratio (SNDR) of a CT ADC is not constant. Due to amplitude quantization harmonics are generated; the SNDR will depend on how many harmonics fall in band. Let's consider the frequency spectrum of a system whose useful bandwidth is 10 kHz plotted in Figure 10 [24]. In Figure 10-(a) the input signal is a 200 Hz full-scale sinusoid wave, fifty-two harmonics fall in-band and the measured SNDR is 47 dB. In Figure 10-(b) the input signal is a 4 kHz full-scale sinusoid wave, since only one harmonic falls in-band the measured SNDR increases to 62 dB.

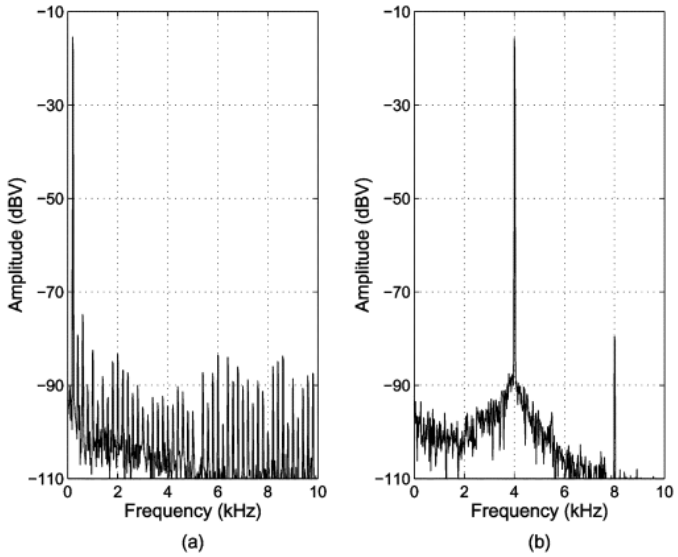


FIGURE 10 OUTPUT SPECTRA OF A CT SYSTEM CONFIGURED AS A 10KHZ LOW-PASS FILTER, GIVEN (A) A FULL-SCALE 200HZ SINUSOID INPUT AND (B) A FULL-SCALE 4 KHZ SINUSOIDAL INPUT [24]

2.4.2 CT Level crossing sampling limitations

To properly track the input and avoid information losses [25], the loop delay in an asynchronous delta modulator needs to be smaller than the minimum time between two consecutive level crossings (T_{gran}). The delay T_{gran} needs to be further divided between the comparators, the digital logic, and the feedback DAC. The bandwidth of a CT level crossing ADC is restrained by the blocks composing the feed-back loop, limiting its implementation in high frequency systems, this problem is particularly important at high ADC resolution values.

Given that the number of samples generated by a level crossing ADC increases exponentially with the resolution *N*, for resolutions above 8 bits a very large number of samples is generated. The large number of samples also has a high impact on the power consumption of the signal processing system, making the resolutions superior to 8 bits, unpractical in terms of energy consumption and surface.

Different alternatives have been proposed: In [26] per edge encoding is used to create a 3-bit ADC that operates in the GHz rate. In [27] the use of a 1-bit DAC to reduce the power consumption of the ADC loop is proposed, optimizing the ADC for very low frequency applications. The main optimization effort has focused on reducing the number of samples generated by a CT ADC, either by using signal dependent variable resolution to process the fast-varying inputs at low resolutions or by using an integrator before the comparators. These approaches are detailed below.

2.4.3 Variable resolution level crossing ADC

Figure 11 from [20] shows the quantization error waveform for a sinusoidal input and its corresponding spectrum normalized to the input frequency. We can distinguish three parts: the baseband corresponding to the input signal spectrum, the bell portion corresponding to the quantization error of the slow varying parts of the sine wave, and the sawtooth that corresponds to the quantization error of the fast varying parts of the input.

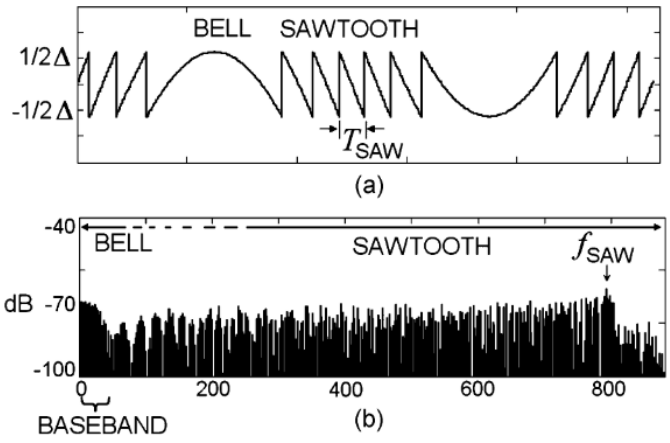


FIGURE 11 SIX BIT QUANTIZER WAVEFORMS FOR A SINUSOIDAL INPUT. (A) QUANTIZATION ERROR WAVEFORM, NORMALIZED TO THE QUANTIZATION LEVEL, (B) SPECTRUM OF THE QUANTIZATION ERROR, NORMALIZED TO THE INPUT FREQUENCY

As stated in section 2.3.1 by increasing the ADC resolution the spectral error is shifted away from the baseband. However, the fast varying parts of the input can be quantized with less resolution without degrading the in-band performance. Thus, a variable resolution (VR) ADC has been proposed by [20] . The principle of the VR ADC

is to reduce power dissipation and ease the hardware timing requirements of the system by aggressively reducing the resolution for fast inputs, according to the input slope. This is done without significantly affecting the in-band error since a higher frequency low-resolution error does not alias into the baseband, in contrast to what would have been the case with discrete-time adaptive resolution systems [20]. VR CT ADC in conjunction with CT DSP offers not only a significant reduction in processor power dissipation but also a reduction in the processor size, achieving high in-band SNDR with lower effective resolution.

A VR CT ADC has been implemented by [28], and the functioning principle is the following. After a sample event, the ADC is placed in its lowest resolution state with V_L and V_H widely separated. A controller then uses timing from a delay line, placed at the output of the comparators, to progressively narrow the interval as shown in Figure 12(a). Inputs with low slope, when they finally cross a level, are quantized with high resolution, whereas high-slope signals hit a comparison level before the comparison window has narrowed, and are thus quantized with coarse resolution Figure 12 (b),(c). This scheme has a negligible impact in signal-to-noise-ratio (SNR) and reduces the power consumption of the zero-crossing detectors (ZCD) used as comparators. As a downfall the complexity and surface of the circuit is increased. Additionally the VR CT ADC produces multi-bit output samples that require a more complex processing than the outputs generated by an ADC with a constant resolution.

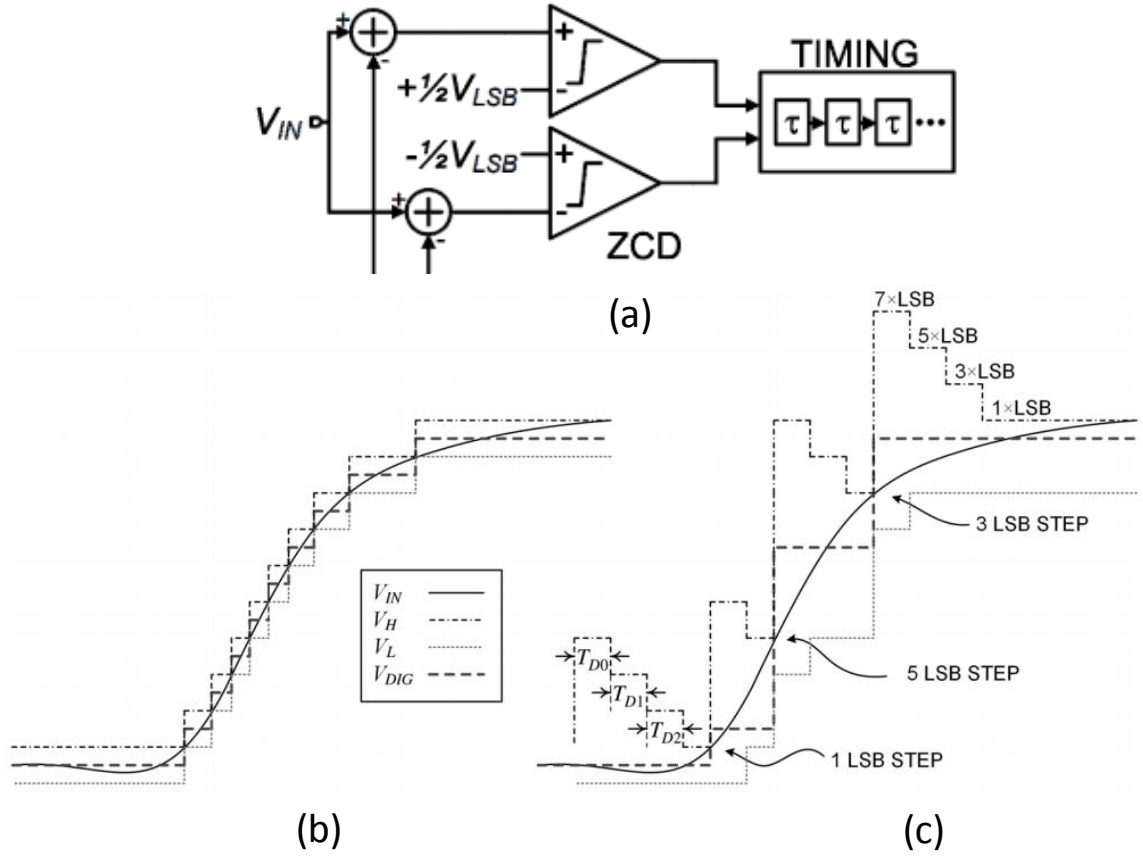


FIGURE 12 VARIABLE RESOLUTION CT ADC PRINCIPLE (A) TIME-VARYING COMPARISSON WINDOW CIRCUIT [28] (B)(C) INPUT SIGNAL (SOLID LINE), UPPER AND LOWER COMPARISON LEVELS (DOT-DASHED AND DOTTED LINES RESPECTIVELY), AND QUANTIZED SIGNAL (BROKEN LINE), FOR (B) CONVENTIONAL LEVEL CROSSING ADC, AND (C) ADAPTIE-RESOLUTION LEVEL-CROSSING ADC [22]

2.4.4 Integrator embedded ADC

In [23] an integrator embedded CT ADC is proposed. The model is shown in Figure 13. The input signal is fed to an integrator, next the output of the integrator is fed to the comparators that generate the increase and decrease signals. The main trait of this architecture is that the number of tokens generated per second given by (4), is independent of the input frequency; rather it increases with the input amplitude. The number of tokens produced per second (NTPS), for this ADC depends on f_0 the bandwidth of the ADC and V_c which is the comparator threshold.

$$NTPS = f_{in} \times \left[\frac{2A_{q,p-p}}{V_{LSB}}\right] \approx \left[\frac{f_0 \times A_{in,p-p}}{V_C}\right] \quad (4)$$

The circuit partially behaves like a voltage-to-frequency converter (VFC). However, unlike a VFC, the ADC produces no pulses when the input is zero. The main advantages are a lower loop delay, a lower NTPS rate and a very low power consumption compared with other CT ADC.

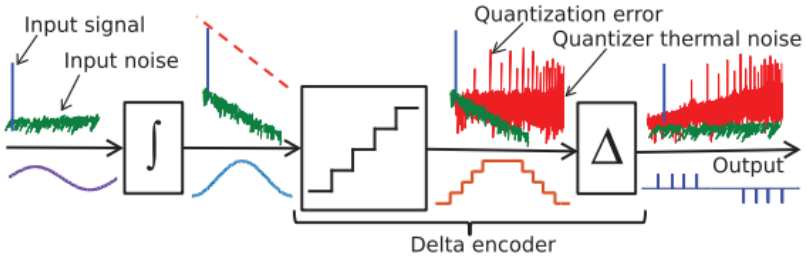


FIGURE 13 MODEL OF THE ADC WITH EMBEDDED INTEGRATOR [23]

2.5 Overall review of CT ADC

TABLE 2 COMPARISON BETWEEN DT AND CT ADC CONVERSION

Parameter	Conventional ADC	CT ADC
Pre-filter required?	Yes (anti-aliasing)	Yes (overslope prevention)
Event rate	Uniform at Nyquist Rate, independent of ADC resolution	Event driven, proportional to 2^N (N -bit ADC)
SNDR	6.02n + 1.76 dB	> 6.02n + 1.76
Power Consumption	Constant	Scales automatically to input signal activity
Main Problems	- Samples are generated independently from signal activity - Aliasing degrades the SNDR	- Sample number increases exponentially with ADC resolution - Feedback ADC architecture inappropriate for high frequency ADCs
Best suit for	Continuous signal streams with high SNDR requirements	Signals sparse in time and variable spectral content

2.6 Digital signal processing in CT

2.6.1 CT Receiver – General considerations

As stated in [23], if we compare CT receivers as existing in State of the Art, they exhibit two major advantages compared to classical receivers:

- No need for antialiasing filter
- No need for a "fast" (Nyquist rate) clock

This is because, in CT ADCs, signals are sampled without a clock. Thus, the spectrum of the resulting output is alias free and contains only components related to the spectrum of the input signal with no quantization noise floor, and with signal to noise and distortion ratio (SNDR) higher than that of conventional systems with the same number of bits [29].

However, a major challenge is to interface CT DSP blocks with the rest of the system which is, by nature, DT digital. In existing implementations, this is done using a continuous time digital to analog converter (CT DAC) followed by a discrete time analog to digital converter (DT ADC).

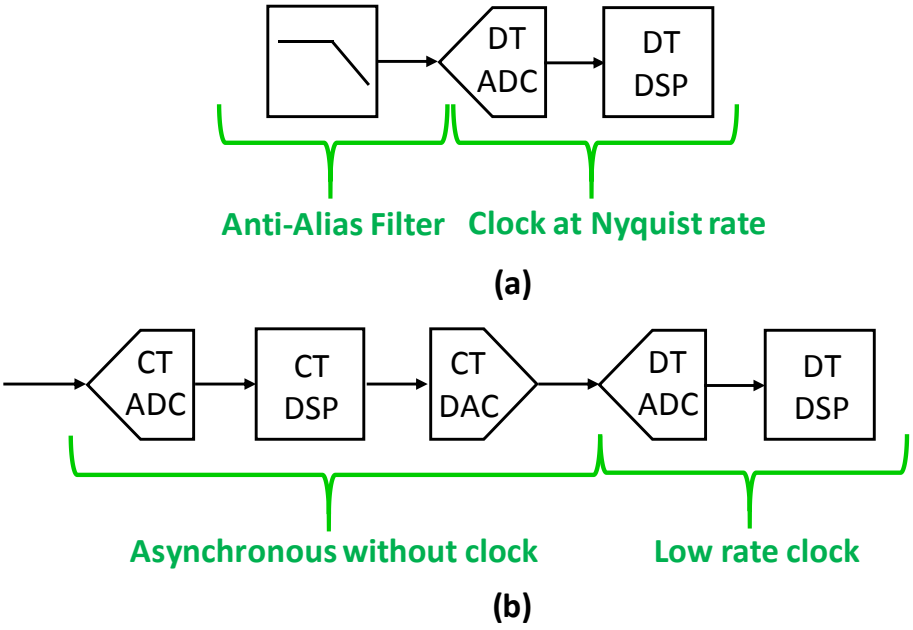


FIGURE 14 CLASSICAL RECEIVER ARCHITECTURE (A) CT RECEIVER ARCHITECTURE (B)

2.6.2 FIR FILTERS

With the exception of [17] all the CT DSP functions that have been demonstrated are finite impulse response (FIR filters). FIR filters are the linear combination of the delayed inputs weighted by a set of coefficients. FIR filters in continuous time are implemented using delays, accumulator-multipliers and adders operating in continuous time. Multipliers and adders are implemented using asynchronous digital circuitry; the delays, consists of a cascade of digital delays elements operating without a clock [18], [24]

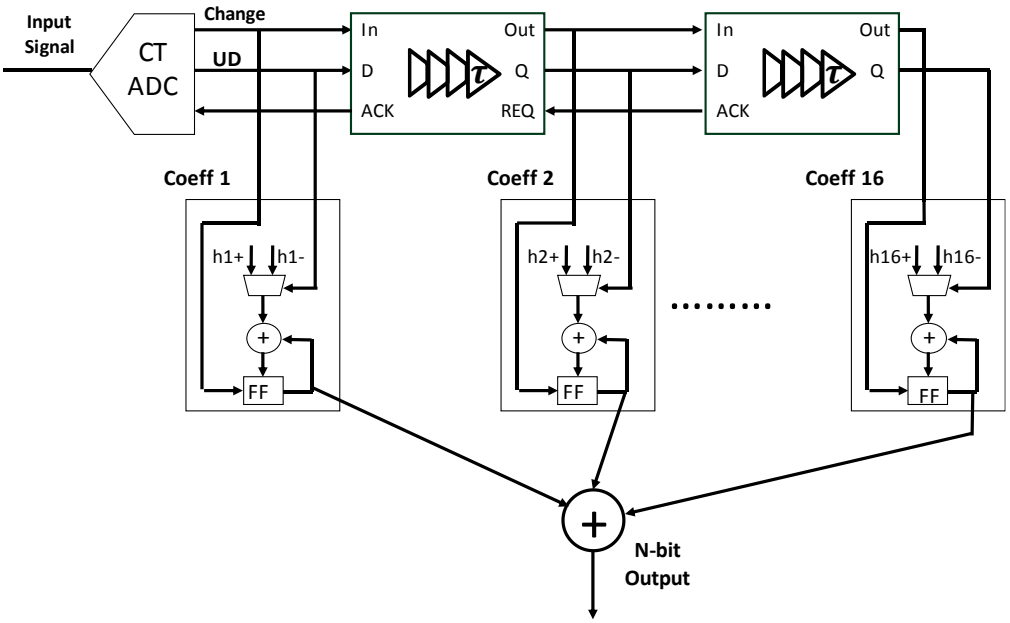


FIGURE 15 CT DSP FIR FILTER SCHEME

CT FIR filters implementation is simple because, due to the fact that CT DSP signals vary only one bit at the time, adders/accumulators can replace the multipliers, as shown by (5), where y[n] is the output of the multipliers, x[n] is the input signal and h_k the coefficient value.

$$\begin{aligned} y[n] &= (x[n-1] \pm 1) \times h_k \\ &= x[n-1] \times h_k \pm h_k \\ &= y[n-1] \pm h_k \end{aligned} \tag{5}$$

The output of a CT Filter also varies only one bit at the time. Due to the absence of a clock CT DSP FIR filters, require power-hungry and area consuming delay lines, their number being equal to the order of the filter. The use of delays has limited the order of the FIR filters implemented in CT. 15 is the highest order FIR filter in CT [24].

2.6.3 CT DELAY TAP

As part of the operation of the FIR filter, the pairs of Change and UD signals generated by the CT ADC known as tokens are delayed by CT delay taps of delay time T_D. The value of T_D is chosen based on the desired frequency response of the system. Because of the use of tap delays, the frequency response of the CT filter is a function of $e^{j\omega T_D}$ and is periodic in the frequency domain. For instance, if the CT filters is configured as a low pass filter, it has repetitive passbands around every $1/T_D$. Which is shown in Figure 16.

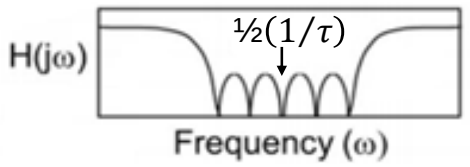


FIGURE 16 THE FREQUENCY RESPONSE OF A CT FILTER IS REPETITIVE AROUND $1/T_D$

The delay tap must be able to accommodate all the tokens generated by the ADC, at any given time [24]. Thus, the total delay T_D is composed of a series of granular delays, whose number is given by T_D / T_{GRAN} Figure 17.

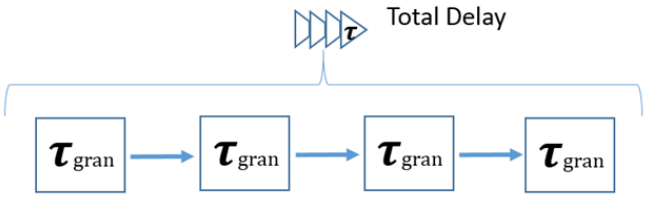


FIGURE 17 STRUCTURE OF DELAY TAPS IN CONTINUOUS TIME

2.6.4 IIR FILTER

In [17] an Infinite Impulse Response (IIR) filter was implemented in CT. An IIR filter is the weighted combination of the previous inputs and the previous outputs. Thus, the main implementation challenge is to prevent parasitic oscillations, from the feedback loops that can lead to instability. Asynchronous logic circuits have been used to alleviate the mismatch of the feedback loops and to synchronize the pulses. A 6th order IIR has been implemented. Due to instability issues IIR filters are not configurable, however compared with a FIR filter of the same order, their stop-band rejection and transition bandwidth are far superior.

2.7 CT DAC

In literature we can find the use of CT DAC to convert the N-bit output of a CT filter, into an analog signal, in this case the signal is not passed to the output DAC as a parallel word. Instead, the N-bit output is exanimated to create a token consisting of two signals ODACChange and ODACUD similar to those generated by the ADC. Given that the filter output only varies one LSB at a time, the DAC will only have to change its output voltage one step at the time. An r-string DAC is used for this purpose. The DAC used in the feedback loop of the CT ADC works with the same principle [24].

2.8 Continuous time to discrete time converter.

Since the vast majority of signal processing systems are discrete in time, for some applications the interaction between CT and DT systems may be required. Thus a processing circuit to convert the CT samples into a DT samples is needed.

Algorithms like Akima, Cubic spline have been proposed to solve this problem. However, these algorithms require a very large computational complexity and above all they require to know the time between samples, thus time coding is required [30].

Another alternative is proposed in [17] where an interpolator employing FIR filters is used to suppress out-of-band distortion and noise power. Given that the filter response is repetitive at multiples of $1/T_D$, the repetitive passband is push away to $16/T_D$. After the interpolator, a conventional DT ADC is used to sample the signal.

2.9 Performance assessment

Figure 18 shows the power consumption of a complete CT system presented in [24] when processing a speech signal. The system consists of an ADC working at an 8 bits resolution, a 16th order filter whose frequency response is sampled at 20 kHz and a DAC.

We can observe that although CT systems have an event driven power consumption that scales with the signal activity, two challenges arise. The first is the high energy consumption during the activity periods, that easily reaches 600μW. The second and most important is the static power consumption that dominates periods of non-activity. As we can see for the system presented in [24], this value is above 200μW, significantly larger than the power consumption of state-of-the-art circuits that also process speech signals [11],[12].

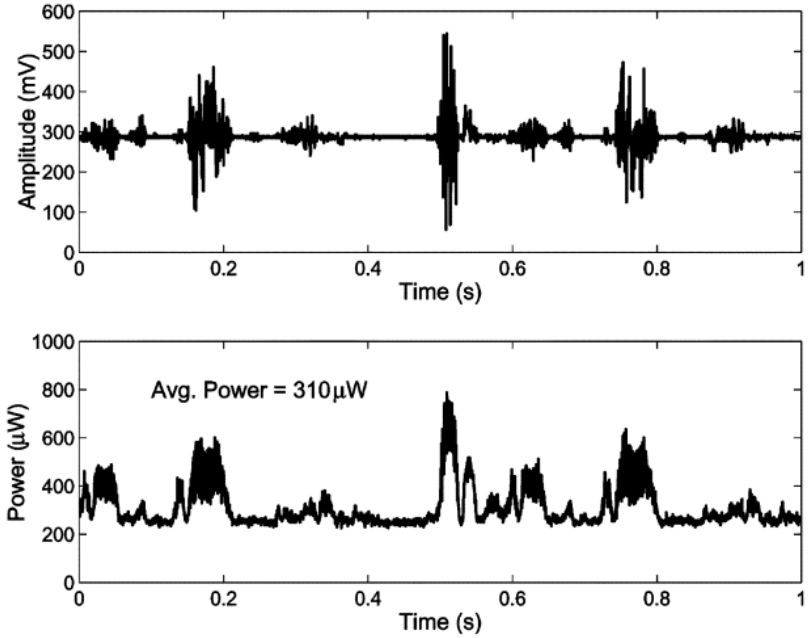


FIGURE 18 PLOT OF INPUT SPEECH SIGNAL (TOP) AND THE INSTANTANEOUS POWER CONSUMPTION (BOTTOM) OF THE CT ADC/DSP/DAC FROM [24]

2.10 Challenges

Here is a list with the main challenges for an efficient implementation of a CT DSP system:

1. **Implementation of efficient ADC architectures.** Current CT ADC architectures are limited in terms of resolution and maximum input frequency. The minimum time between samples T_{gran}, imposes a heavy burden on the circuitry. In addition, the large number of samples generated, imposes a large constraint on the digital processing of CT systems. Variable resolution CT ADC have shown promising results in terms of reducing the number of samples generated, however their performance in terms of power consumption and surface still lags behind that of their discrete time counterparts. In the other hand although integrator embedded ADCs have shown interesting performance, as the number of samples no longer depends on the input frequency of the signal.
2. **The implementation of CT functions other than FIR and IIR filters,** to this day the only functions implemented in CT are FIR and IIR filters. This severely limits the domains of application of CT DSP systems.
3. **Efficient implementation of high order FIR filters.** CT DSP FIR filters require power-hungry and area consuming delay lines, their number being proportional to the order of the filter. At the time of writing, a 15th order CT DSP FIR filter is the most complex CT-FIR demonstrated which limits the potential applications

of CT DSP filters. However, the implementation cost of a CT FIR filter can be reduced if an efficient method to cascade filters is demonstrated.

4. **Efficient granular delay elements.** Granular delay elements are responsible for almost 70% of both power consumption and surface occupation in a CT DSP system. Efficient delay elements with a large tuning range are fundamental to improve the performance of CT systems.
5. **Interface with a synchronous world.** Currently the only feasible interface with a synchronous world is to reconstruct the output of a CT Filter with a CT DAC, and to resample the signal with a DT ADC. Therefore, CT systems are limited to function as analog filters. An efficient interface with a synchronous world could open the door for the use of CT systems in new applications.

2.11 Conclusion

In this chapter, we have studied the basic principles of analog-to-digital conversion, amplitude quantization and sampling. CT systems offer interesting properties because signals are only quantized in amplitude. We explained the principles of continuous-time systems, and we have studied each building block that constitutes them. Here are some of the main points to retain:

1. CT ADC are based on a LCS architecture, in which comparators constantly track the input signal, and generate a sample each time that a level has been crossed.
2. CT ADC generates a very large number of samples, which limits their resolution and their maximum input frequency. CT ADC with variable resolution or with an embedded integrator, have been studied and show significant improvements in power consumption
3. Further study is required to quantify the impact of integrator embedded ADC, in which the number of samples generated no longer depends on the input frequency.
4. Low order FIR filters and a 6^{th} order IIR filter are the only functions that have been implemented in CT. Therefore, the current CT systems is limited to the replacement of analog filters.
5. Given the absence of a clock, CT DSP requires the use of energy hungry delay lines.
6. An efficient interface with a synchronous world still required.

Chapter 2. Continuous-Time Digital Signal Processing Systems

Chapter 3
Feature Extraction for audio signals

3.1 Description

In previous chapters, we introduced the general concepts of feature extraction and continuous-time digital signal processing (CT DSP) systems. We concluded that CT DSP systems are best fitted, for the processing of signals whose activity is sparse in time, and whose frequency content varies drastically over time. These characteristics are met by audio signals. We will begin this chapter, by discussing about the different applications that require audio feature extraction like voice activity detection (VAD) and automatic speech recognition (ASR). We will discuss the characteristics of audio signals and the functioning of human hearing, by which most audio recognition applications are inspired. Next, we will continue by discussing the state of the art of audio feature extraction, in both analog and digital domains. We will conclude with a discussion about the advantages and limitations of each of these approaches.

3.2 Audio feature extraction applications

Thanks to recent technological advances, voice control and speech recognition applications have transformed, the way users interact with their devices and have gained huge popularity over the last years [31]. Systems employing speech recognition can be found in retail, attraction parks, smart homes, smart cars, etc.

Audio recognition applications are not only limited to voice or speech recognition. For instance in war zones the American army deploys ultra-low power audio sensors to detect the presence of certain vehicles like tanks or trucks, that are approaching a military base or strategic position [32].

The conventional architecture of audio recognition systems is shown in Figure 19. When the amplitude of the input signal is larger than a given threshold, a circuit used to monitor the signal activity generates a wake-up signal. Once, the wake-up signal has been generated, the system begins to generate a large amount of raw data, which is digitized and either processed locally or sent via a communication interface to an external processor.

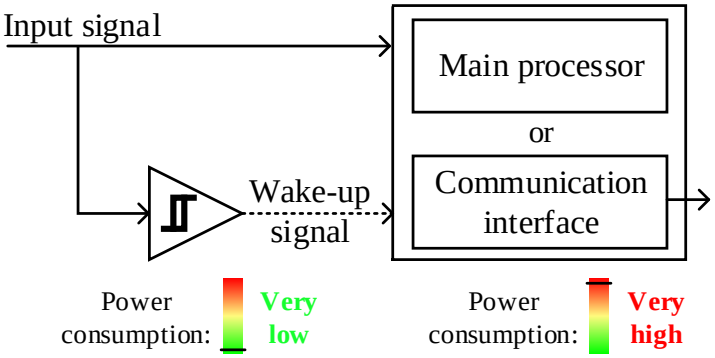


FIGURE 19 TYPICAL ARCHITECTURE OF AUDIO RECOGNITION SYSTEMS

Although this approach is very robust in the sense that it provides an almost perfect detection rate, it is not power efficient for two reasons: The first one is that all raw data is processed the same way. This is especially deleterious for applications that

require power-hungry functions like speech recognition, because the processing units are activated by the presence of noise. The second one is that, in the case of high-complexity recognition tasks, the processing units are situated externally, and thus it is necessary to send all the raw data for processing. However, the transmission of data is typically the largest portion of the energy budget of a sensor node [32].

Thus, two paradigms have been introduced: Near sensor computing relies on processing data locally to reduce the amount of information transmitted and thus the transmission power. The second paradigm is power-proportional sensing introduced by [11]. This paradigm can also be called power-scalable sensing because its premise is that the power consumption of a sensing system must scale proportionally with the complexity of the sensing task. In the next section the details of power-scalable sensing will be discussed.

3.3 Power-scalable sensing

Instead of implementing a single processing unit that performs a complex task all the time, [11] proposes a hierarchical operation of tasks according to their complexity, where each processing stage performs a more complex task than the previous stage. Figure 20 illustrates this principle. The task of speaker identification is hierarchized into three tasks, VAD, gender recognition and speaker identification. Energy is saved because the only processing units that are always-on are those required for VAD and the high complexity algorithms are activated only when strictly required.

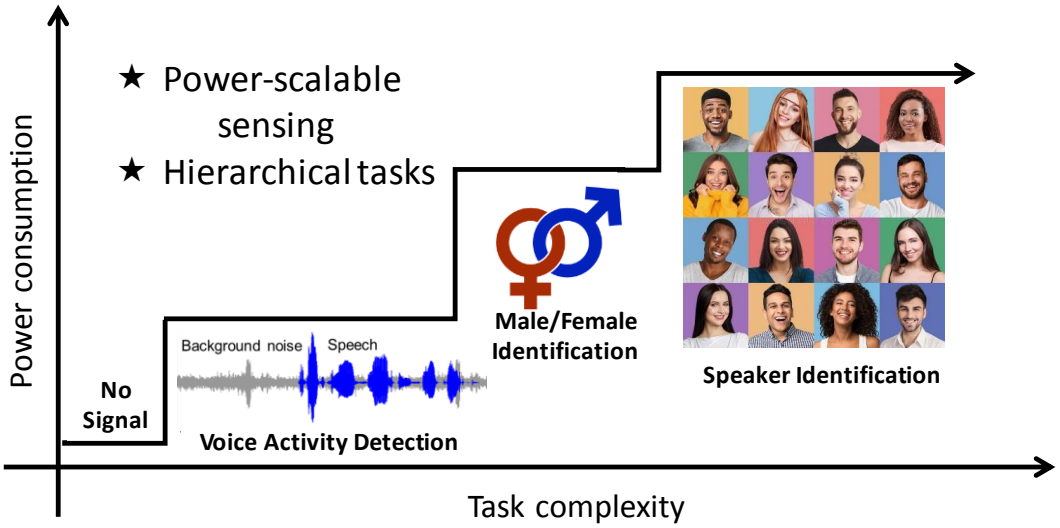


FIGURE 20 POWER SCALABLE SENSING INSPIRED BY [11]

The complexity of a task also varies as a function of the background and context, for instance VAD is easier to perform in a quiet room than in a noisy environment like a restaurant or a concert room [11]. Depending on the context some features are more discriminative than others [33]. Therefore, context awareness helps to reduce power consumption, by activating or deactivating the particular hardware of the most discriminative features in each context.

Hierarchical operation of complex tasks and context awareness open a path for the implementation of near-sensor computing. Figure 21 shows the architecture of a near-sensor computing system. The threshold-based wake-up signal is used to turn a local pre-processing unit on, which performs simple recognition tasks and/or pre-process the input signal, thus simplifying subsequent recognition tasks. Whenever relevant information is detected, the pre-processing unit generates another wake-up signal, to turn on either a sub-sequent processor or a communication interface.

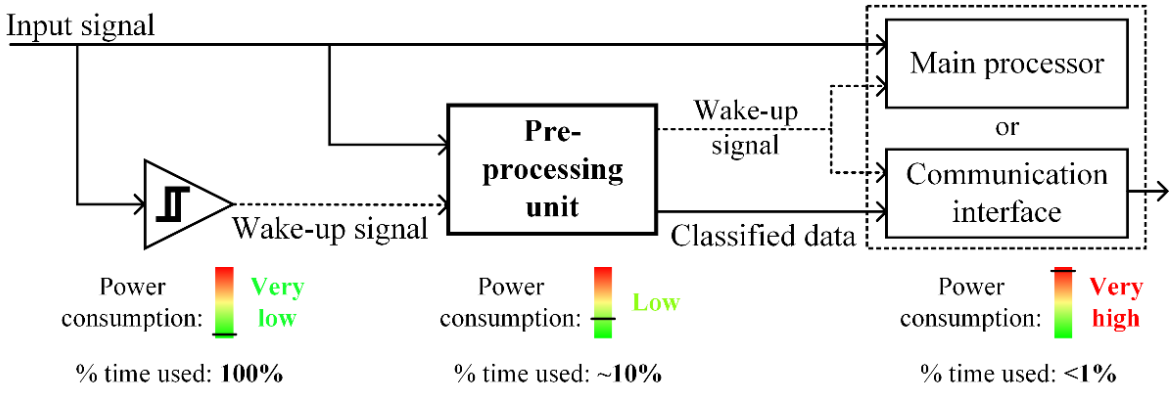


FIGURE 21 ARCHITECTURE OF A HIERARCHICAL AND/OR NEAR SENSOR COMPUTING RECOGNITION SYSTEM

3.4 Characteristics of audio signals and human hearing

The human hearing spectrum range extends from 20 Hz up to 20 kHz [34]. In this section we recall the characteristics of audio signals and the human hearing structure which inspired most audio recognition applications.

3.4.1 Characteristics of audio and voice signals

Voice can be represented with a spectrogram, which is a graph of energy per frequency band versus time. Figure 22 shows a spectrogram of spoken vowels. The number of features in an audio recognition system corresponds to the number of frequency bands of the spectrogram. Sound recognition is done by identifying the frequency bands with the highest energy value.

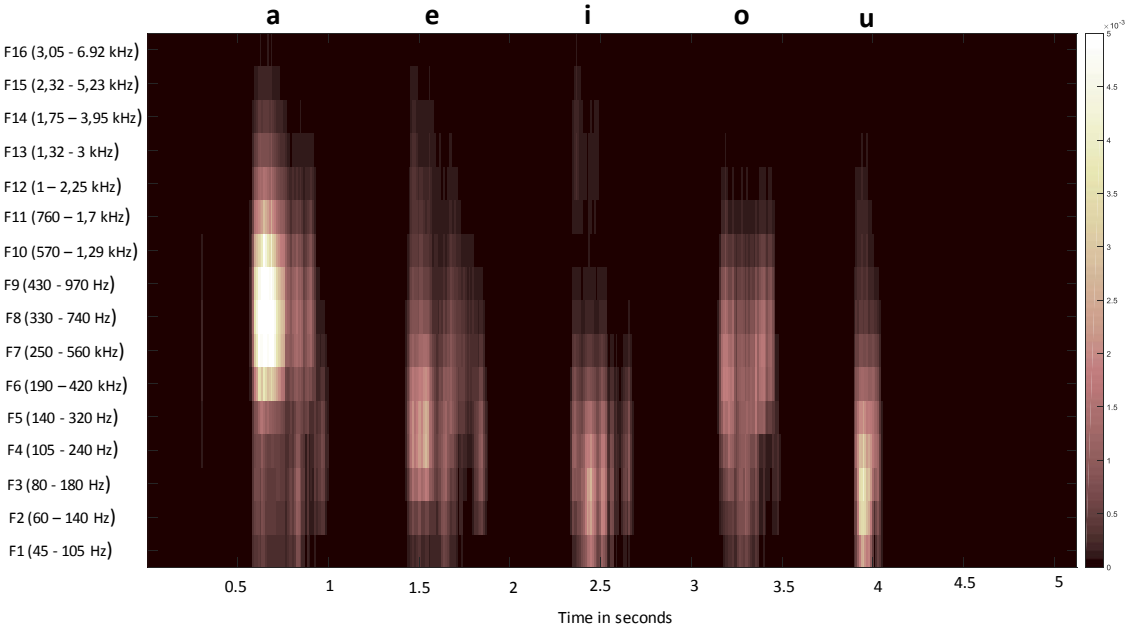


FIGURE 22 SPECTROGRAM OF SPOKEN VOWELS

In the spectrogram of a speech signal the high energy value bands serve to identify the formants which are the broad spectral peaks in a speech signal. These peaks correspond to the resonances of the vocal tract, and their frequency values depend on the morphology of the speaker [34]. The pronunciation of each letter, in a given language, requires a unique shape of the vocal tract. Therefore, the spectrum of each letter has different formant frequencies. The two first formants are important for vowel recognition and the three first formants for consonant recognition [35], [36]. Figure 23 from [37] shows an audiogram of consonant sounds, indicating the sound intensity at one-meter distance (typical conversation) and the frequencies of different consonant sounds for the English language. We can see for instance that the letter “R” has three formants: the first between 600-800 Hz, the second between 1000-1500 Hz and the third between 1800-2400 Hz.

The audiogram also shows although that “F” and “CH” share a common formant (4-5 kHz), but we can differentiate between this sounds because “CH” has an additional lower frequency formant (1.5 – 2 kHz).

Two important features can be highlighted from the audiogram: the first is that the highest frequency consonant sound “th” has a maximum frequency formant at around 6 kHz. This is the reason for considering in most VAD and ASR systems that audio signals have a bandwidth limited to 8 kHz although the human hearing range extends up to 20 kHz, The second characteristic is that the majority of sound formants are concentrated at frequencies between 250 Hz and 2.5 kHz, while the number of formants above 3 kHz is reduced.

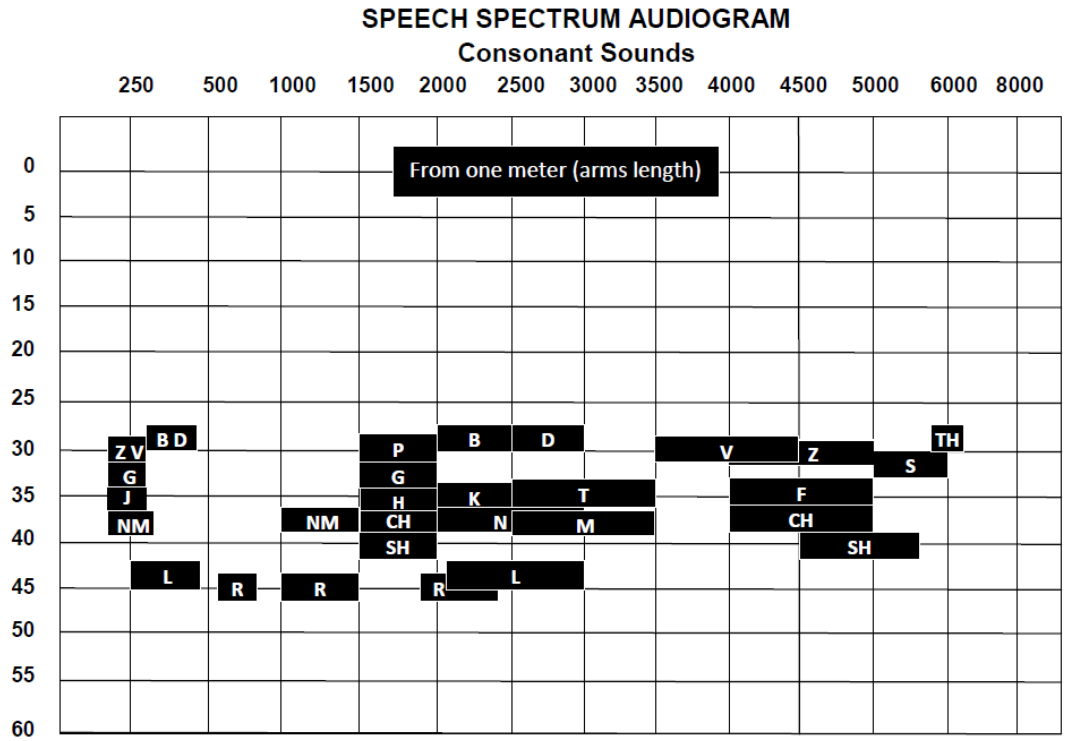


FIGURE 23 SPEECH SPECTRUM AUDIOGRAM OF CONSONANT SOUNDS [37]

Figure 24 shows the formant frequencies of vowels in the English language, the detection of vowel formants is a common method employed in VAD applications [11].

The main characteristic is that each vowel has 2 formants one at lower frequencies (300 – 850 Hz) and another at higher frequencies (1.5 – 3 kHz).

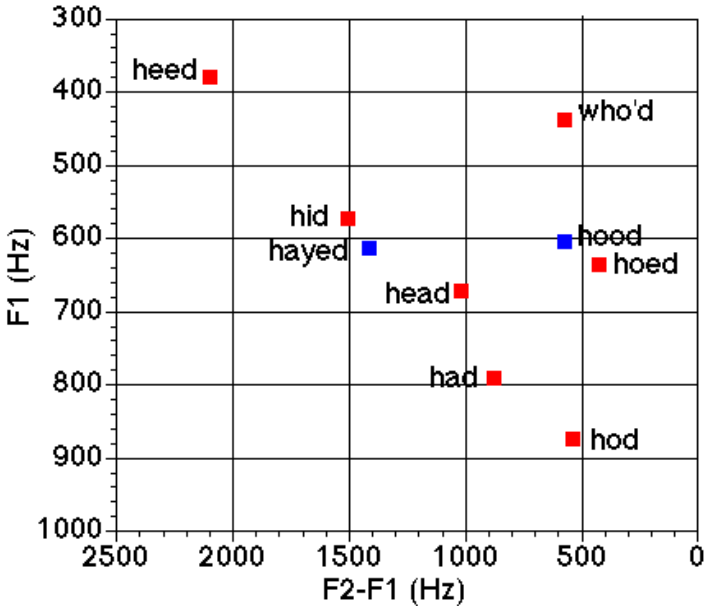


FIGURE 24 FORMANT FREQUENCIES OF VOWELS [38]

In addition to the corresponding formants of each consonant and vowel, there is another low-frequency formant that is present in speech signals which is shown in Figure 25 [39]. This formant is unique to each person, it allows to differentiate between female and male speech, and thus it is useful for speaker recognition tasks. It is in a range between 85-180 Hz for most men and between 165-255 Hz for most women. The frequency spectrum of human speech varies slightly according to the spoken language and the accent associated with each region [40].

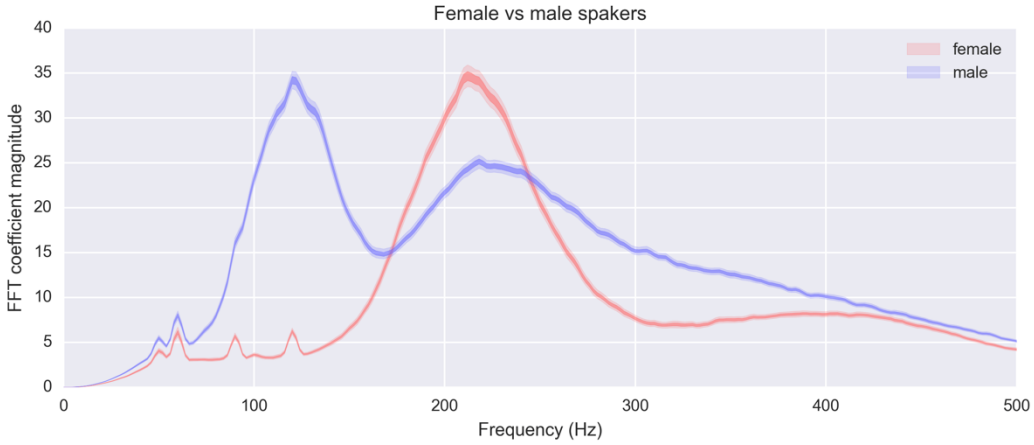


FIGURE 25 FREQUENCY SPECTRUM OF MALE AND FEMALE SPEECH [39]

Another important characteristic which must be considered to generate a meaningful spectrogram of speech sound is the time resolution. A detailed work on the time properties of speech is presented in [6]. Speech can be classified in three categories: normal, fast and slow. For instance, the average duration of a normal syllable is 156.2 ms while the duration of a fast syllable is 93.9 ms.

The average duration of a consonant in normal speech is 50ms and the average duration of a vowel is 99 ms. For fast speech these values are decreased to 33.2ms and 62.7 ms respectively.

The consonants with the lowest average duration are “d”, “b” and “p” with respective duration of 13.9, 14.3 and 17.4 ms. In contrast the lowest average duration vowel is “i” with an average duration of 85.3 ms.

These values are particularly important for ASR systems based on spectrograms, because they will determine the specifications of the spectrogram (number of bands and time resolution) needed to recognize a spoken sequence. If the spectrogram has a coarse step, it cannot accurately represent the consonant sounds.

3.4.2 Human Hearing functioning

A common algorithm to obtain the spectral representation of a signal is the Fast Fourier Transform (FFT), it offers a high spectral resolution with a constant step, however it requires the use of high-energy consuming memories. Thus, many low power integrated applications for audio recognition are inspired by the functioning of the human hearing system. Figure 26 from [34] shows a simplified diagram of the human ear consisting of the outer, middle, and inner ear. Sound undergoes a series of transformations as it travels through the outer ear, middle ear, inner ear, auditory nerve and into the brain. The outer ear catches acoustic pressure waves, which then travel through a narrow passageway called the ear canal and are converted to mechanical vibrations by the eardrum which is a small membrane. The eardrum vibrates a series of small bones in the middle ear. These bones amplify, the sound vibrations and send them to the cochlea. The cochlea which is situated in the inner ear is a snail-shaped cavity filled with fluid that transforms the mechanical vibrations to vibrations within the fluids of the cochlea. These vibrations lead to displacements of a flexible membrane, called the basilar membrane, situated inside the cochlea. These displacements contain information about the different frequencies of the acoustic signal [34],[41]. Finally, the displacements of the basilar membrane, causes the movement of hair cells that are attached to the basilar membrane inside the cochlea. Each time the hair cells move an electrochemical reaction is produced and an electric impulse is sent to the brain [34], [41].

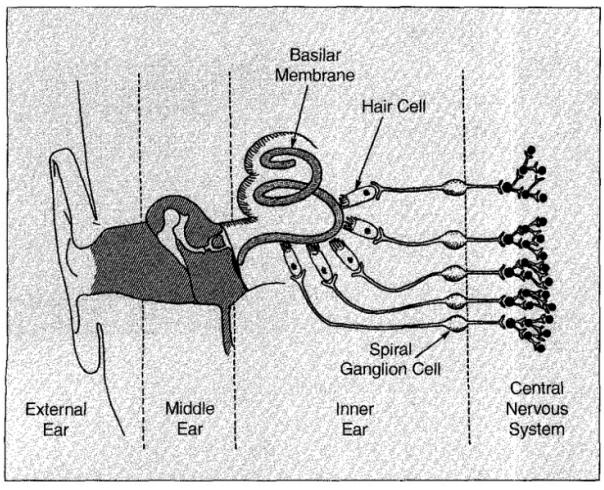


FIGURE 26 DIAGRAM OF THE HUMAN EAR [34]

Georg von Bekesy showed in the 1950s that the basilar membrane in the inner ear acts as a spectrum analyzer, being responsible for decomposing the sound waves into their spectral frequency components. Different frequencies cause maximum vibration amplitude at different points along the basilar membrane. The base of the cochlea responds to higher-pitched sounds. Towards the top of the spiral, the cochlea responds progressively to lower-pitcher sounds. At the very top, the cochlea vibrates in response to the lowest frequency sounds. [34],[41]. If the acoustic signal contains several frequencies, the sound wave will create maximum displacement at different points along the basilar membrane. Figure 27 from [34] shows a diagram of the basilar membrane showing the base and the apex. The position of maximum displacement in response to sinusoids of different frequencies (in Hz) is indicated. The basilar membrane has a logarithmically distributed spectral resolution, with a higher spectral resolution for low-frequency sounds and a lower resolution for high-frequency sounds.

The advantages of creating spectrograms using filters mimicking the human cochlea instead of a standard FFT algorithm are:

- A logarithmically distributed spectral resolution will directly fit the distribution of speech sound formants, simplifying the voice detection.
- A reduced power consumption: As FFT calculates linearly spaced frequency contributions a very small frequency step will be needed to capture low frequency bands. This will result in significant consumption, not only due to computation but also due to memory access. On top of that, further processing will be needed to aggregate FFT results in order to create the spectrogram.
- Recognition algorithms are simplified because they process a reduced number of features.

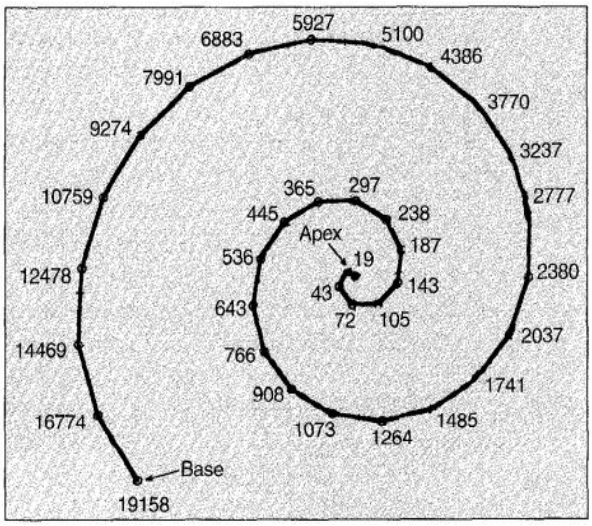


FIGURE 27 DIAGRAM OF THE BASILAR MEMBRANE [34]

3.5 State-of-the-Art of audio feature extraction applications

Figure 28 shows the basic architecture of a speech recognition system. It consists of a feature extraction unit to obtain the spectrogram, followed by a recognition algorithm, which processes the spectral content to extract information. The complexity of both units depends on the amount of information that will be extracted.

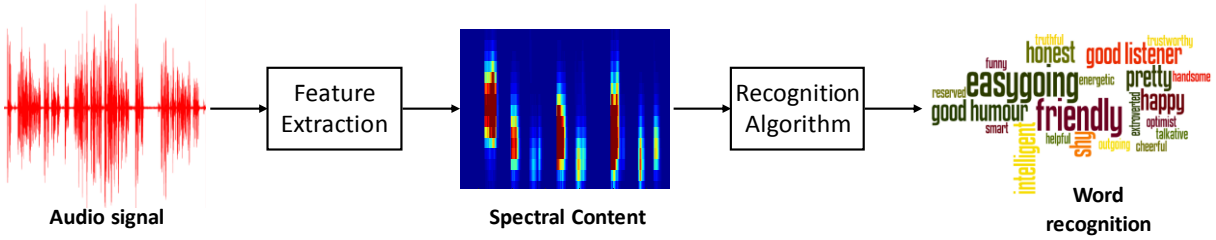


FIGURE 28 BASIC ARCHITECTURE OF AUDIO RECOGNITION SYSTEMS

For simpler tasks like VAD the feature extraction unit consumes a higher amount of resources than the recognition algorithm. For instance, in [42] where a VAD system is implemented, the feature extraction unit has a consumption approximately 15 times higher than the recognition algorithm. However, as the task complexity increases, the recognition algorithm increases its power consumption and becomes the main consumer of power. The work presented by [43], gives a clear illustration of this phenomena. In Table 3 we can see how the power consumption of an automatic speech recognition (ASR) system is increased by several orders of magnitude, as the task complexity increases. For instance, the detection of a small digit's vocabulary of 11 words consumes 172µW, while ASR of a food diary vocabulary of 7000 words consumes 4.67mW.

TABLE 3 POWER CONSUMPTION OF DIFFERENT AUDIO FEATURE EXTRACTION TASKS [43]

Task	Vocabulary	Number of Words	Power Consumption
VAD	---	---	22.3 µW
ASR	Digits	11	172 µW
ASR	Weather	2k	4.7 mW
ASR	Food Diary	7k	4.67 mW
ASR	News 1	145k	7.78 mW

3.5.1 VAD Applications

In VAD applications, the feature extractor dominates the energy consumption and area of the overall system. The two approaches to obtain the spectral content of a signal are shown in Figure 29. The digital approach consists in the use of an ADC to sample the signals and send the information to a complex processing unit. ADC sampling frequency is dictated by the voice signal bandwidth (8 kHz) or the human hearing range bandwidth (20 kHz). This approach uses a Fast Fourier Transform (FFT), to obtain the spectrogram of the input signal [42],[43]. The analog approach consists in the implementation of signal pre-processing in the analog domain, and subsequent sampling using an ADC. ADC data rate is reduced to the information change rate (10 – 100 Hz) [7]. This approach is known as Analog-to-Information and has been implemented by [11] and [12]. The spectrogram is obtained with a logarithmically distributed analog filter bank that emulates the functioning of the human cochlea. The filter bank approach is also used in circuits used as cochlear implants and hearing aids, for this application filters bank have been implemented in both the digital [9] and the analog domains [44].

The recognition algorithms employed differ among the different VAD implementations. In some cases, like [11] a decision tree is used, in this case the energy values at the output at each band are quantified and compare with a certain threshold. Consumption and latency times depend on the number of decision tree

nodes. For instance, in [11] three decision nodes are used, the specific features employed to make the decision vary in function of the context.

Other implementations like [12] and [43] used neural networks as recognition algorithms.

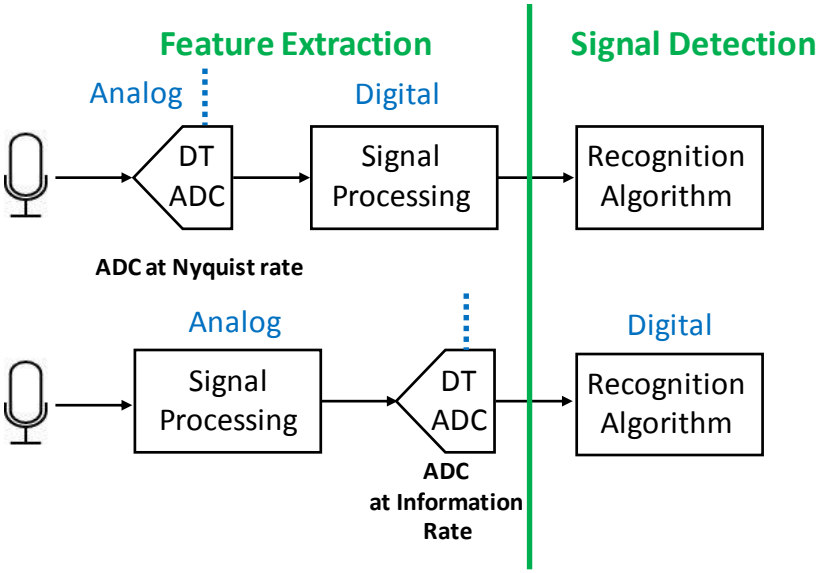


FIGURE 29 APPROACHES TO COMPUTE THE SPECTRAL CONTENT OF A SIGNAL TOP: DIGITAL AT NYQUIST RATE. BOTTOM: ANALOG-TO-INFORMATION. IMAGE INSPIRED FROM [32]

As described by [45] the main tradeoff in a VAD system is between accuracy and power consumption. The tradeoff includes the choice of the number of features, the complexity of the recognition algorithm and in the case of digital implementations the ADC resolution.

The key to power/accuracy optimization lies in finding the proper compromise between overfitting and underfitting. To better grasp these concepts, we can refer to the work of [45], where the accuracy of a recognition system is determined as a function of the ADC resolution and the size of a decision tree algorithm.

Overfitting occurs for low ADC resolutions and a high complexity decision trees, accuracy is abruptly degraded because the information content is poor and the decision tree classifies noise.

The opposite phenomenon called underfitting occurs when using a high-resolution ADC and a simple decision tree. In this case accuracy is degraded because the information is rich in context, but the decision algorithm does not have enough capability to process it.

Thus, a high configurability system is desired because it allows to tune the system parameters in order to always attain the best compromise between accuracy and power consumption.

In [13], a study was conducted to optimize the resolution of a discrete time ADC, used in a simple speech recognition system. An 8-bit resolution can be used without degrading the accuracy of the recognition system.

The requirements of the filter bank were determined in a study conducted by [46]. The study found that the level of accuracy depends on both the spectral resolution

(the number of filters/bands) and the spectral amplitude resolution, which is the number of quantification levels used to quantify the energy at each filter output. Therefore, it can be possible to keep the same level of accuracy by reducing the number of filters and increasing the resolution of energy quantization. In [46], the number of channels needed to understand speech was evaluated. Results are in accordance with the study made in [13]. A system with 8 to 16 filters and an energy quantification resolution of 6 to 8 bits are appropriate to obtain a classification accuracy of approximately 90%.

Table 4 shows a comparison of different circuits that use the filter bank approach to obtain the spectral content of a signal for VAD applications and cochlear implants. Analog implementations have a lower power consumption and occupy a larger area than digital implementations, particularly if digital implementations are not event driven and do not exploit the large sparsity of voice signals. However digital implementations occupy a smaller area and can be integrated in advanced nanometer technologies.

Except for [44] whose frequency range covers the whole human hearing spectrum, most applications are limited to cover the spectrum of voice signals with frequencies between 75Hz and 8kHz.

TABLE 4 STATE OF THE ART OF LOW POWER INTEGRATED FEATURE EXTRACTOR FILTER BANKS

Feature Extractor	Yang, 2018 [12]	Yang, 2016 [44]	Badami, 2016 [11]	Yip, 2015 [10]	Wu, 2017 [9]
Technology	180nm	180nm	90nm	180nm	65nm
Feature Type	Analog filters	Analog filters	Analog Filters	Digital Filters	Digital Filters
Channel Number	16	64x2	16	8	18x4
Frequency Range (Hz)	100 – 5k	8 – 20k	75 – 5k	300 – 5.5k	160 – 8k
Power (µW)	0.38	55	6	1.59[a]	13.8
Power/Channel (nW)	24	430	380	199	190
Area/Channel (mm^2)	0.1	0.26	0.13	0.19	0.00934
Time spectral resolution (ms)	10	NA	<100	NA	< 0.05

a. Does not take into account the clock and the ADC power consumption nor the output data redundancy

3.6 Conclusion

In this chapter, audio feature extraction was studied. There is a diverse range of audio feature extraction applications that demand high configurability to operate in very different backgrounds and low power consumption.

To meet this this challenge, the concept of power scalable sensing has emerged, whose premise is that the power consumption of a sensor should scale in accordance to the complexity of the task it performs, in terms of both task utility and task complexity. To achieve this premise a hierarchical implementation of tasks and a flexible, adaptable systems are required.

The content information of audio signals is found in its frequency spectrum, with each sound having a corresponding set of frequencies and intensities, like consonant and vowels sounds that have two or three distinct frequency components. Thus, the task of audio signals recognition is divided in two parts: The first one is the signal processing required to obtain the spectral representation of the audio signal, called spectrogram. The second one is the processing used to extract and classify the information present in the spectral representation.

In the case of low complexity recognition tasks like VAD, the signal processing to obtain the signal spectrum consumes most of the system resources. The most efficient approach is the use of a filter bank to mimic the human cochlea, which can be implemented in both the analog and digital domains. The advantages and drawbacks of analog and digital implementations are complementary. Therefore, it would be interesting to find an approach that can combine the best of both domains.

To further improve the system performance and expand its capabilities beyond VAD, in order to perform other tasks like specific word recognition or male/female differentiation, some slight modifications can be made to the system architecture. For specific word recognition a spectrogram with a maximum time resolution of 20ms is needed to avoid missing short duration consonants, while for male/female differentiation it is required to cover the 85-180 Hz and 165-255 Hz frequency bands where the male/female voice formant are present.

Chapter 4
Proposed System: Continuous Time filter bank suitable for audio feature extraction

4.1 Introduction

We concluded the previous chapter by showing the state of the art of feature extractor circuits for audio applications. In this chapter, we propose a Continuous Time (CT) digital signal processing, filter bank to obtain the spectrogram of an audio signal. We will describe the system and present the architectural choices that were made for each block. Next the implementation challenges will be discussed. Next, we will study into detail different filter structures and a novel circuit allowing to cascade FIR filters in CT. A novel function to calculate instantaneous energy in CT will also be presented. Finally, we will speak about the delay elements needed to implement the filter functions. We will conclude this chapter by an estimate of the complete system performance in terms of silicon area and power consumption.

4.2 Proposed CT DSP system for feature extraction

In chapter 3, we concluded that the analog and digital filter banks used to obtain the spectrogram of an audio signal, have advantages and drawbacks that are complementary. Thus, we propose a novel approach, shown in Figure 30, to obtain the spectrogram of a signal. The analog signal from the audio source is converted to the continuous time digital domain. Unlike a DT ADC, the CT ADC is event driven and does not requires a clock, therefore the power consumption scales with the input activity. Since signals are quantified in amplitude, the filters and the digital processing functions are implemented in the digital domain. Making the system configurable and scalable with advanced technology nodes. The filter bank and instantaneous power functions are implemented in CT DSP. Since the energy computation block requires a uniform time reference over which the instantaneous power will be accumulated, a very low-frequency clock is used for this purpose. The use of this clock does not degrade the advantages of CT processing regarding aliasing and an event-driven power consumption and allows interfacing of the system with a discrete-time recognition algorithm. The proposed approach can be considered an “analog-to-information” interface, because the data rate at which digital information is generated, is dictated by the information rate and not by the Nyquist theorem.

The proposed feature extraction unit is shown in Figure 31. The system is composed of a CT ADC, followed by a bank of FIR filters in CT, whose frequency response is inspired by the human cochlea. FIR filters are chosen over IIR filters because their frequency response is easily configurable and always stable. To generate the spectrogram of the audio signal, the instantaneous signal power is calculated in CT at the output of each filter. Finally, to calculate the energy of the signal a very low-frequency clock (typically 10-100 Hz) gives the period over which the power is accumulated. At the end of each clock cycle, the final calculated energy value is latched and transferred to the recognition system and the accumulator is reset. This value is chosen to avoid information loss in accordance with the duration of spoken letters.

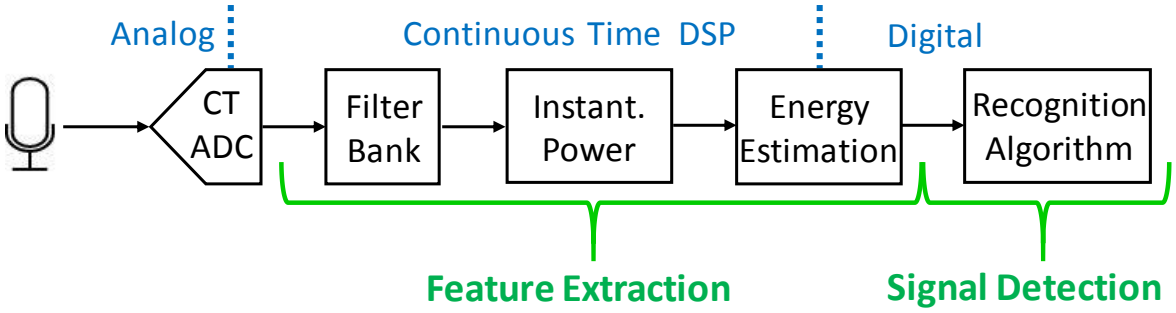


FIGURE 30 PROPOSED APPROACH TO COMPUTE THE SPECTROGRAM OF A SIGNAL

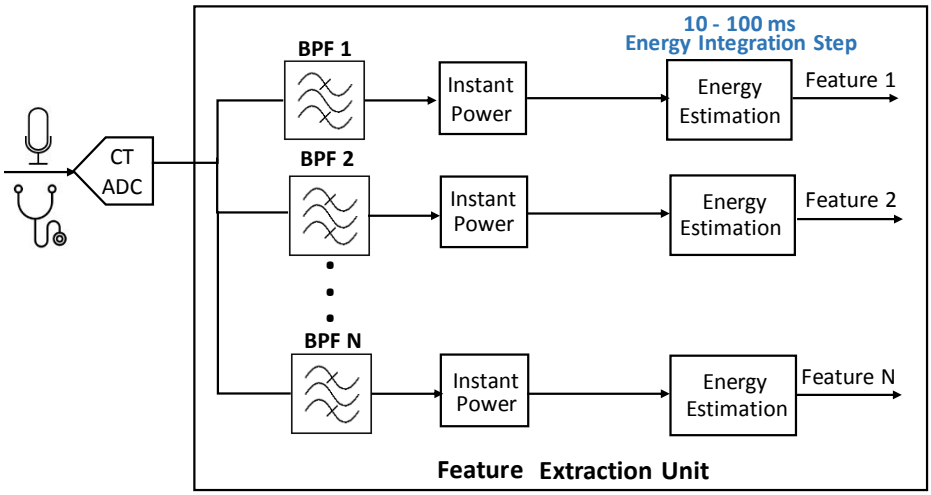


FIGURE 31 FEATURE EXTRACTION SYSTEM IN CT DSP

4.2.1 Targeted performance

The targeted performance of the proposed system with regards to the existing state-of-the-art implementations are the following:

1. A low power consumption, between 1 – 5 µW, in the range of existing analog filter bank implementations.
2. A total silicon area inferior to 1.5 mm², which represents a significant reduction with regards to existing analog filter banks.
3. High configurability and adaptability.

4.3 Architectural choices

4.3.1 ADC Type

Given the event-driven nature of a CT system, the ADC dictates the number of samples that will be generated and thus is fundamental to dimensioning the subsequent signal processing unit. Thus, the ADC parameters are required to estimate the overall surface and power consumption of the system.

We analyze the following architectures introduced in Chapter 2 to determine which one is best suited for the targeted application, feature extraction of audio signals:

- Level crossing sampling
- Level crossing sampling with a variable resolution
- Level crossing sampling with an embedded energy integrator

Table 5 shows a comparison of the different CT ADC architectures. For the processing of audio signals which are low frequency signals with drastically varying spectral content, it is important that the ADC adapts to the spectral content variations. Despite its poor performance with regard to power consumption, the level crossing sampling ADC has been chosen for the following reasons: it has a simple structure, produces one-bit samples that ease the implementation of the arithmetic functions for signal processing and has a large SNR. Although the ADC with embedded energy integrator has the highest energy efficiency, its SNR is moderate, and it does not adapt to the spectral content. The LCS with variable resolution produces multi-bit output samples and is therefore not appropriate for the envisaged architecture.

TABLE 5 COMPARISON OF DIFFERENT CT ADC ARCHITECTURES

	Level Crossing Sampling (LCS)	**LCS with Variable resolution**	**LCS with embedded Energy integrator**
Complexity	Low	High	Moderate
Adapts to Spectral content	Yes	Yes	No
1 Bit Processing	Yes	No	Yes
Number of Samples Generated	High	Moderate	Moderate
SNR	High	High	Moderate
Energy efficiency	Poor	Moderate	High

4.3.2 ADC Dimensioning

In CT, the resolution of the ADC is inversely proportional to the granular timeT_{gran}, which is the minimum time between two samples. For each extra bit of resolution, the number of samples generated increases by a factor 2. Table 6 shows the dynamic power consumption of the delay elements needed to implement a 15TH order FIR filter as a function of the ADC resolution. The considered ADC has a maximum input frequency of 8 kHz, this bandwidth is used by most filter banks used for audio feature extraction and is enough to contain the information related to speech and human voice. The input signal is a full scale 2 kHz sinusoidal wave and granular delays are considered to have a 20fJ/event power consumption.

TABLE 6 DYNAMIC POWER CONSUMPTION OF THE DELAY ELEMENTS IN A CT SYSTEM IN FUNCTION OF THE ADC RESOLUTION

System bandwidth	**Total Delay (µs)**	**Delay Taps**	**ADC Resolution**	**Granular Time**	**Granular Delays**	**Dynamic Power Consumption µW**
8 kHz	62.5	15	5	1.24 µs	50	3
			6	621 ns	100	12
			7	310 ns	200	48
			8	155 ns	400	194
			9	77 ns	800	776

**** Energy consumption of a full scale 2 kHz sinusoidal input. Granular delay elements are considered to have a dynamic power consumption of 20fJ/event.**

For each extra bit of resolution, the dynamic power consumption increases by a factor 4 for a full scale sinusoidal input, since both the number of events and the number of delay cells required to process the signals are multiplied by a factor 2. Therefore, the ADC resolution is a critical choice for power consumption and silicon area.

In Figure 32, the number of samples generated by a CT ADC is shown for different input signals. For human speech the number of samples generated is low, while full scale babble noise generates a significantly higher number of samples. The explanation can be seen in Figure 33 and Figure 34. Male speech is a low frequency signal that contains many pauses, during which the CT ADC system does not generate any new sample. Babble noise on the other hand is a continuous signal that has no pause, and thus generates a larger number of samples. For resolutions higher than 8 bits we don't obtain any significant data reduction compared with an implementation that uses a discrete time ADC at a 16 kHz sampling rate.

An important aspect to highlight is that for speech signals quantified with resolutions lower than 6 bits, the number of samples generated does not follows a linear progression. The reason for this phenomena is that many of the speech signal variations are smaller than the LSB amplitude at 4 or 5 bits resolution, therefore these variations are discarded and do not generate any event.

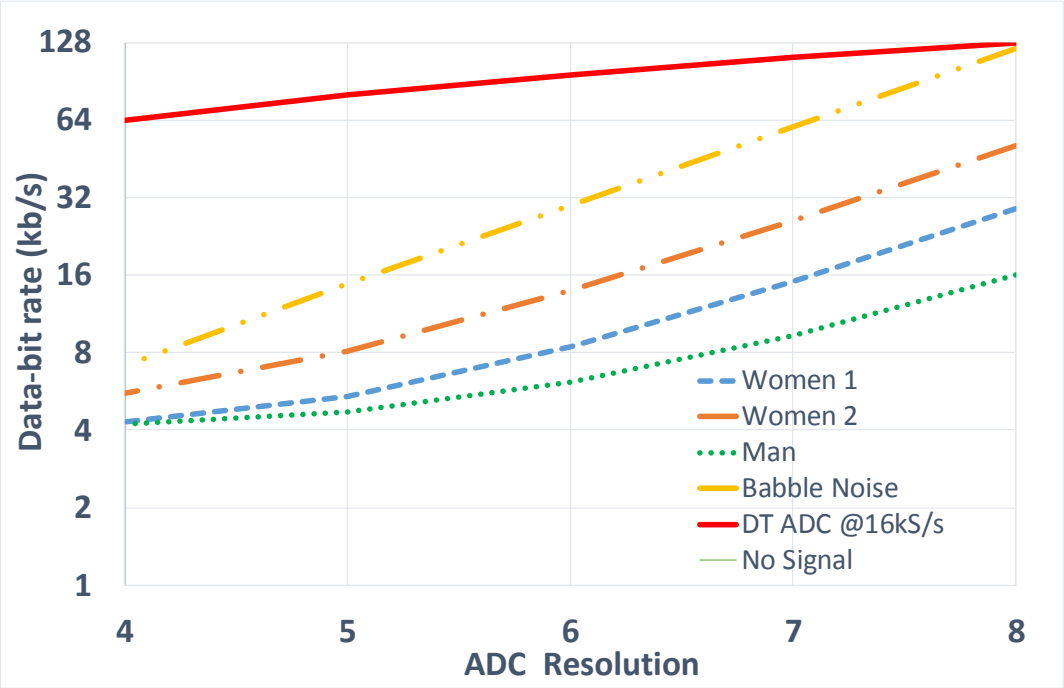


FIGURE 32 NUMBER OF SAMPLES GENERATED BY A CT ADC FOR DIFFERENT INPUT SIGNALS

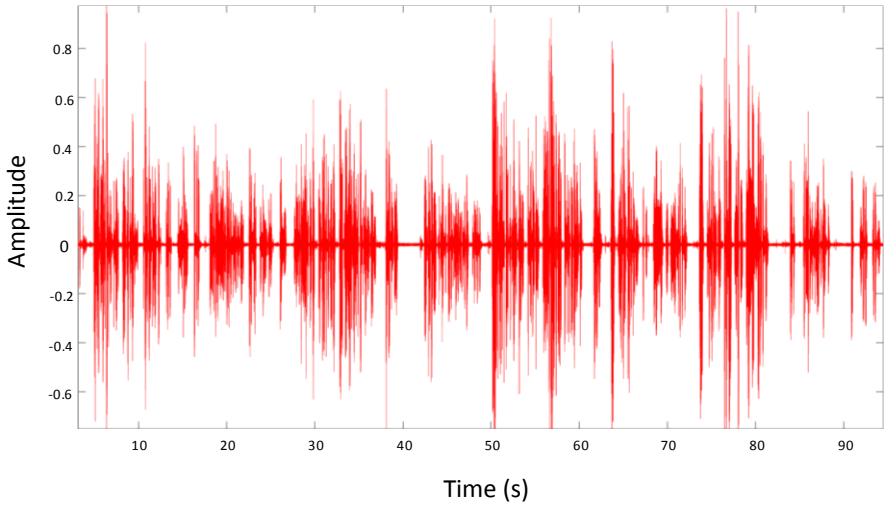


FIGURE 33 MALE SPEECH PLOT USED TO GENERATE FIGURE 32

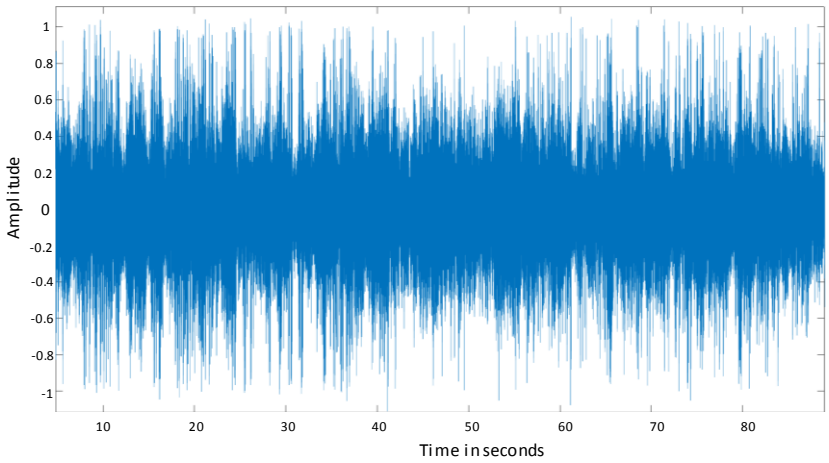


FIGURE 34 BABBLE NOISE PLOT USED TO GENERATE FIGURE 32

4.3.3 Filter bank dimensioning

The filter specifications are obtained from [46]. All filters must have a stop band rejection of at least 20dB, without overlap between their pass-bands at their 3dB cut-off frequencies. We propose the use of 12 band-pass filters, whose frequencies are given in Table 7. For a 12 filter implementation [46] proposes a frequency range of 192 Hz – 5.5kHz. We have increased this frequency range to 60Hz - 6.4 kHz using 12 filters, to best cover the frequency spectrum of the human voice and human speech. The center frequencies of the filters and their distribution mimic the human cochlea. The central frequencies of filters below 1 kHz are spaced by a factor varying from 2.1 to 1.5. For filters above 1 kHz, the central frequencies are spaced by a factor ranging from 1.3 to 1.25.

To attain a better selectivity between features, the pass-bands and the stop-bands of each filter were chosen, to guarantee that each filter has a rejection of at least 20dB at the central frequency of the filters adjacent to their closest neighbors as shown in Figure 35. For instance, filter 7, centered at 1.95 kHz, has a stop-band rejection of at least 20 dB at 1.1 kHz and 3.2 kHz, which are the central frequencies of filters 5 and 9 respectively.

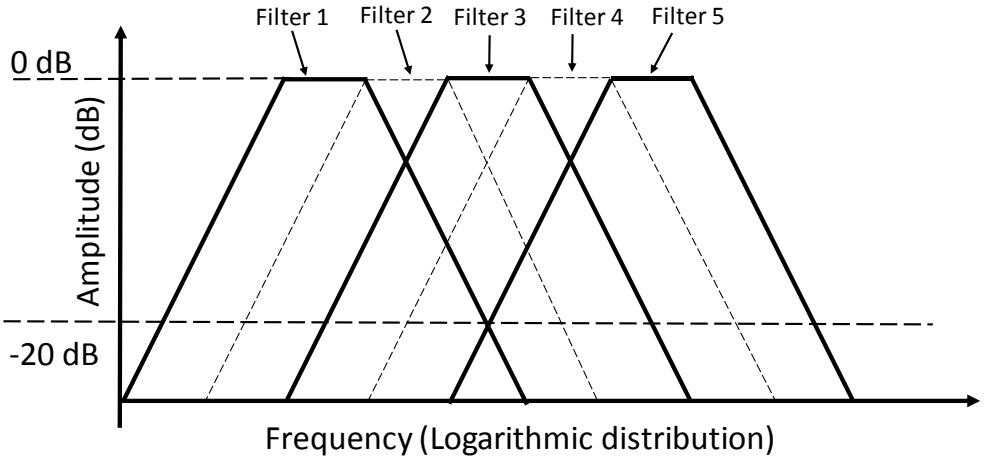


FIGURE 35 FILTER BANK REPRESENTATION, ILLUSTRATING THE 20 DB STOP BAND REJECTION AT THE CENTRAL FREQUENCIES OF THEIR SECOND CLOSEST NEIGHBORS

The transistion band, also called the skirt, is a range of frequencies that allows a transition between a passband and a stop band of a filter [47]. The transition bandwith ratio is the width of the transition band divided by the total bandwith of the input signal To meet these requirements high order FIR filter functions are required. For instance, a 778th order FIR filter is required to implement filter 1, which is centered at 105 Hz, and has a transition bandwidth of 0.012.

TABLE 7 CHARACTERISTICS OF THE PROPOSED FILTER BANK

Filter number	Center Frequency	Bandwidth
1	105 Hz	90 Hz
2	225 Hz	150 Hz
3	450 Hz	300 Hz
4	750 Hz	300 Hz
5	1.1 kHz	400 Hz
6	1.5 kHz	400 Hz
7	1.95 kHz	500 Hz
8	2.55 kHz	700 Hz
9	3.2 kHz	600 Hz
10	4 kHz	1 kHz
11	5 kHz	1 kHz
12	5.9 kHz	800 Hz

4.4 Filter bank implementation

4.4.1 Global architecture

The filter bank architecture is shown in Figure 36. All filters implemented are 16th order FIR filters to maintain an efficient implementation in terms of energy and surface. At the output of each filter, energy is computed to obtain the spectral representation of the input signal. Each filter has a unique set of coefficients that were obtained using Matlab® with an equiripple design method, additionally coefficients were calculated to maintain the passband as flat as possible. The frequency response of the filter bank is shown in Figure 37.

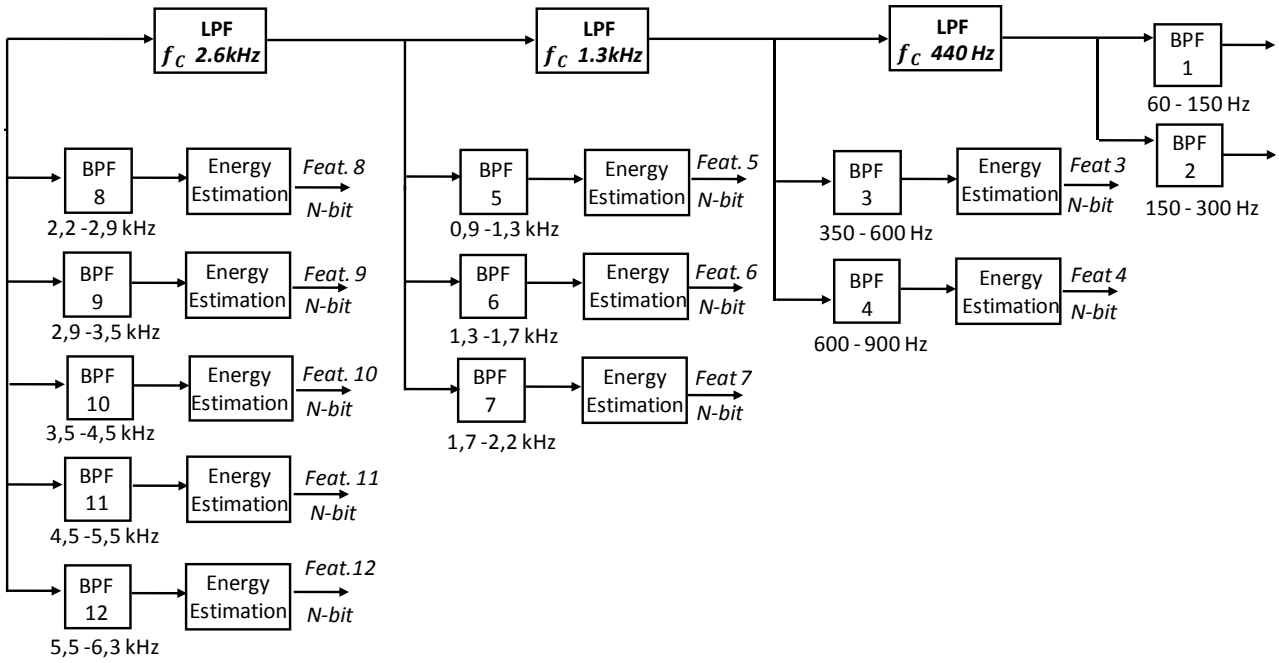


FIGURE 36 COMPLETE SYSTEM IMPLEMENTATION

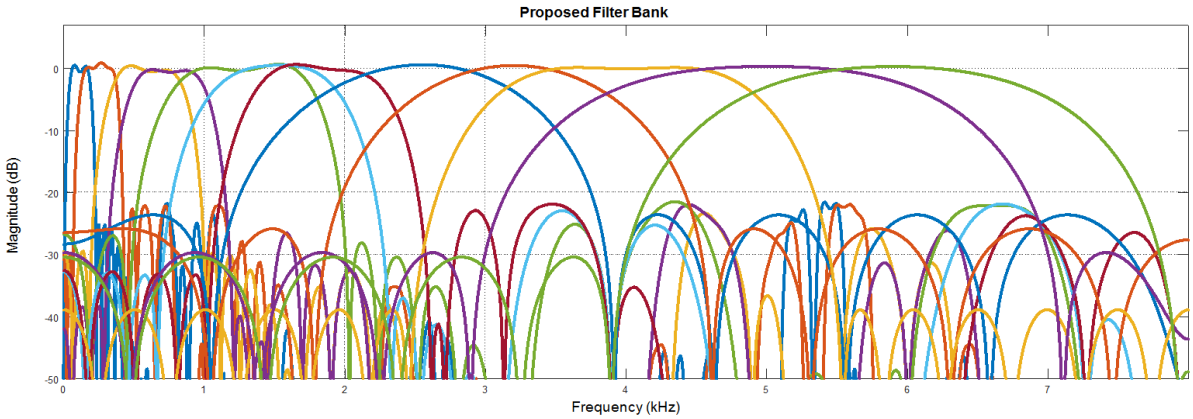


FIGURE 37 FREQUENCY RESPONSE OF THE FILTER BANK

To efficiently implement high order FIR filter functions, the use of cascaded filters is proposed. For this purpose, the proposed system contains three low pass filters that serve as decimators and reduce the bandwidth. Decimation also allows energy savings, because information is processed with a lower number of events.

The implementation challenges of the proposed CT filter bank suitable for audio feature extraction applications are the following:

1. The implementation of cascaded FIR filter functions;
2. Energy efficiency;
3. Silicon area reduction;
4. Instantaneous power and energy estimation functions in CT DSP.

The architecture shown in Figure 36 proposes to cascade CT DSP filters. At each stage, a low pass filter reduces the signal bandwidth, which is the first step required to reduce the event-rate and thus increase the granular time T_{gran} of subsequent filters. The first and the third decimator reduce the bandwidth by a factor

3, while the second decimator reduces the bandwidth by a factor 2. At each stage, we have thus $^1/_3$, $^1/_6$ and $^1/_{18}$ of the original bandwidth. The bandwidth is decimated by a factor 2 or 3, because these are the larger integer values for which the band can be decimated using a 16th order FIR filter and still attain a -20dB stop-band rejection.

The low pass filters in Figure 36 have a special constraint: their output must be in the form of a pair of Change and UD signals to be able to feed it to the input of other filters. However, the standard low-pass filter output is an N-bit word. Given that the input signal to the LP filter only varies by one LSB at a time and assuming unity gain in the pass-band, the output of the filter should also not vary by more than one LSB at each event. It is thus possible to convert the N-bit filter output into a pair of Change and UD signals that will be fed to the next filter. The principle is shown in Figure 38 where a state machine called delta encoder examines the changes of the two least significant bits of the filter output and generates at each transition a pair of Change and UD signals that will be fed to the following filter.

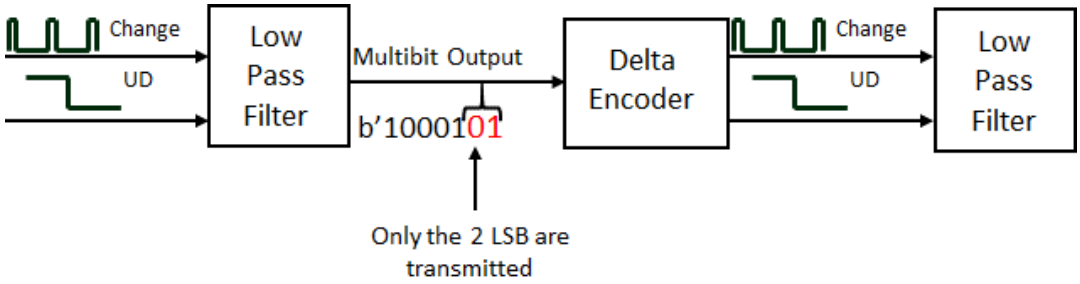


FIGURE 38 PROPOSED LOW PASS FILTER CASCADE OPERATION

The delta encoder block is shown in Figure 39. The delta encoder allows to maintain the advantages of the single bit processing: The simplified carrying of information across the delay taps of the subsequent filter, and the implementation of the filters coefficients only with adders and accumulators.

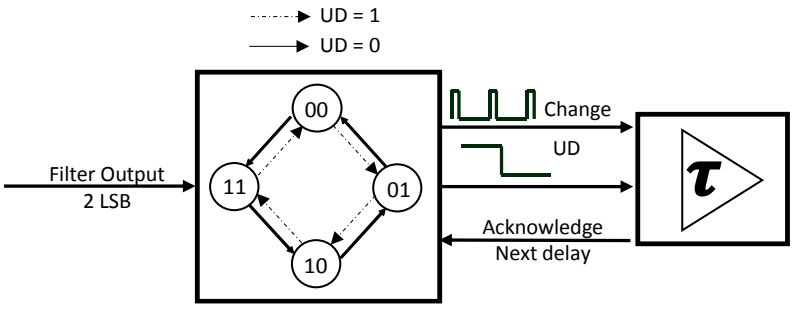


FIGURE 39 DELTA ENCODER CIRCUIT

4.4.2 Design of 16th order filters

A CT DSP FIR filter is a linear combination of the delayed input. Figure 40 shows an implementation of one of the four CT DSP FIR filter section that compose our system. Three components are needed to implement FIR filters: delay elements, multiplier-accumulators, and an output adder. The filter order is equal to the number of delay elements used. The filter coefficients are M-bit numbers. Thus, the output of the adder is coded on N+M bits (where N is the ADC resolution). Given that the input signal only varies by one LSB at a time, the multiplier circuit is simplified to an accumulator and a multiplexer [15]. The output of the filter is obtained by truncating the output of

the adder to the N most significant bits. Due to the absence of a clock, the samples generated by the CT ADC are delayed by CT delay taps of delay timeT_D. To accommodate all the samples and to preserve information, the delay taps are constituted by sequences of granular delays, with value T_{gran} as shown in Figure 40.

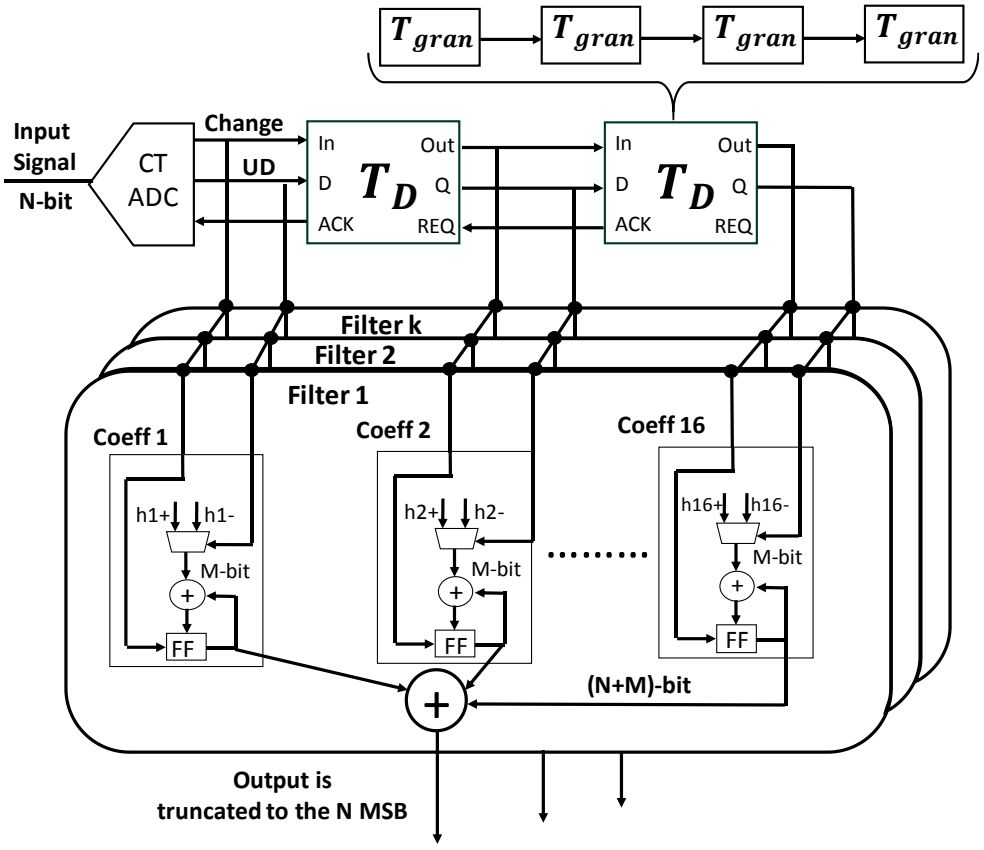


FIGURE 40 16TH ORDER CTDSP FIR FILTER SECTION

Since our proposed system, has three decimators, we need four different delay taps, ranging from T_{gran} to $18 \times T_{gran}$ each constituted by 16 delay taps. To ease the implementation, the use of a single digital delay cell with a large tuning range that complies with all delay taps is preferable. The characteristics of the granular delay elements that composed the delay taps depend on the ADC resolution. Table 8 compares the delay taps characteristics for a 6-bit and an 8-bit ADC resolution.

TABLE 8 CHARACTERISTICS OF THE DELAY TAPS FOR DIFFERENT ADC RESOLUTIONS

Branch	Decimation Factor	ADC Resolution Bits	Delay Tap	Number of granular elements	Granular delay time
1	1	8	62.5 µs	400	156 ns
		6		100	468 ns
2	1/3	8	187 µs	400	937 ns
		6		100	2.8 µs
3	1/6	8	375 µs	400	624 ns
		6		100	1.87 µs
4	1/18	8	1.12 ms	400	3.75 µs
		6		100	11.25 µs

Apart from tuning range, other specifications that must be considered when designing the digital delay elements are power consumption, mismatch and area. Among these specifications, mismatch is the only one that can jeopardize the correct functioning of the system and is measured as σ/μ (standard deviation σ over average μ), of the delay time. Figure 41 shows the effects of delay tap mismatch on the filter's frequency response. For this simulation, the delay taps had a σ/μ value of 10%. Two effects are present: The first is the shift on the filter cut-off frequency, whose value correspond to the standard deviation of the delay taps; the second one is the degradation of the stop-band rejection.

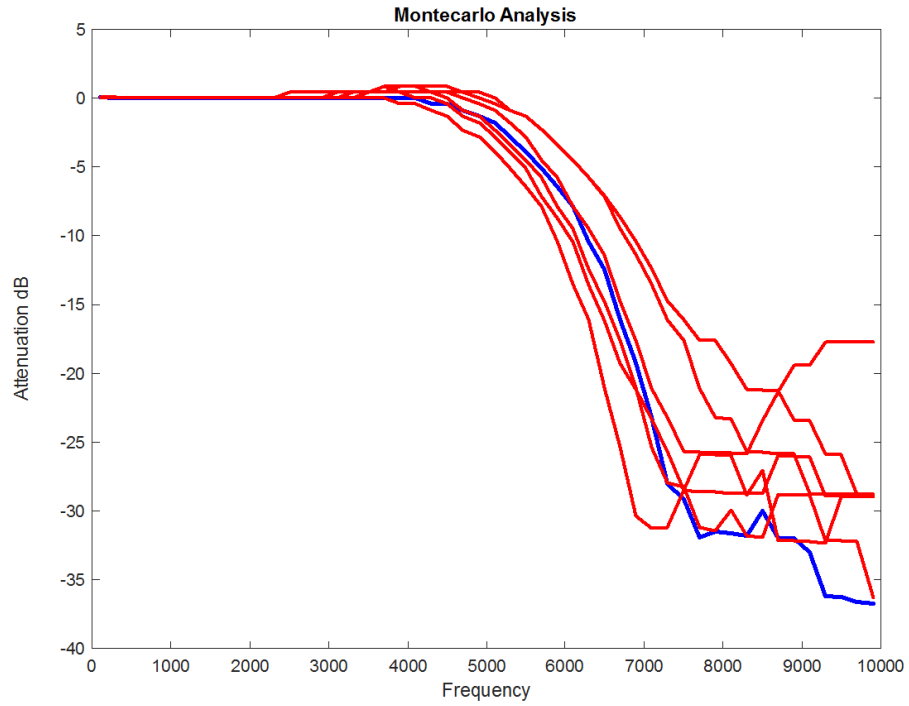


FIGURE 41 EFFECTS OF MISMATCH IN THE FREQUENCY RESPONSE FOR A DELAY TAP MISMATCH VALUE (σ/μ) OF 10%.

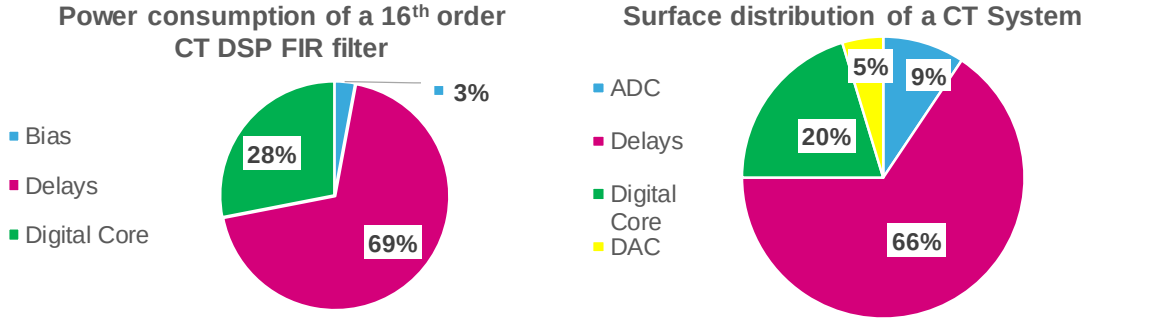
Given that the total delay tap is the sum of smaller granular delays of time T_{gran}, the global delay value follows the probability density described in equations (6-8). Assuming independent random distribution of delay variations, the variance ${\sigma_{sum}}^2$ of the sum of n equal elements, is equal to the sum of the variances of each element ${\sigma_n}^2$. And thus the σ/μ of the sum of elements corresponds to the σ/μ of the single element divided by the square root of the number of elements. For example, if n = 100 and the n element has a σ/μ = 0,12 the sum will have a σ/μ = 0,012. Therefore, given that the system can tolerate a maximum delay tap mismatch value of 10%, the specification for the granular delay mismatch is set a 30%, for a delay tap composed by 100 granular delays. This specification allows to have an important margin between performance and the limit operating point.

$$\sigma_{Tap}^2 = n * \sigma_{Delay\ gran}^2 \quad (6)$$

$$\sigma_{Tap} = \sqrt{n} * \sigma_{Delay\ gran} \quad (7)$$

$$\sigma_{Tap}/\mu_{Tap} = \sqrt{n} * \sigma_{gran}/n * \mu_{gran} = \frac{1}{\sqrt{n}} * \frac{\sigma_{gran}}{\mu_{gran}} \quad (8)$$

To estimate the silicon area and surface occupied by the delay lines we use [24] as a reference, where a complete CT system composed by a 15th order FIR filter, a CT ADC and a CT DAC was implemented in a 90nm CMOS technology. The contribution of each system block to the area and power consumption is shown in Figure 42. The delay lines contribute 70% of the overall system power consumption and occupy 66% of the silicon area. These values will be used for power consumption and silicon area estimations of the proposed system. The delay elements are the key elements that needs to be optimized, in order to attain energy efficiency and reasonable silicon area.



Bias	Delays	Digital Core	Total
0,05 mW	1,18 mW	0,48 mW	1,71 mW

ADC	Delays	Digital Core	DAC	Total
0,06 mm²	0,42 mm²	0,13 mm²	0,03 mm²	0,64 mm²

Figure 42 Power consumption and surface of a CT DSP System

Considering the targeted performance for the complete system, we estimate the required granular delay specifications for different ADC resolutions which are shown in Table 9.

TABLE 9 DELAY ELEMENT SPECIFICATIONS FOR DIFFERENT ADC RESOLUTIONS

Criteria	Total system performance	Granular delay @ 8 bits ADC	Granular delay @ 6 bits ADC
Tuning range	62.5 µs – 1.12ms	156 ns – 3µs	625 ns – 11.24µs
Mismatch σ/µ	<10%	<60%	<30%
Surface	≤ 0.1 mm² per channel	≤ 25 µm²	≤100 µm²
**Dynamic Energy Cons.	150 -175 nW per channel	≤ 1.5 fJ/event	≤ 25 fJ/event

**Targeted energy consumption for a typical speech signal

The delay element specifications are very stringent for an 8-bit ADC resolution in terms of surface an energy consumption. If we reference to Figure 32 we observe that a resolution of 6 bits offers a good compromise between performance, delay tap complexity, and events generated. At lower resolutions like 4 bits, the quantizarion step is large and information may be lost. Thus we have chosen a 6 bits resolution to model the system.

Since the delay tap is the main contributor to the power consumption of the system, the energy per channel can be reduced by increasing the number of filters at each stage of the filter bank. The same statement is also valid for the surface per channel value.

4.4.3 Design of event-reduction and encoding circuits

The LP filter output can be subject to very closely spaced transitions. These glitches can be classified into two categories. The ones caused by delays in the combinational logic are "false" values and must be ignored (removed). The ones coming from the event-driven nature of the signal (delayed events can reach the adder in Figure 40 almost at the same time, creating two events in less than T_{gran}) are "true" transitions and should be processed by the next filter stages. However, these high-speed transitions do not carry critical information and will impact the robustness of the next filter, the time interval between consecutive events being shorter than T_{gran}. They can thus be removed as well. This can be done by eliminating successive Up and Down events appearing inside the granular delay time of the subsequent filtering stage. A glitch eliminator at the LP filter output is proposed here to both suppress the glitches and to adapt the event rate to the signal bandwidth. The modified concept for the re-encoding of the LP filter output is shown in Figure 43.

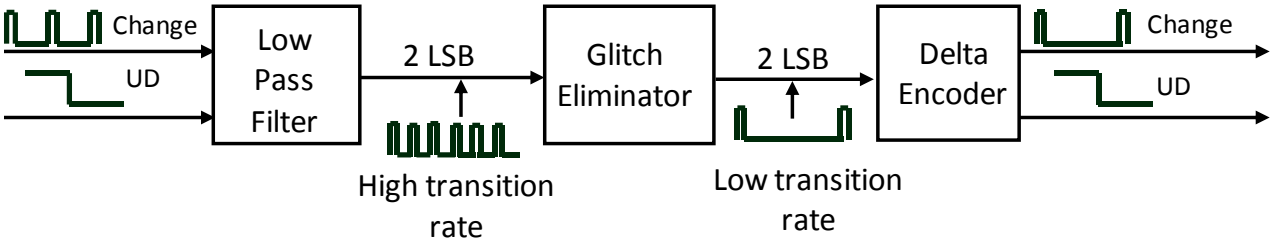


FIGURE 43 CONCEPT OF THE GLITCH ELIMINATOR ACTING AS A DECIMATOR

The proposed circuit implementation for the glitch eliminator is shown in Figure 44. When the LSB of the filter multibit output changes, the 2 LSBs are saved for a period of time given by a delay matched to a value slightly inferior to T_{gran}, due to mismatch considerations. Once this time has passed, the previously stored value is compared with the current value by the circuit logic block, which is composed of basic logic gates. If the values are identical, the transition is considered valid and the 2 LSB values are sent to the delta encoder. If the 2 LSB values are different, the transition is discarded, and the output of the glitch eliminator maintains its previous value.

This approach of processing the filter output offers an alternative to the asynchronous protocols previously used to eliminate glitches [24], where handshaking between the filter coefficients and the global adder, was implemented through matched tap delays inside each coefficient. The system we propose does not require matched delays at each coefficient, nor asynchronous protocols.

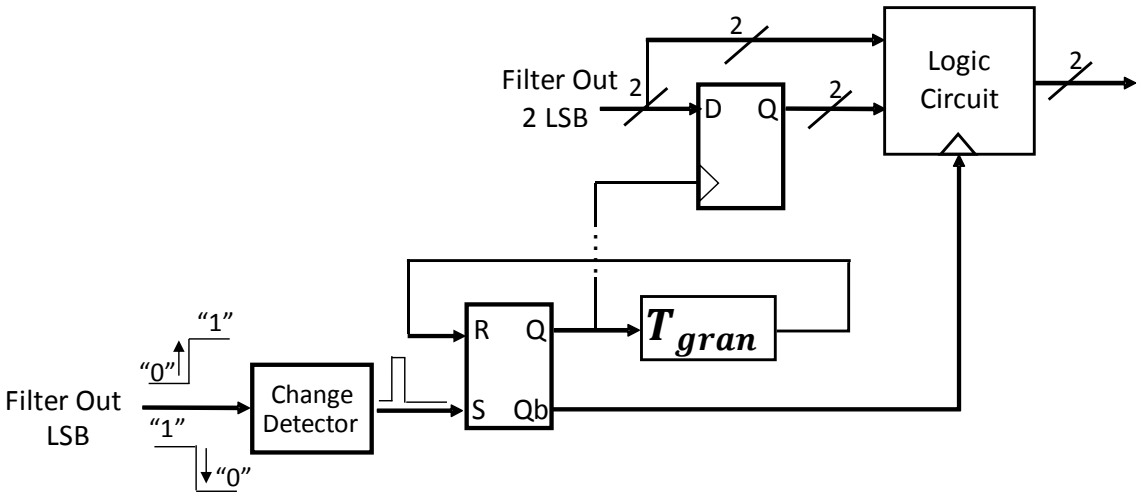


FIGURE 44 GLITCH ELIMINATOR CIRCUIT

To quantify the impact of the glitch eliminator, we have simulated the cascade of the three low-pass filters used as decimator filters in our system. The ADC has a maximum input frequency of 8 kHz and an 8-bit resolution. A full-scale signal composed of two sine waves of equal amplitude at 1 kHz and 7 kHz was fed into the system. The number of transitions was quantized at the output of each filter, over a period of 5 ms. The results are shown in Table 10. We can see that the glitch eliminator drastically reduces the number of transitions at each filter output by approximately a factor of 10, thus preventing over consumption in subsequent stages.

TABLE 10 COMPARISON BETWEEN NUMBER OF TRANSITIONS WITH AND WITHOUT A GLITCH ELIMINATOR CIRCUIT QUANTIFIED OVER A 5MS PERIOD

	Transitions without Glitch eliminator	**Transitions with Glitch eliminator**	**Ratio**
Input Signal 1kHz + 7 kHz Full scale	7112	7112	1
Output Filter 1 LPF @ 2.6 kHz	8820	994	0.11
Output Filter 2 LPF @ 1.3 kHz	11028	1194	0.10
Output Filter 3 LPF @ 440 Hz	5242	366	0.07

An analysis has been conducted to see the impact of the glitch eliminator. a full scale signal from -1 to +1, composed by two equal amplitude frequency components at 1kHz and 7kHz was fed into the system. The spectrum of the input signal can be seen at large frequency span 0-50 kHz on Figure 45.

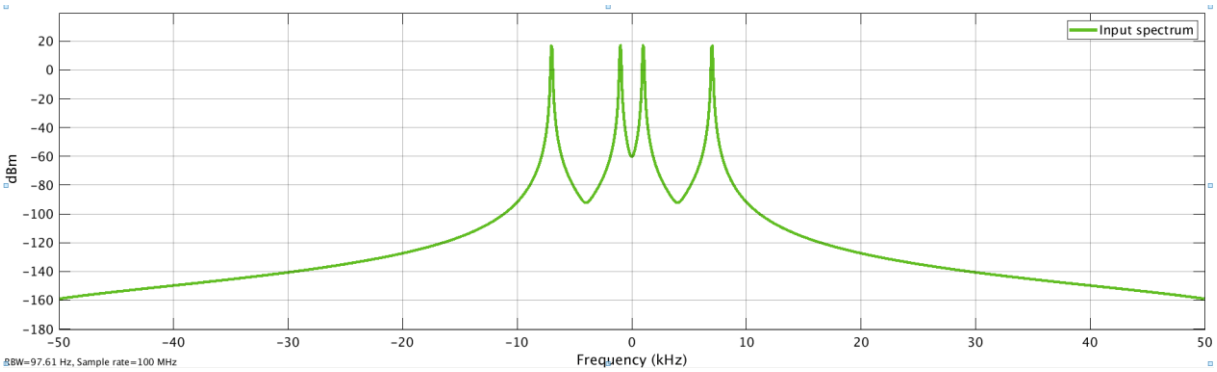


FIGURE 45 SPECTRUM OF THE SIGNAL USED TO EVALUATE THE GLITCH ELIMINATOR

The signal was fed into a two ADC one with a 6 bits' resolution and another one with an 8 bits' resolution. Figure 46 shows the spectrum of both quantized signals and the isolated quantification noise; we can see that the spectral components of the 6bits quantification error are approximately 20dB larger than the components of the 8 bits' quantification error.

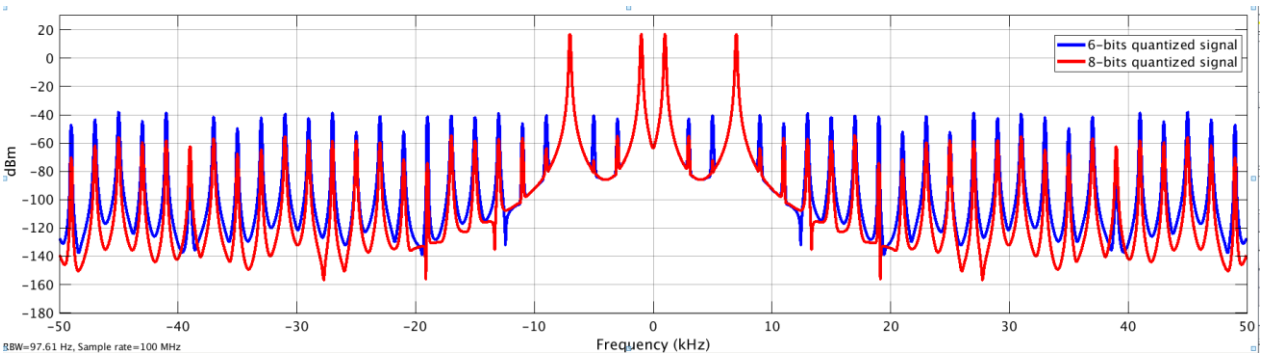


FIGURE 46 SPECTRUM OF THE QUANTIZED SIGNAL USED TO EVALUATE THE GLITCH ELIMINATOR

The 6-bit signal was fed into a CT filter, Figure 47 shows the spectrum at the filter output/glitch eliminator input and the spectrum at the glitch eliminator output. There is not any significant difference between the original spectral components of the input. The 1kHz component is not attenuated while the 7kHz component is attenuated. However, there is an important difference in the noise spectrum.

We see that the filter output in areas corresponding to the zones where the passband of the filter is replicated, the quantization spectral components aren't attenuated and remain at the same value. In the other areas spectral components are attenuated approximately 20dB, in accordance with the filter response.

At the glitch eliminator output, given the use of a delay to hold the signal value the glitch eliminator causes aliasing, increasing the spectral components due to quantization noise, that now have a constant value 20dB higher.

For our application the aliasing caused by the glitch eliminator doesn't pose a mayor issue because we target a low SNDR.

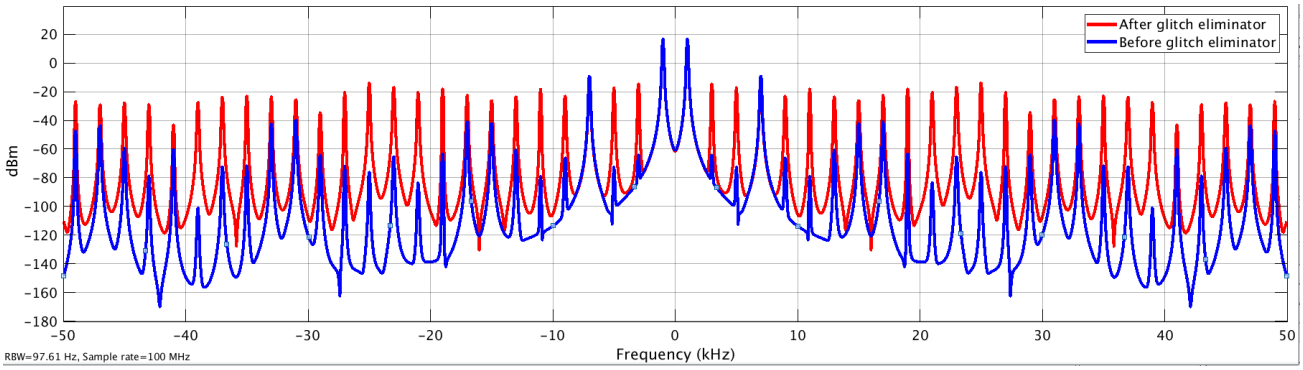


FIGURE 47 SPECTRUM AT THE INPUT AND OUTPUT OF THE CLITCH ELIMINATOR

4.5 Energy estimation

4.5.1 Energy estimation principles

The output of a CT filter is an N-bit word whose value represents the signal's amplitude, at a given time. However, this value does not provide by itself meaningful, relevant information about the spectral content of the input signal.

To generate the spectrogram required by the VAD recognition algorithm, we need to quantify the energy at the output of each filter over a given time interval. In [44] a delta encoder is used at the output of each analog filter to generate events, which are counted over a period of time to estimate the energy within each channel. In our system, this could be done by counting the Change pulses generated by the delta encoder at the output of each filter. However, the number of events is not directly correlated to the energy. Therefore, we propose an improved method to compute the energy of a signal in CT, which is described in the following sections.

4.5.2 CT instantaneous power estimation

The first step to compute the energy of a signal is to compute its instantaneous power. The CT DSP filter output is an N-bit word, whose value represents the amplitude of the signal and only varies one LSB at the time. Therefore, we can consider the current amplitude value $x(n)$ as:

$$x(n) = x(n-1) \pm 1 \tag{9}$$

The instantaneous power of a signal (*P)* is proportional to its amplitude squared:

$$P = \propto V^2 \tag{10}$$

where *V* is the amplitude of the signal therefore the instantaneous energy value of the CT DSP filter's output $x(n)^2$can be computed using the binomial square formula:

$$x(n)^2 = x(n-1)^2 \pm 2x(n-1) + 1 \tag{11}$$

Given that the filter output is subjected to glitches, we use a simplified version of the delta encoder previously as shown in Section III, in order to convert the filter's output into a pair of signals Change and U/D free of glitches and unwanted transitions. The implemented circuit is shown in Figure 48, and its functioning is the following: Using the Change and U/D signals, the values $x(n)$ and $x(n-1)$ are reconstructed. A multiplexer driven by UD is used to choose between the two different scenarios: The first when $x(n) = x(n-1) + 1$, UD = 1, in which the $2x(n-1)$ multiplication is implemented by simply wiring the N bits at the multiplexer output to the N bits of higher

weight of the adders $\pm 2x$ input; the second one when $x(n) = x(n-1) - 1$, UD = 0, in which the two's complement of $x(n-1)$is required. The two's complement of a number x C_2^x is given as

$$C_2^x = C_1^x + 1 \tag{12}$$

Where C_1^x is the one's complement of x and is given as

$$C_1^x = \overline{x} \tag{13}$$

The one's complement of $x(n-1)$, is obtained with an inverter, and it is also multiplied by two by wiring the N bits at the multiplexer output to the N bits of higher weight of the adders $\pm 2x$ input. A second multiplexer is used to add the +1 part of the binomial equation. Given that $2 * C_2^x = 2C_1^x + 2$, +3 needs to be added in the second case. Consequently, when UD = "1" the +1 value is selected thus completing the binomial square formula, otherwise when UD = "0" the +3 value is selected.

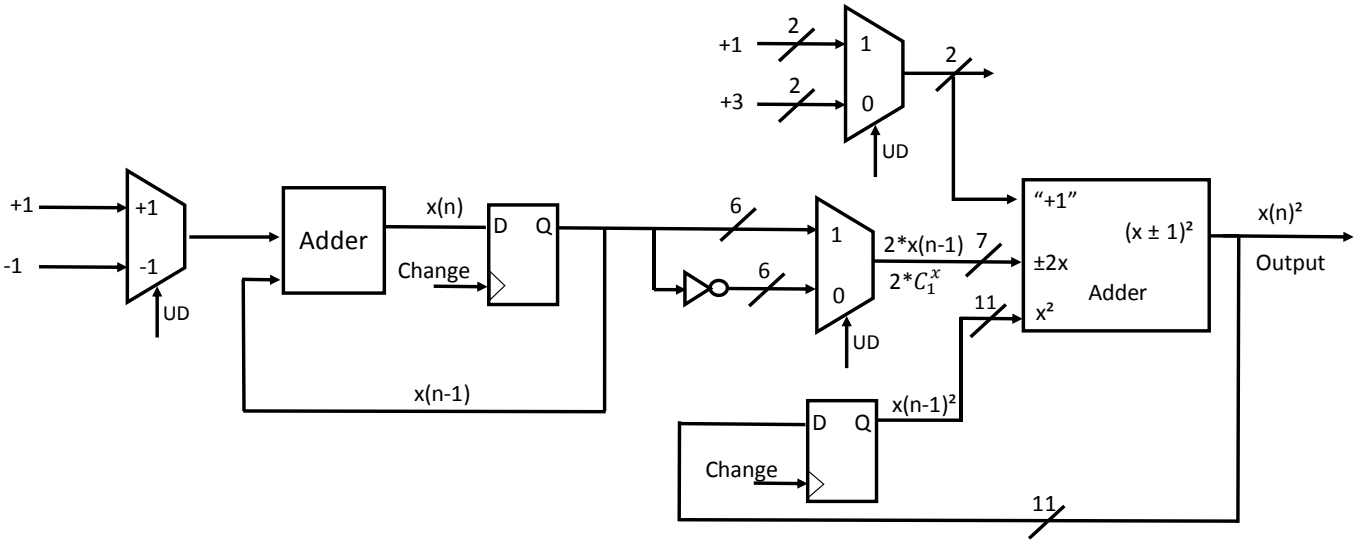


FIGURE 48 PROPOSED CIRCUIT TO CALCULATE THE INSTANTANEOUS POWER OF A SIGNAL IN CT

4.5.3 Energy accumulation

Finally, the instantaneous power is accumulated to compute energy E(n) within a given time slot. There are two approaches to perform this operation. The first is an event-based approach where the instantaneous power is accumulated at each Change pulse. The second is a time-based approach where the instantaneous power value is sampled and accumulated using a clock at the Nyquist rate. Although the time-based approach is more accurate, the event-driven approach has been chosen to keep consumption low. A very low-frequency clock (typically 10-100 Hz) gives the period over which the power is accumulated. Figure 49 shows the circuit used for this purpose, at the end of each clock cycle, the final calculated energy value is latched and transferred to the recognition system and the accumulator is reset. The use of this low-frequency clock, much lower than the Nyquist rate, has a negligible impact on the total power consumption of the system.

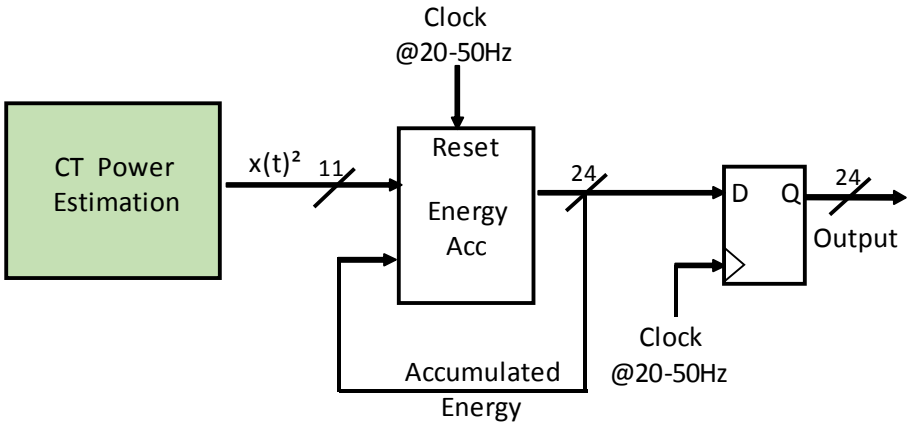


FIGURE 49 CIRCUIT TO COMPUTE THE ENERGY OF A SIGNAL BY ACCUMULATING INSTANTANEOUS POWER

Figure 50 shows the comparison between the proposed method and the event counting approach to evaluate energy. Simulations of the 440 Hz LPF decimator in Fig. 2 were made for different input frequencies logarithmically distributed from 10 Hz to 10 kHz. The energy was accumulated over a 10ms period. The energy values were normalized with respect to the maximum in band-value. We can see that our proposed solution better complies with the theoretical filter response, improving by approximately 20dB the rejection in the stopband. The drop of measured energy at low frequencies in the pass band is due to the reduced event rate for very-low-frequency sinusoidal signals. In the presence of noise or complex signals this effect disappears.

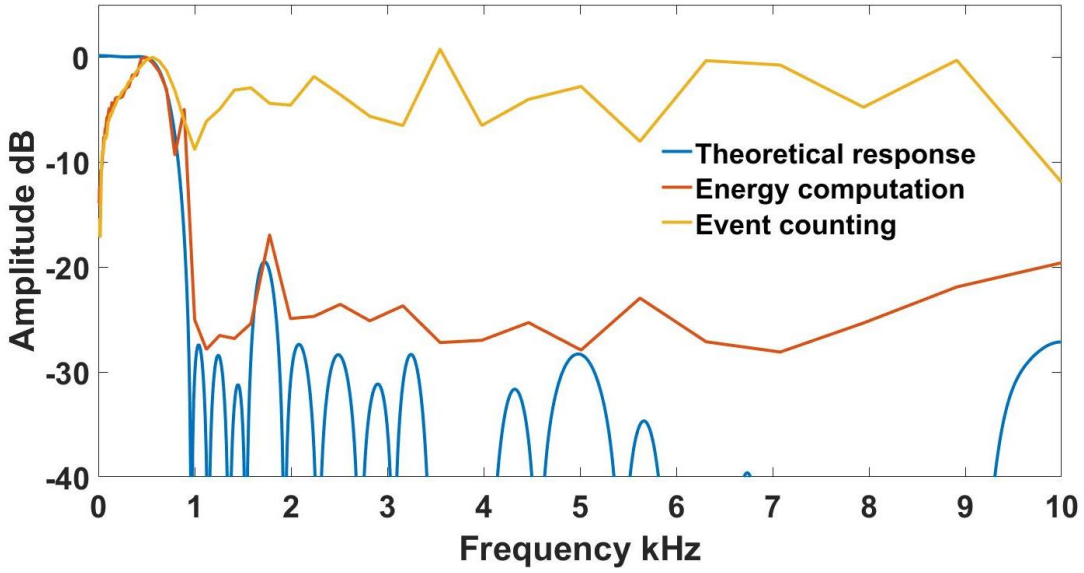


FIGURE 50 COMPARISON OF FILTER TRANSFER FUNCTIONS OBTAINED FOR THE EVEN-COUNTING METHOD AND THE PROPOSED ENERGY CALCULATION METHOD, WITH RESPECT TO THE CONTINUOUS-TIME EQUIVALENT FILTER FUNCTION (BLUE CURVE)

Figure 51 shows an alternative to interface the energy accumulator block with the synchronous world. In this circuit the output of the circuit used to accumulate instantaneous power is compared with a threshold value, using an asynchronous comparator. Each time the energy accumulator value is larger than the threshold, a reset pulse is generated. A asynchronous counter counts all the pulses generated by the comparator. The counter is read and set to zero with a low-frequency clock.

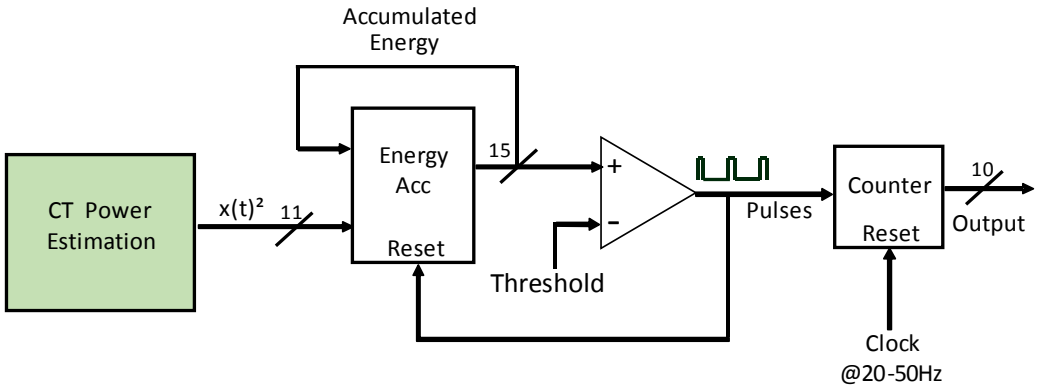


FIGURE 51 ALTERNATIVE CIRCUIT TO COMPUTE THE ENERGY OF A SIGNAL IN CT

4.6 System performance

To validate our system, we generate a spectrogram from an audio recording of spoken vowels, which is a popular case used for VAD systems [4]. The recording was generated with standard CD quality (16-bit at 44.1 kHz sample rate). Figure 52 (a) shows the spectrogram of 3 spoken vowels obtained with our proposed filter bank with a 6-bit ADC resolution, and energy quantization over a period of 20 ms with a 6-bit resolution. Figure 52 (b) is computed with a 16-bit 512 point FFT at a sampling frequency of 16kHz and integration of the result over the same frequency bands and time period, with final energy results quantized to 6 bits. Both graphs are very similar, showing the performance of our system approach. The aliased noise causes the undesired presence of energy in some features but remains more than 20 dB below full scale.

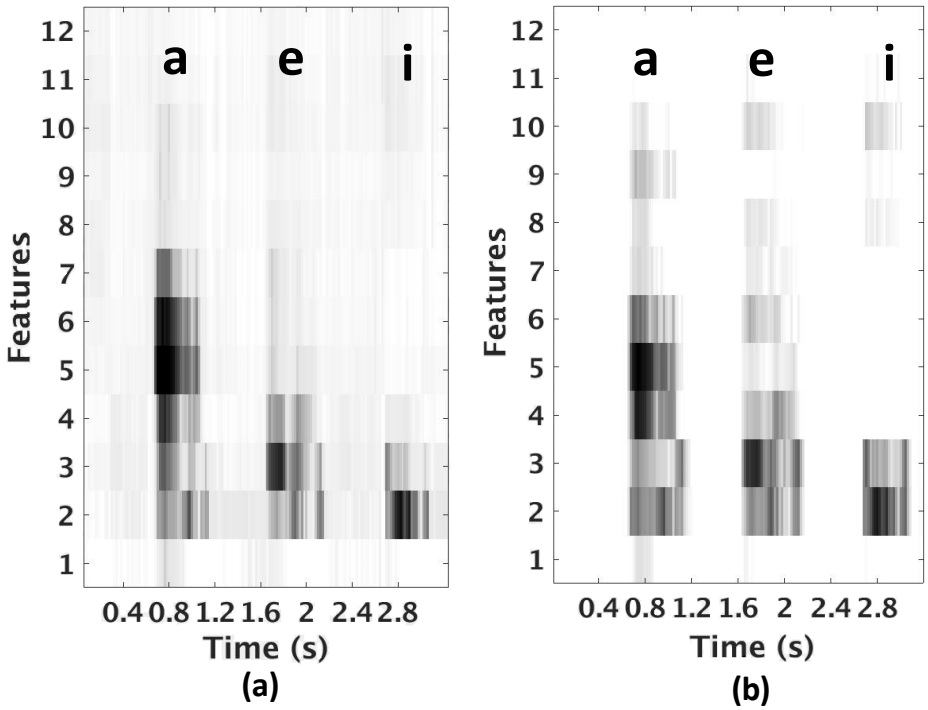


FIGURE 52 COMPARISON OF SPECTROGRAMS OF SPOKEN VOWELS GENERATED WITH (A) OUR PROPOSED SYSTEM, AND (B) AN FFT ALGORITHM

We make estimations on the energy consumption and surface of the system to quantify its performance. We quantify the performance per channel and the overall system performance.

Given the event-driven nature of the system its energy consumption has two components: the static consumption which is dominant during stand-by periods of non-activity, and the dynamic consumption, that varies in function of the activity at the input. To estimate the dynamic energy consumption, we considered a worst-case scenario, in which a full-scale signal at a particular frequency is present at the input and a typical scenario in where male speech or female speech is processed.

To evaluate a worst case scenario power consumption, we considered a 440Hz single tone, full-scale amplitude signal. Table 11 and Table 12 show the overall dynamic power consumption of the system in a worst-case and a typical scenario respectively, decomposing the contribution of the delay elements and the digital signal processing. The 440 Hz signal will go through the 3 decimators of the cascade without suffering any attenuation, and thus the consumption of each of the four delay taps will be identical. For purpose of the estimation we considered that each delay element has a dynamic power consumption of 25 fJ/event. To estimate the energy consumption of the digital processing unit, we considered a dynamic power consumption of 0.7fJ per event per equivalent NAND-gate. In total, the dynamic energy consumption of the digital signal processing (DSP) unit is estimated to be 2.78pJ per event.

TABLE 11 WORST CASE SCENARIO DYNAMIC POWER CONSUMPTION

ADC Resolution Bits	Delay Consumption (µW)	DSP Consumption (µW)	Total Consumption (µW)	Consumption per Band (µW)
6	14.2	3.7	18	1.5

TABLE 12 TYPICAL CASE SCENARIO POWER CONSUMPTION

ADC Resolution Bits	Delay Consumption µW	DSP Consumption nW	Total Consumption µW	Consumption per Band µW
6	1.62	0.420	2.04	170 nW

Since, most of the time, no activity will be present at the input, it is important to reduce the static power consumption as much as possible. Arbitrarily, we considered that the static power consumption value should be at most 20% of the dynamic power consumption of a typical speech signal. Thus, we target a static power consumption value of 400nW.

4.7 Alternative implementation approach by using low pass filters and energy subtraction

The energy estimation functions open the way to implement a new method to create band-pass filters functions, which is shown in Figure 53. In this method low pass filters are cascaded; next the energy is estimated at the output of each low pass filter. Finally, the energy difference between the filters energy is quantified. Allowing to creating an equivalent band-pass filter function.

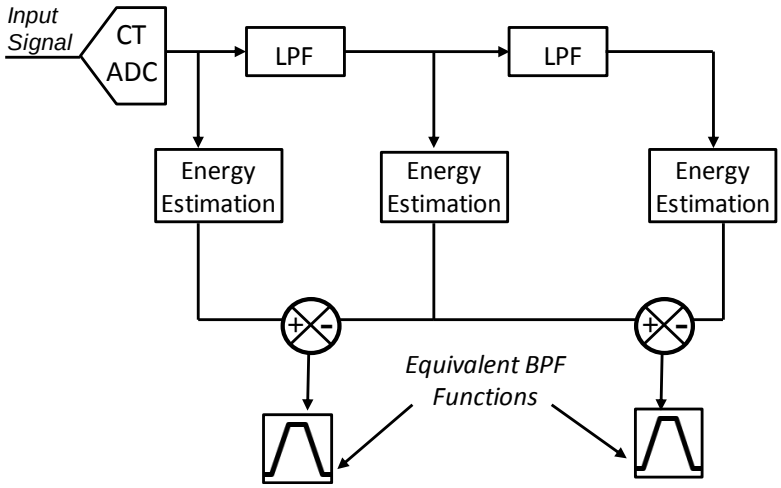


FIGURE 53 DIAGRAM OF BAND-PASS FILTERS CREATED BY SUBTRACTING ENERGY OF LOW PASS FILTERS

The principle was implemented using a 10th order FIR filter shown in Figure 54. The filter coefficients were obtained in MATLAB, so that filter ripple is minimized. The LPF filter function was replicated to obtain four low-pass filters whose frequency response is shown in Figure 55, all filters have a pass-band with no ripple and a stop band rejection of at least 20 dB. Figure 56 shows the theoretical frequency response of the band-pass filters, obtained when subtracting the frequency responses of the low-pass filters.

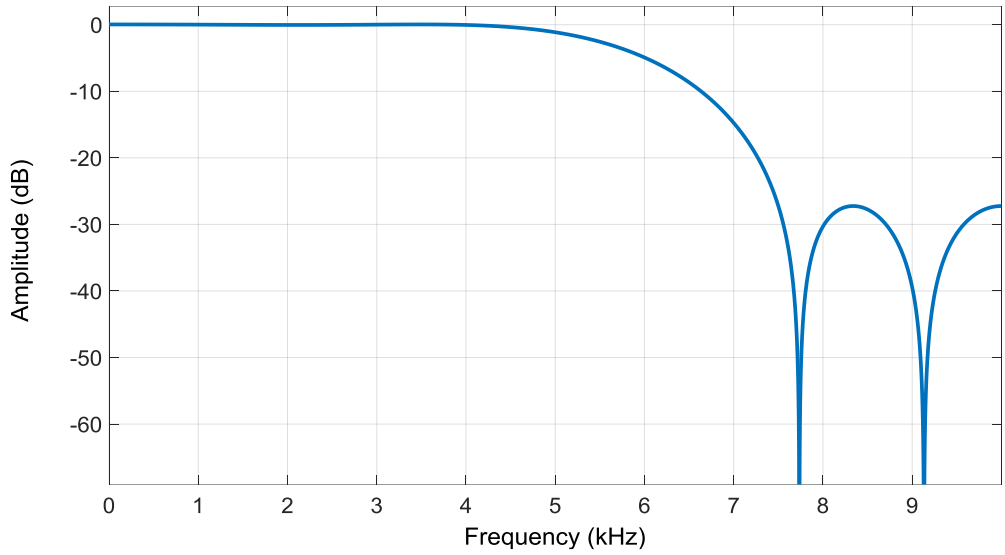


FIGURE 54 LOW-PASS FILTER USED TO CASCADE FILTERS

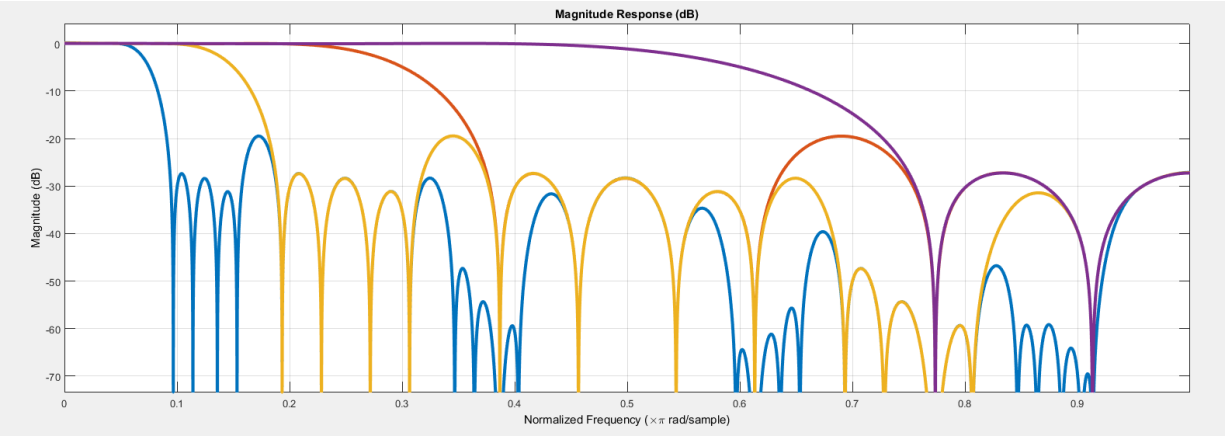


FIGURE 55 FREQUENCY RESPONSE OF THE CASCADED LOW-PASS FILTERS USED TO CREATE BAND-PASS FILTERS

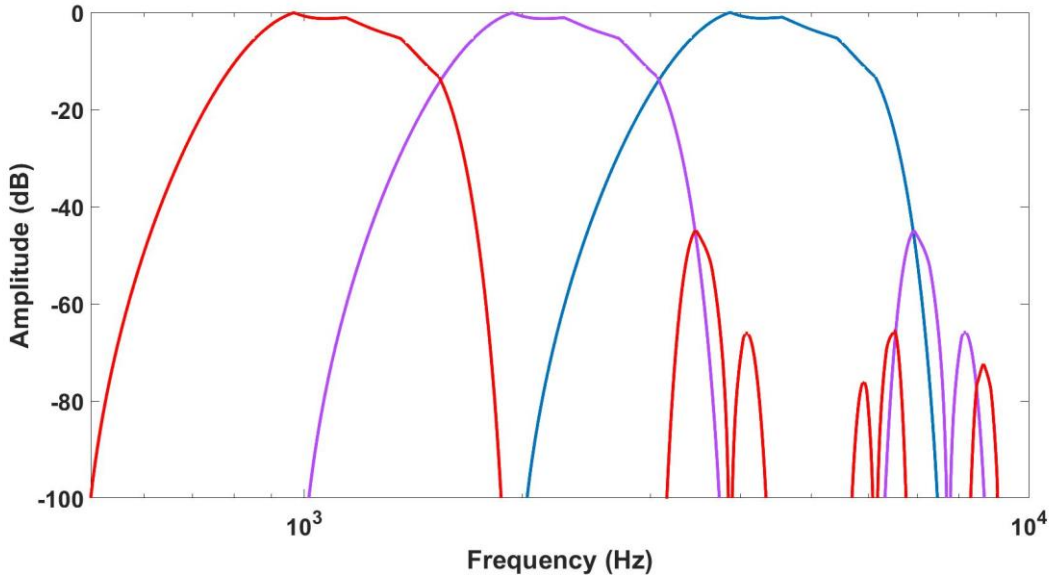


FIGURE 56 FREQUENCY RESPONSE OF THE THEORETICAL BAND-PASS FILTERS

Figure 57 shows a comparison between the theoretical frequency response, the frequency response obtained when events are quantized, and the frequency response obtained with the energy estimation function for a band-pass filter, between 625 Hz to 1.25 kHz. Results are coherent with Figure 50. The energy estimation function allows having a stop-band rejection in accordance with the theoretical transfer function. At frequencies below 100 Hz, the stop band rejection of the filter function obtained by computing energy is perturbed, because the period of the input signal is equal or larger than the integration step used to quantify energy. There are three different factors affecting the stop-band rejection of filters, and have to be optimized:

1. The normalization
2. The length of the integration step
3. The ADC resolution

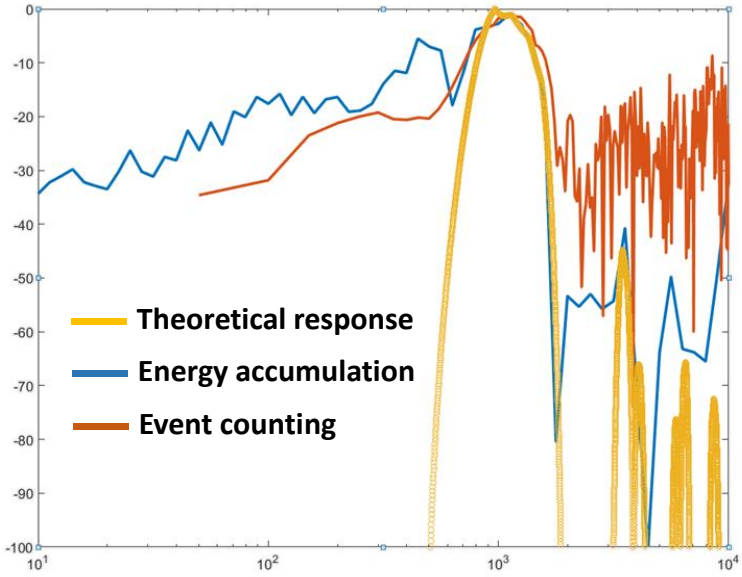


FIGURE 57 COMPARISON BETWEEN EVENT COUNTING AND ENERGY ACCUMULATION

Figure 58 compares two different normalizations, the response in green was obtained by normalizing each point with regards to their corresponding frequency value, the response in red was normalized by dividing the energy quantified at each frequency, by the maximum energy value obtained, and that corresponds to the central frequency of the filter pass-band. Results are coherent with those obtained in Figure 50.The stop-band rejection at lower frequencies is drastically improved (around 12 dB) by normalizing each energy value with the maximum energy value. At higher frequencies, both normalizations offer at least 20 dB stop band rejection, however energy normalization considering frequency values offers a better stop-band rejection.

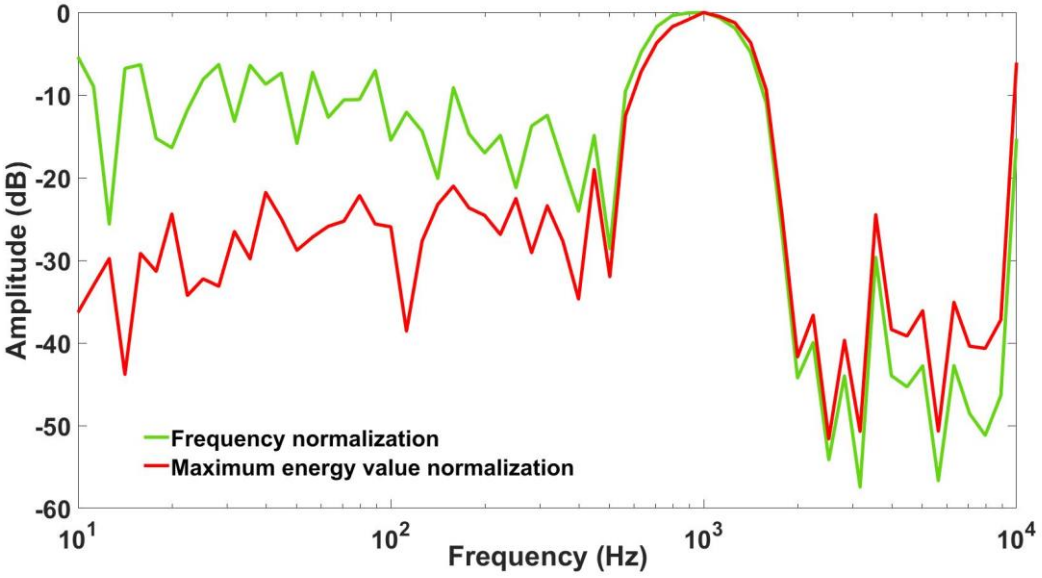


FIGURE 58 COMPARISON BETWEEN DIFFERENT NORMALIZATIONS

Figure 59 shows a comparison of the filter response, obtained for two different quantization time steps. The response in red was obtained with a 10-millisecond quantification, while the response in blue was obtained with a 20 millisecond quantification. The difference in both responses can be found at lower frequencies, where the stop band rejection is improved by the larger integration time by around 5 dB. At higher frequencies, the integration step does not impact the stop-band rejection.

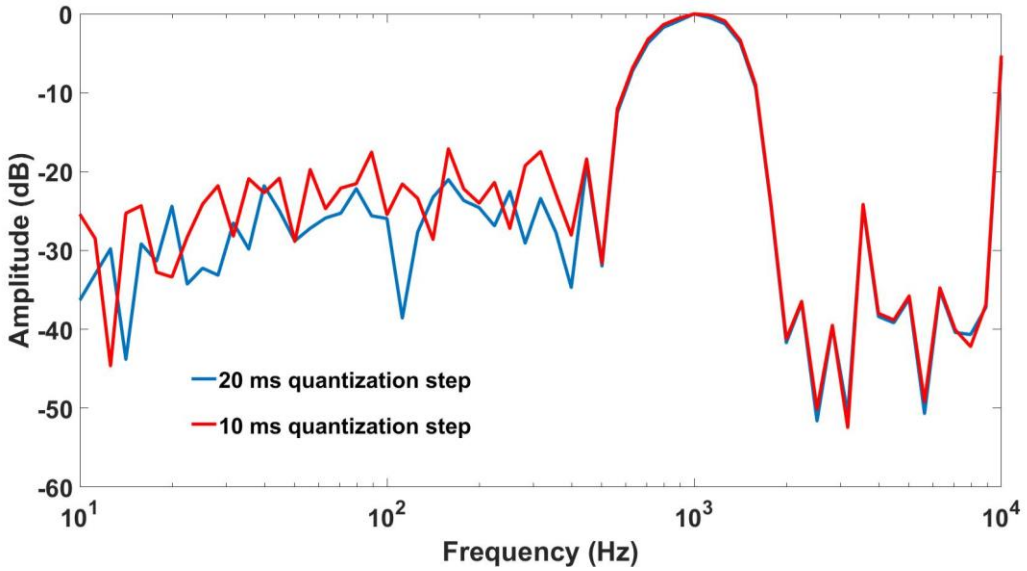


FIGURE 59 COMPARISON OF FILTER FUNCTION OBTAINED WITH ENERGY ESTIMATION FOR DIFFERENT INTEGRATION STEPS

Figure 60 shows a comparison of a filter transfer function, obtained by computing the energy of a signal for different ADC resolutions. A higher ADC resolution allows to have a better stop-band rejection at higher frequencies. Nonetheless a 6 bits' resolution still offers a 20 dB stop-band rejection.

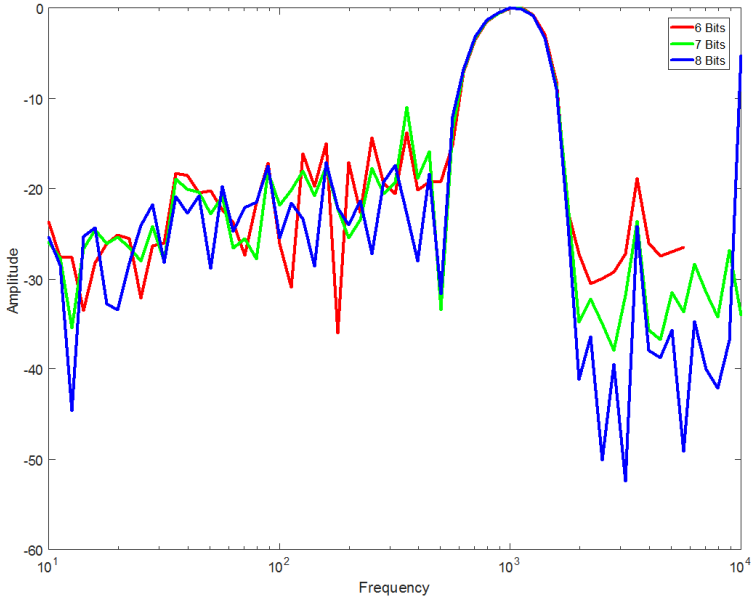


FIGURE 60 COMPARISON OF FILTER FUNCTION OBTAINED WITH ENERGY ESTIMATION FOR DIFFERENT ADC RESOLUTIONS

The approach to create equivalent bandpass filters functions by sustracting the energy at the output of 2 cascaded low-pass filters, has been demonstrated. However in the case of the proposed fiter bank, this approach is not the most efficient, since 12 cascaded delay taps will be required to implement the system. Since the delay taps are the most hungry block, the energy consumption will increase significantly, for this reason this approach has been discarded for this application.

4.8 Conclusions

In this chapter, we proposed the implementation of a CT filter bank with energy extraction functions to obtain the spectrogram of an audio signal for audio recognition applications. The system combines the event driven nature of analog implementations with the configurability of digital implementations, while being scalable with advanced technology nodes.

Architectural choices of the ADC and the filter transfer functions were made. To comply with the system demands, novel circuits to cascade FIR filters efficiently and to compute the energy of a signal in continuous-time were designed. Energy quantization at the filter outputs with efficient hardware implementations allows transferring at low cost the features to the recognition system.

The granular delay element is the key element to improve the energy consumption and silicon area of the system. For each additional bit of ADC resolution, the number of delays and the number of events doubles, resulting in increase of the dynamic consumption of the delay lines by a factor 4. Therefore, the delay elements are dimensioned to accommodate the samples generated by an ADC with a 6-bit resolution, which is the minimum resolution required to generate a proper spectrogram to send to the recognition algorithm.

Although estimations show that for a worst-case scenario the system power consumption is higher than existing analog and digital implementations, its event-driven nature allows the system to efficiently process most typical cases signals with a low data rate, and thus comply with the overall power consumption target (1 – 5 µW).

In order for the system to attain its desired performance, the key element is a delay cell with a set of specifications that will be used as a target for the rest of the thesis.

Chapter 5
Digital Delay Cells in 28 nm FDSOI technology

5.1 Description

The previous chapter concluded that our proposed continuous-time filter bank for feature extraction requires the use of digital delay cells, with very constraining performance requirements. Thus, the focus of this chapter is the design and implementation of the digital delay cells. We begin this chapter with a brief state-of-the-art of digital delay elements. The choice of the technology, the 28 nm fully depleted silicon-on-insulator (FD-SOI) planar ultra-thin body and box (UTBB) will be discussed versus design requirements. A detailed description of the proposed delay-cell will follow. To demonstrate the functionality and performance of this delay, a circuit has been designed. This circuit contains individual delays and eight delay taps (several hundreds of delay elements in total) that can be used to build a CT filter demonstrator, using an FPGA for digital processing. The chapter concludes with a comparison of the delay element performance with regards to the current state-of-the-art and with an estimation of the overall system performance.

5.2 Demonstrator strategy

To demonstrate the functionality of the proposed concepts we aim to fabricate a demonstrator in 28-nm FD-SOI technology. The FD-SOI node has been selected because its structure offers a lower leakage current that conventional bulk technologies and allows us to control small current values.

Our objective is to demonstrate the functionality of the following:

- Energy efficient digital delay cells with a large tuning range.
- Cascaded FIR filter functions in CT.
- Instantaneous power and energy estimation functions.
- Complete system validation.

For the sake of simplicity, we aim at designing an integrated circuit containing the digital delay taps. It will be later integrated with a Field Programmable Gate Array (FPGA) containing the digital processing functions.

Our step by step test plan is the following:

1. Characterization of the unitary delay line elements.
2. Characterization of the delay taps.
3. Implementation of a single FIR filter and energy estimation functions.
4. Implementation of cascaded filters.
5. Complete system implementation.

5.3 Digital delay line state of the art

Figure 61 shows the block level schematic of a typical delay cell in a CT digital filter. The delay cell is composed of three blocks: the delay block providing a delay tuned by a control current, the handshaking block to validate the correct propagation of the CHANGE signal through the delay tap and a D-latch to propagate the data bit.

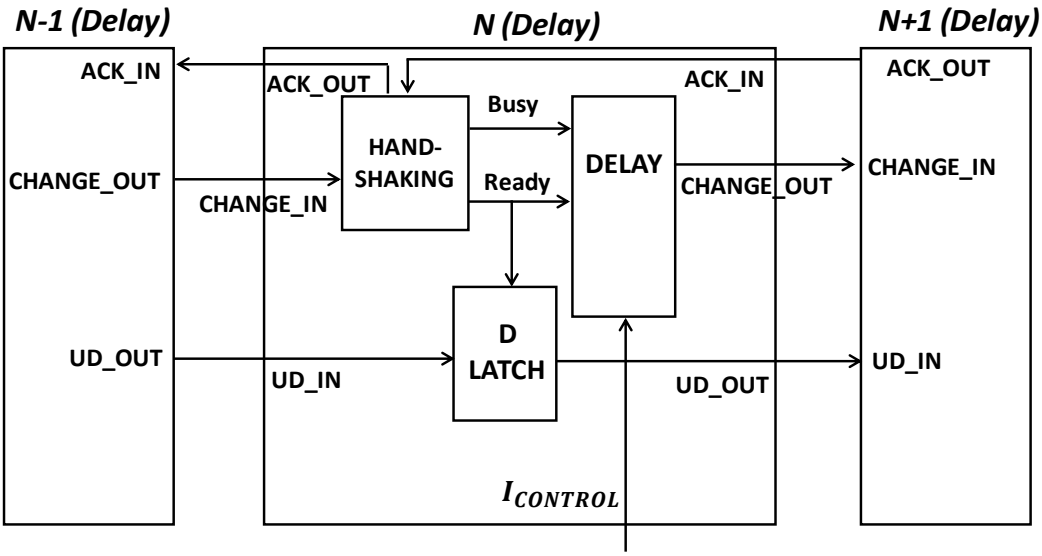


FIGURE 61 TOP LEVEL SCHEMATIC OF A DELAY CELL IN A CT DIGITAL FILTER

The principle to create a delay in electronics is given in (14). An analogy can be made with a sand clock. The capacitor is the recipient; the flow of sand is the current. Typical implementations are listed in [48] straightforward chain of inverters, capacitive shunting, current starving and thyristor-based delay. All these implementations use the same principle of controlling the flow of a current that charges or discharges a capacitor. The delay time (τ) is directly proportional to the capacitor value (C) and the voltage dynamic (ΔV), and inversely proportional to the current flow (I).

$$\tau = \frac{C*\Delta V}{I} \tag{14}$$

The dynamic energy consumption E, is given by

$$E = C * {\Delta V^2}/{2} \tag{15}$$

Thus, to reduce dynamic power consumption the capacitor value and the voltage dynamic should be minimized. For a given delay value, this is only possible with a current reduction. Therefore, the minimum energy consumption of the delay cell elements is limited to our ability to control the smallest possible current value.

Our digital-delay cell will have to fulfill the following constraints as determined in Chapter 4.

- It must delay pulses whose width is between 10 to 50ns over a period as long as 16µs.
- To cope with the asynchronous nature of signals, and delay variability a proper handshake mechanism is required.
- Ideally, to increase configurability and to ease implementation, the same block should have a tuning range covering at least from 800 ns to 16µs.
- A mechanism to propagate the UD signal across the tap

None of the existing delays in the state-of-the-art, comply with all our requirements.

The first paper proposed by [49], presents an architecture implemented in 65nm technology, as shown in Figure 62.

The main feature that distinguishes the delay is the presence of a robust hand-shaking mechanism, that guarantees both synchronization and mismatch robustness. The paper also focuses on reducing delay mismatch. This is achieved by using a long transistor (L=5µm) as the current source to guarantee a large output resistance. The transistor is split into five transistors connected in series to improve the reset conditions. The mismatch is also improved by operating the current source in strong inversion, which is a penalty when targeting low power-consumption and low area.

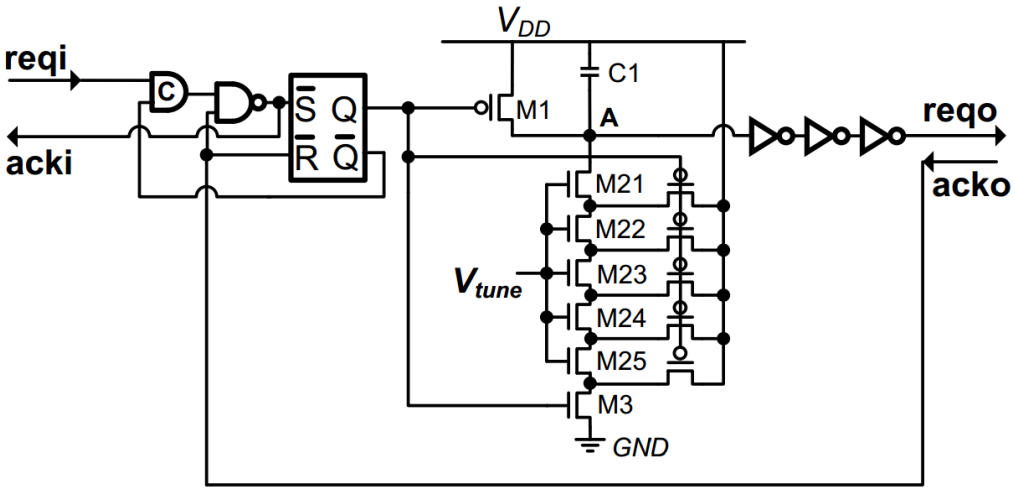


FIGURE 62 DELAY ELEMENT PROPOSED BY [49]

The second is a CMOS thyristor-based delay element proposed by [50] shown in Figure 63, implemented on 90nm technology.

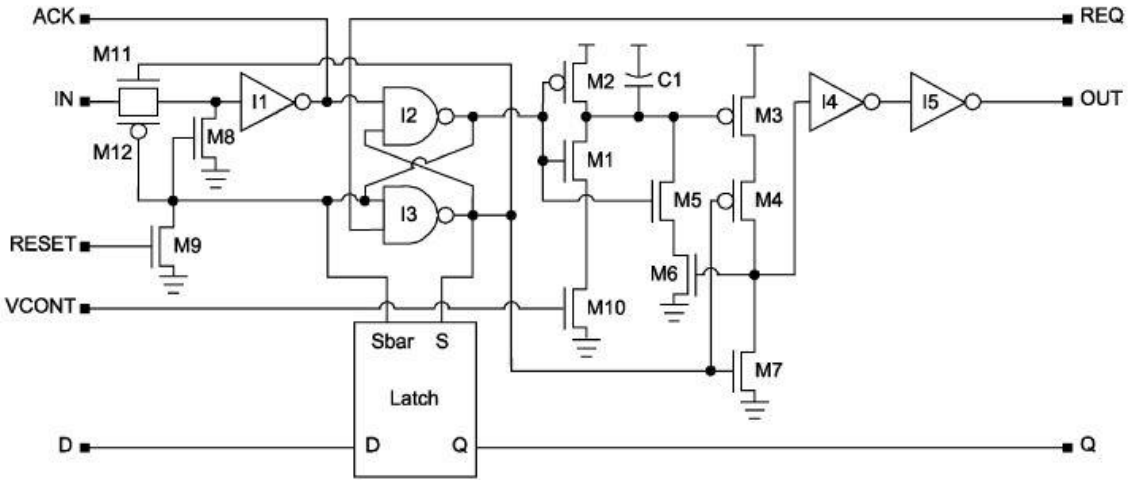


FIGURE 63 DELAY ELEMENT PROPOSED BY [50]

The CMOS thyristor is a circuit implemented using a pair of NMOS (M6) and PMOS (M7) transistors that are activated when a certain threshold is crossed as shown in Figure 64. The functioning principle shown in Figure 65 is the following. In an initial state both transistors are off. The PMOS gate is connected to VDD and the NMOS gate is connected to ground. The threshold value is given by the PMOS. The switches are controlled by a pair of complementary signals Ready and Busy generated by a Set-Reset (SR) latch. When a pulse arrives at the input, the capacitor that fixes the voltage at the PMOS gate, begins to charge through a slow charge-current source. When the voltage at its gate drops below the threshold level, the PMOS transistor behaves like a closed switch, thus activating the NMOS transistor, which behaves as a fast discharge current source, which rapidly finishes the discharge of the capacitor, completing the operation. Once the operation has been completed, an acknowledge signal is generated to reset the SR latch, set the thyristor delay to its initial state.

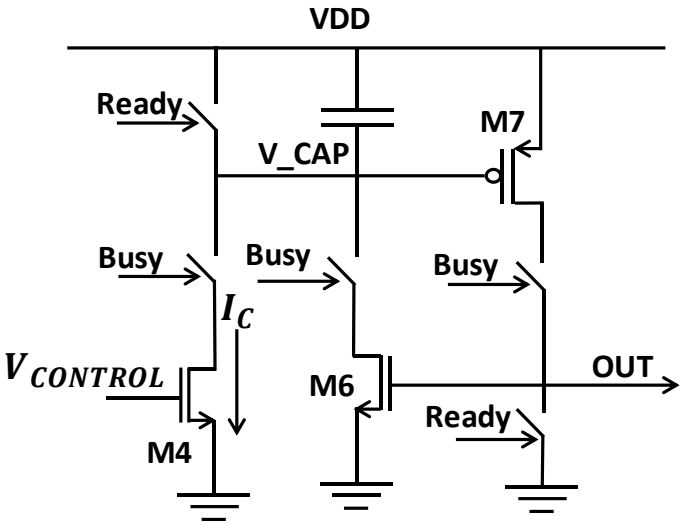


FIGURE 64 THYRISTOR BASED DELAY ELEMENT

The advantages of the delay presented by [50] are a reduction in surface and power consumption obtained by the use of a transistor in weak inversion as current source. On top of that the use of a thyristor will give more flexibility to reduce voltage dynamic, further improving the energy consumption. The disadvantage of existing implementations is that the technology node limits the ability to control low currents, preventing to generate delays longer than 1 µs. Additionally existing implementations of this structure do not have a robust mechanism for handshaking.

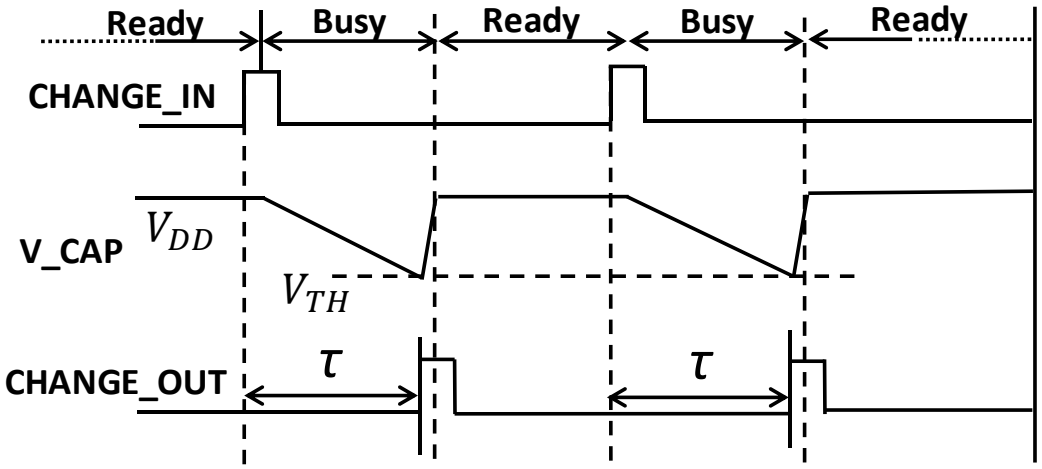


FIGURE 65 FUNCTIONNING PRINCIPLE OF THE THYRISTOR BASED DELAY

The state of the art of digital delay elements is shown in Table 13.

TABLE 13 STATE-OF-THE-ART OF DIGITAL DELAY ELEMENTS

Parameter	Vezyrtzis, 2014 [51]	Chang, 1996 [52]	Schell, 2008 [50]	Sourikopoulos, 2016 [53]	Kurchuk, 2010 [54]	Chen, 2018 [49]
V_{DD}	1 V	2V	1V	1V	1.2V	1.2V
Tuning range	15-500 ns	2.5 ns - 76.3 ms	5 ns – 1 µs	110ps - 4.5 ns	0.3 - 300 ns	5 ns - 10µs
Matching	-	-	-	-	12.4%	2.3%
Energy/delay	50 fJ	60fJ	50 fJ	12.5fJ	20 fJ	83fJ
Area	-	-	36 µm²	21 µm²	-	97µm²
Input-independent Delay	Yes	No	Yes	Yes	Yes	Yes
Robustness	Potential Hazard	No handshaking	Potential Hazard	No handshaking	No handshaking	Hazard-free handshaking

5.3.1 Conclusion

To fulfill our requirements, we will need to design a dedicated circuit, using a handshaking structure similar to [49] and a thyristor based delay, with a current source operating in weak inversion, taking advantage of an advanced technology to deliver a very low current (inferior to 1nA).

5.4 Technology choice

The identified criteria to select the technology are: leakage, mismatch, and a good subthreshold slope to attain a large tuning range while keeping the transistor used as current source in subthreshold operation. The combination of these factors allows the control and generation of low currents.

There are four candidates: “conventional” bulk, partially depleted silicon-on-insulator (PD-SOI), FD-SOI and Finfet technologies. We begin the discussion by speaking about bulk technologies, from where the other technologies have been derived.

Bulk technologies have not been selected for our application, because their structure produces leakage currents from the drain and the source to the substrate. These leakage currents become significant when targeting ultra-low power consumption and a large body factor (relationship between the coupling capacitances from the gate to the channel and the channel to the bulk), that limits the subthreshold slope (S) given by

$$S = n\,\frac{kT}{q}\ln(10) \tag{16}$$

The PD-SOI technology consists in the use of a layer of oxide below the transistor that limits the leakage currents to the substrate. However, it does not improve neither subthreshold slope because the dimensions of the channel are comparable to bulk CMOS, nor mismatch due to channel random dopant fluctuations (RDF).

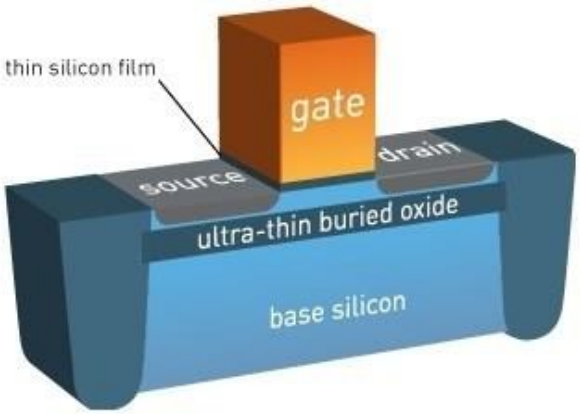


FIGURE 66 STRUCTURE OF AN FDSOI UTBB TRANSISTOR

FD-SOI shown in Figure 66 can be seen as a variant of PD-SOI that uses the ultra-thin body and Box (UTBB) technology, where the film (silicon above the oxide) is very thin (some nm) and fully depleted. FD-SOI take advantage of the ultra-thin oxide to reduce leakage as shown in Figure 67 and has low mismatch thanks to the undoped channel. On top of that the control over the channel is improved due to the thin film, giving the transistor a body factor *n* approximately equal to 1.05 (compared to around 1.5 for bulk CMOS) resulting in a near-ideal subthreshold slope of 60mv/Dec (Equation 16).

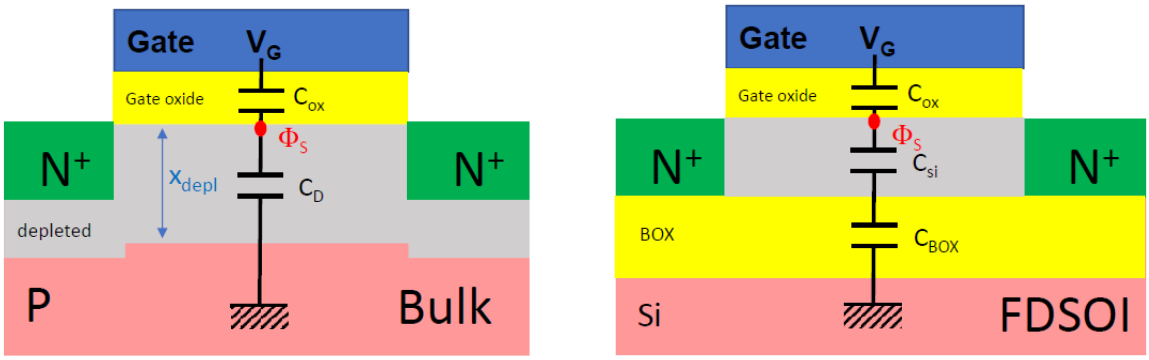


FIGURE 67 COMPARISON BETWEEN BULK AND FD-SOI TECHNOLOGIES

The properties of Finfet technology transistors are similar to those of FD-SOI, and thus Finfets could also be a good candidate for our application. However, FD-SOI was readily available for prototyping and has therefore been chosen for prototype implementation.

In our implementation we can benefit from other aspects of FD-SOI:

- Improved drain induced barrier lowering DIBL (thanks to thin film)
- Capability to tune threshold voltage trough body biasing
- Larger triode and saturation currents
- Larger transconductance for a given current

$$I_D = \mu C_{ox} \frac{w}{L}\left[(V_G - V_{TH})V_D - \frac{1}{2}nV_D^2\right] \quad (17)$$

$$I_{Dsat} = \frac{1}{2n}\mu C_{ox} \frac{w}{L}(V_G - V_{TH})^2 \quad (18)$$

$$\frac{g_m}{I_D} = \sqrt{\frac{2\mu C_{ox}\, W/L}{nI_D}} \quad (19)$$

Since our application does not requires a very high speed operation, the gate length is not a critical parameter, and thus we select the largest FD-SOI technology available the 28nm node.

5.5 Digital-delay cell design

5.5.1 Current source in weak inversion

The delay time of our circuit will be given by a capacitor that is charged with a current source. From our architecture study, we target current values in the range of 250 pA as shown in chapter 4. Since our power estimations only consider the consumption for charging and discharging the capacitance, we made the decision to size the current source to be able to operate properly down to 100pA. The transistor that will be used as a current source is polarized in weak inversion or subthreshold operation, in order to attain the largest possible delay for a given capacitance,

The drain current I_D of the transistor operating in weak inversion is given by:

$$I_D = I_{D0}\frac{W}{L}e^{V_{gs}/nV_t}(1 - e^{V_{ds}/V_t}) \quad (20)$$

Where:

$n \approx 1.05$ for a FD-SOI transistor

$V_t = \frac{kT}{q} \approx 26\,mV$ at $T = 300°K$

$I_{D0} = \mu_n C_{ox} V_t^2 e^{-V_{TH}/(nV_t)}$

The following considerations are taken into account when sizing the transistor that will be employed as a current source and when choosing the operating point:

- In the I_{D0} equation the impact of the channel length modulation is ignored, thus a long transistor is chosen. Due to the low current targeted length L will be likely larger than width W.
- To eliminate the dependency of the Ids current on the VDS voltage, the VDS voltage must be several times larger than the V_t voltage. Therefore, VDS must be at least 100mV. In which case we can assume that the transistor is working in the saturation region for weak-inversion operation.
- To assure that the transistor is operating in subthreshold operation the V_{gs} voltage must be at maximum $V_{gs} \leq V_{TH} - 100mV$
- We choose L to be as high as possible without compromising mismatch, and thus we select L = 500nm, according to the Avt curve shown in Figure 68.

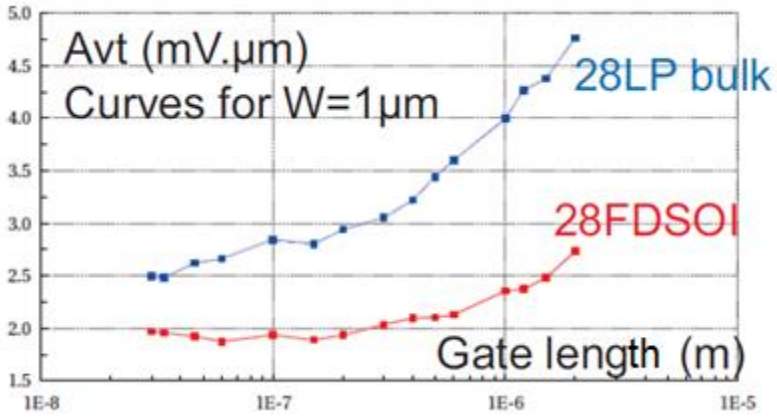


FIGURE 68 AVT COMPARISON BETWEEN THE 28NM BULK AND THE 28NM FD-SOI TRANSISTORS [55]

The 28nm FD-SOI technology provides two flavors of MOS transistors, low voltage threshold (LVT) and regular voltage threshold (RVT). For weak inversion operation, a RVT transistor is chosen because of smaller leakage currents (Ioff).

Several simulations were run for different W values; we finally select a W=300nm. Figure 69 shows the curves of Ids current vs Vds voltage for this transistor

size. Results have shown that we obtain a current value within the targeted range and a proper flat saturation region in weak inversion operation.

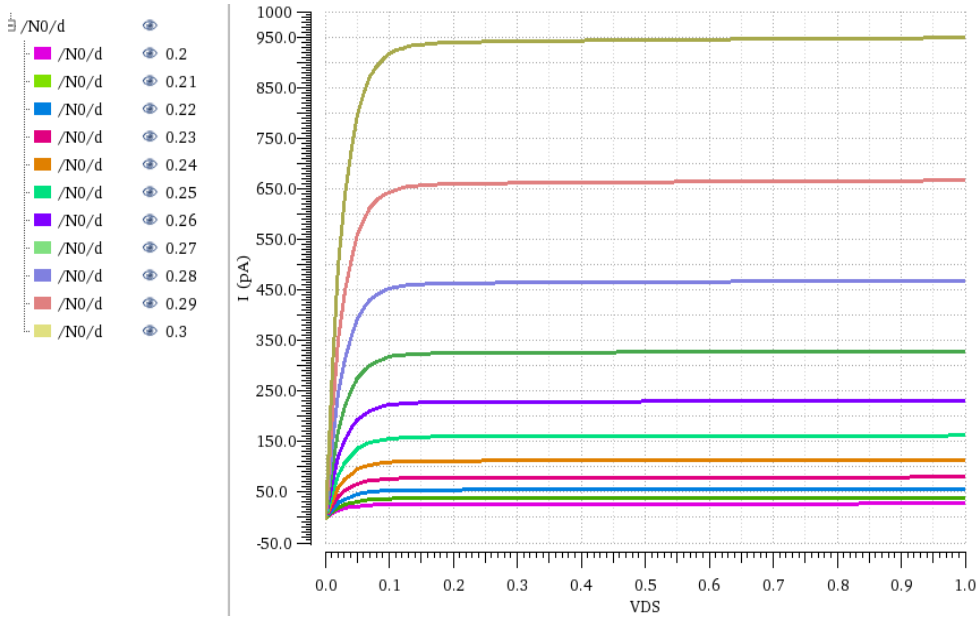


FIGURE 69 DC ANALYSIS OF A W=300NM L = 500NM FD-SOI TRANSISTOR

The subthreshold slope for this transistor is shown in Figure 70. The subthreshold slope represents the increase in gate voltage that is required, to increase the current flowing through the transistor channel by a factor 10. The transistor has a subthreshold slope that varies from 65 mV/Dec, to 70mV/Dec, which is near the ideal theoretical value of 60 mV/Dec. The subthreshold slope value allows us to have a large tuning range while maintaining the transistor in the week inversion regime. Since, the subthreshold slope is not constant; the effects of mismatch will vary as a function of the bias voltage.

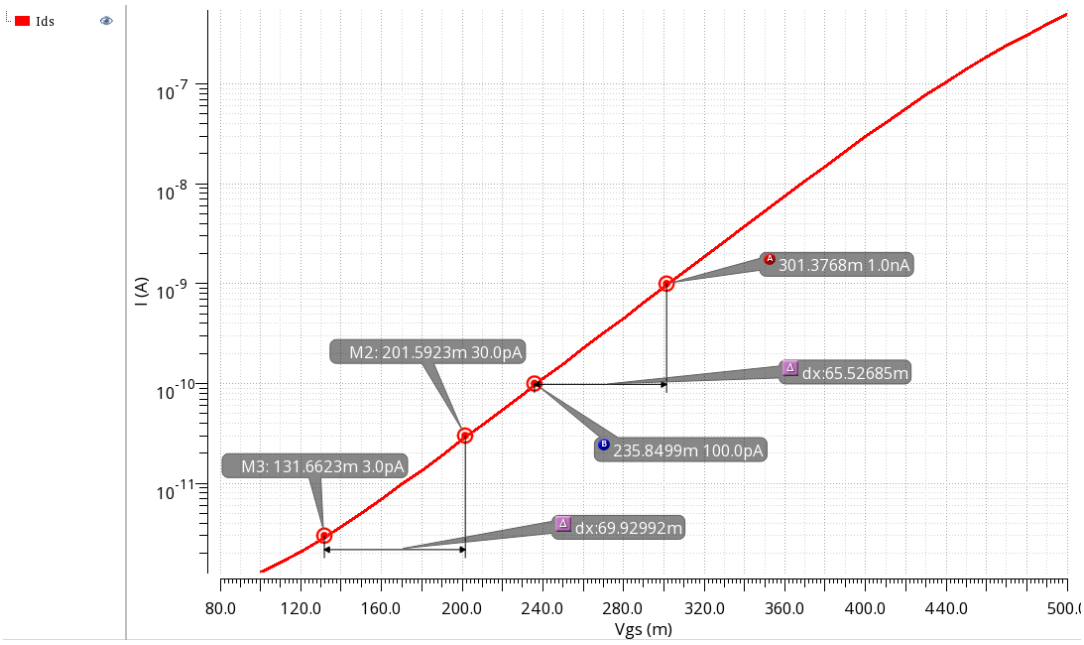


FIGURE 70 SUBTHRESHOLD SLOPE OF A W=300NM L=500NM TRANSISTOR

5.5.2 Analysis of capacitance charging

We started our analysis by simulating the charge of a capacitor using through our current source transistor. For this simulation, ideal switches replace the transistors that serve for control. The transistor was biased with a 250mV voltage at its gate, for an expected current of 161 pA. Several capacitance sizes were simulated. Results are shown in Figure 71.

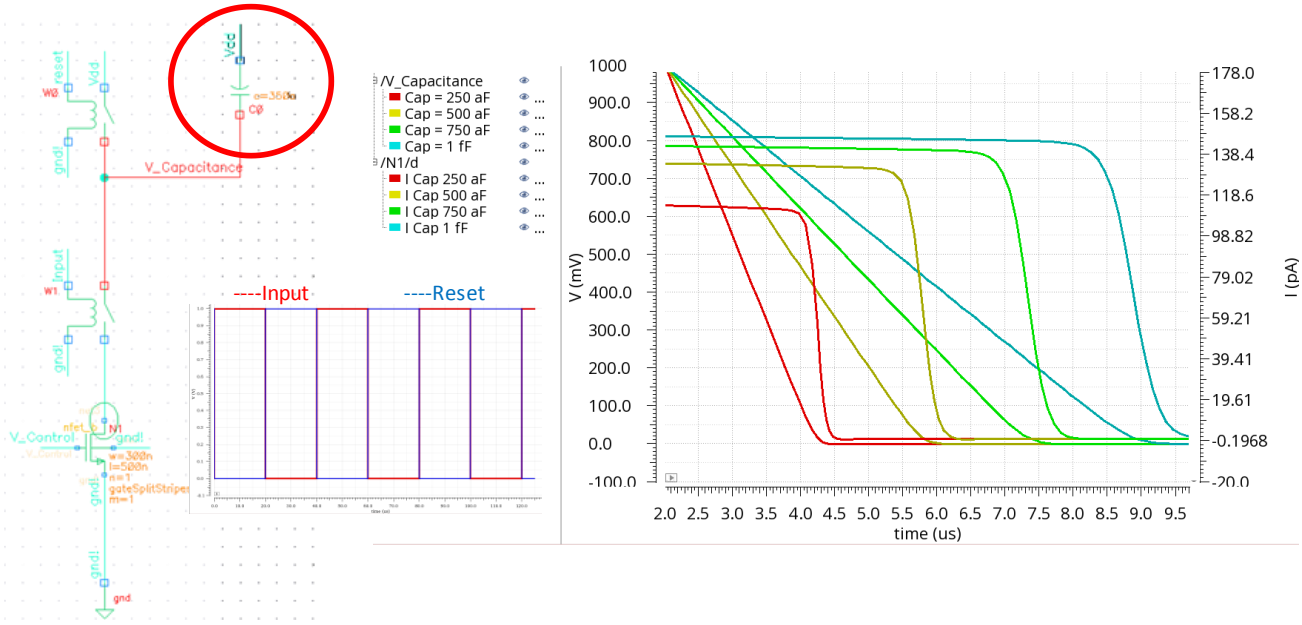


FIGURE 71 SLOW CHARGE CURRENT SOURCE VARIATIONS DUE TO CAPACITOR COUPLING

Due to a coupling between the capacitance and the parasitic capacitance of the transistor, the current flowing through the capacitance depends on the capacitance value (capacitive current division). To prevent undesired effects affecting the current source, we have opted to replace the switch with a cascode transistor. This consists of biasing the gate of the “switch” transistor with a voltage lower than VDD. The results are shown in Figure 72. Thanks to the cascode transistor the control current is constant for all cases.

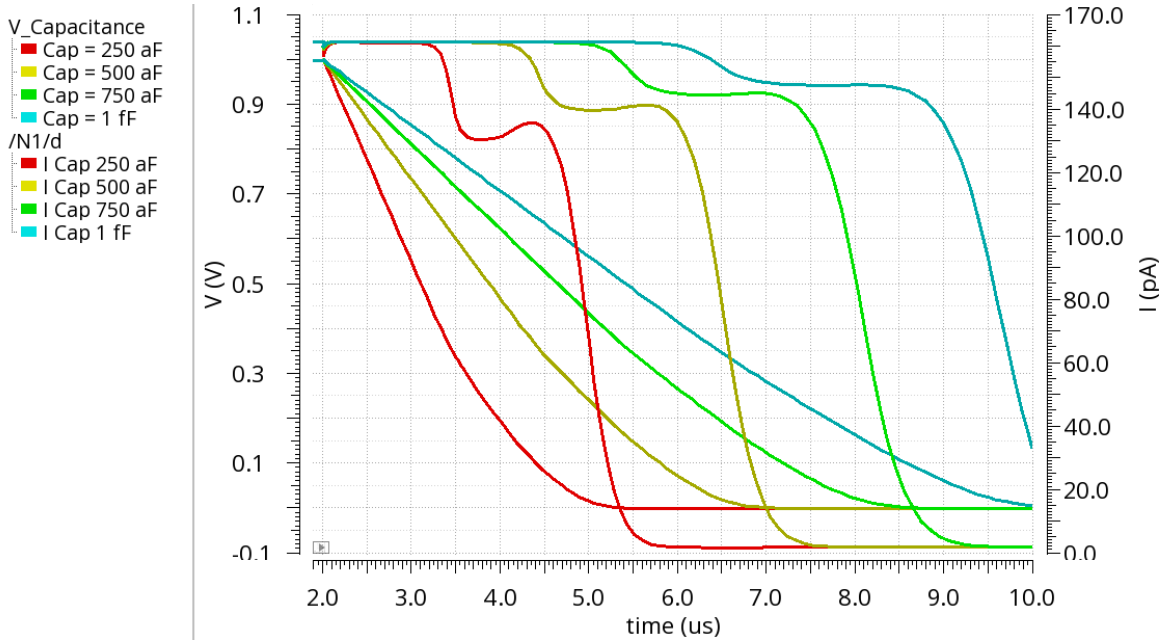


FIGURE 72 SLOW CURRENT SOURCE VALUE WITH A CASCODE PAIR

5.5.3 Complete delay element evaluation

Shown in Figure 73 is the digital delay element used for the top evaluation. In the following sections its performance will be analyzed, and the proposed modifications to improve its performance will be presented.

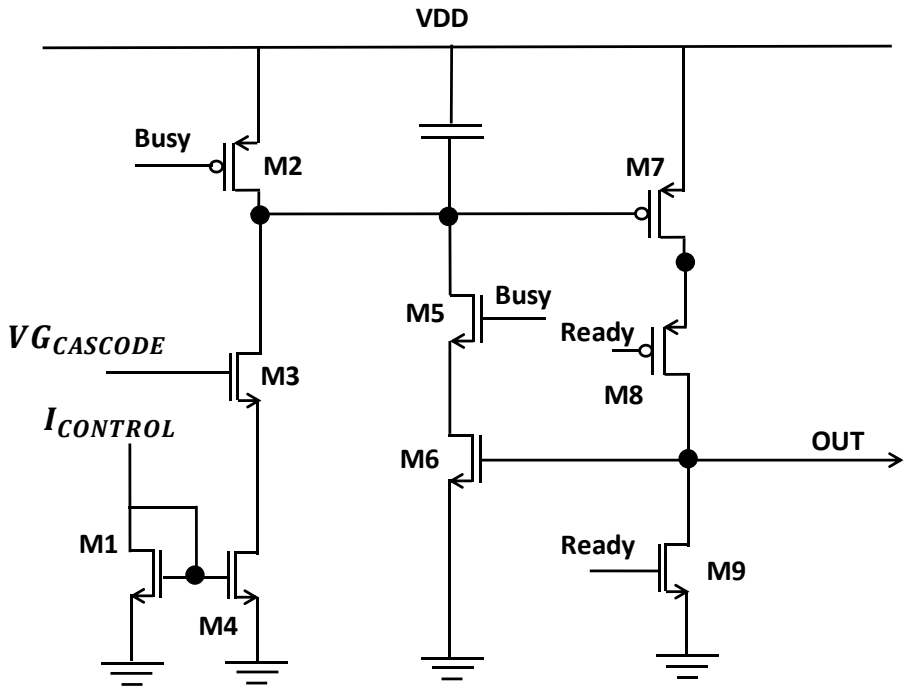


FIGURE 73 DELAY ELEMENT FIRST VERSION

The simulated delay time for different control currents, for the first version of the delay element is shown in Figure 74. We can highlight that the delay achieves a

maximum delay of approxiamtely 10µs, which is insufficient to met the longest delay value required of 11.24 µs presented in Table 9.

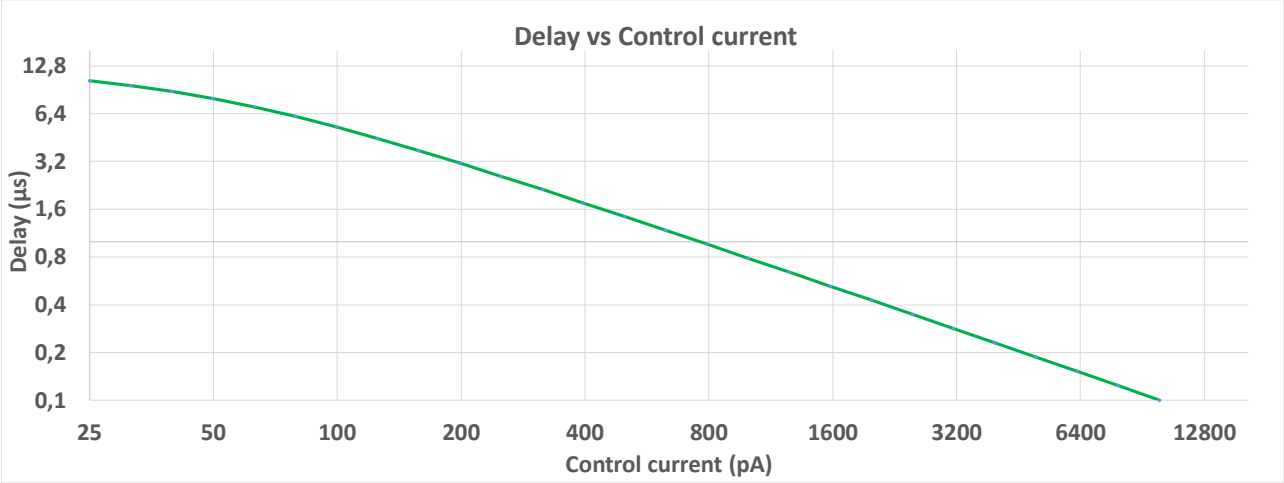


FIGURE 74 DELAY TIME VS CONTROL CURRENT FOR THE FIRST DELAY ELEMENT VERSION

5.5.4 Input signal dependency

The delay time should always be constant independently on the input signal. A simulation was performed to verify the input signal dependency. An initial pulse was presented at the delay, then after a time t_{period} a second pulse was presented at the input as shown in Figure 75. t_{period} is varied and the delay time for each pulse is quantified. The results are shown in Figure 76. The delay time is not constant for shortly spaced pulses; this difference decreases as the time between pulses is increased and disappears for pulses spaced at least 250 µs. This behavior leads us to the hypothesis that one of the circuit's nodes was not properly reset to its initial state after the completion of a delay operation.

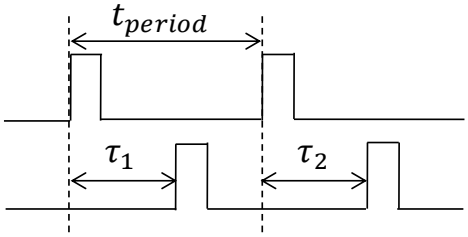


FIGURE 75 INPUT SIGNAL DEPENDENCY ANALYSIS

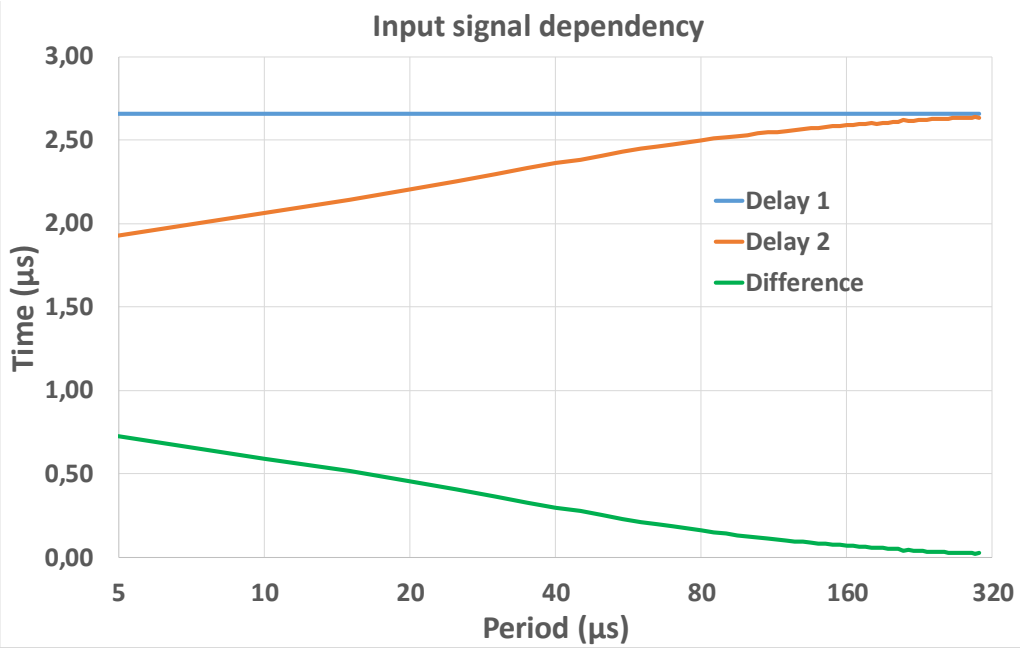


FIGURE 76 INPUT SIGNAL DEPENDENCY FOR THE FIRST DELAY ELEMENT VERSION

We analyze each node to verify our hypothesis, that a node was not properly reset. Indeed, the node that links transistors M7 and M8, stores charge as seen in Figure 77. When a new delay operation starts the stored charge goes directly to the output node, thus reducing the total delay time. However, when the pulses are sufficiently spaced on time, the charge is eliminated as leakage current trough M9.

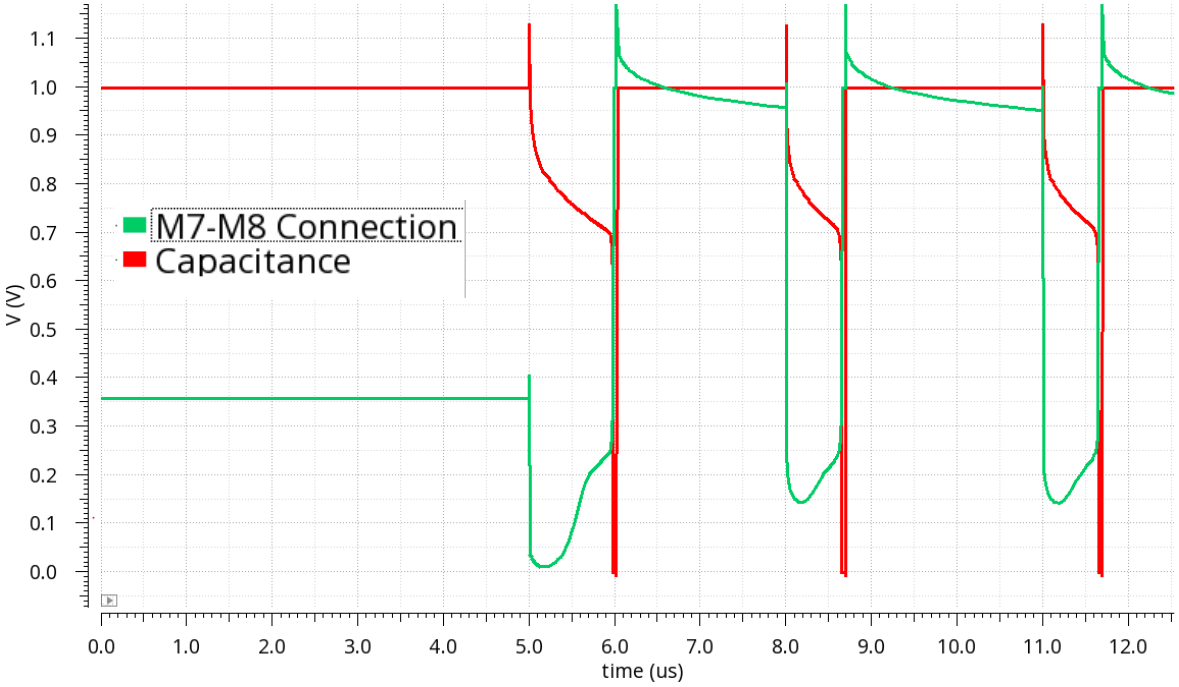


FIGURE 77 STORED CHARGE IN THE NODE THAT CONNECTS THE TRANSISTORS M7 AND M8

The addition of the NMOS transistor, to convert the PMOS switch into a MOS pass gate, allows reducing the input-signal dependency. However the node is still not

totally reset to its initial state as shown in Figure 78.

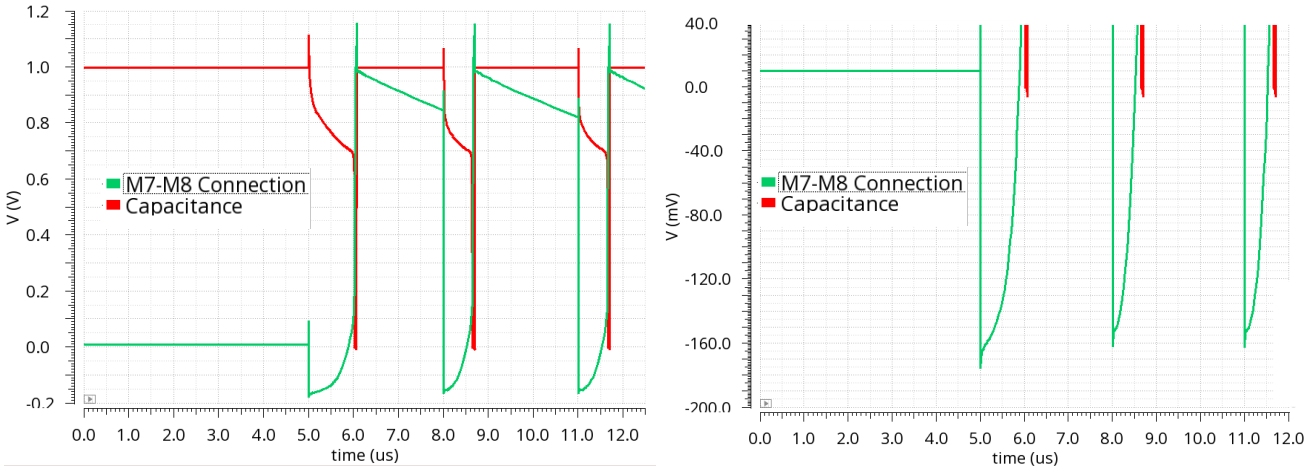


FIGURE 78 (A) EFFECT OF ADDING A NMOS TRANSISTOR IN PARALLEL TO M8 TO CREATE A MOS PASSGATE; (B) SAME FIGURE AS (A) WITH A DIFFERENT SCALE TO BETTER APPRECIATE THE VOLTAGE VARIATIONS

We have added an NMOS transistor used as a switch controlled by the Reset signal, in parallel to the MOS passgate to guarantee that the voltage on this node is set to zero, and that no remaining charge will affect the delay time as shown in Figure 79. A new simulation was run to quantify the input signal dependency after the modifications; results are shown in Figure 80. Input signal dependency has been eliminated.

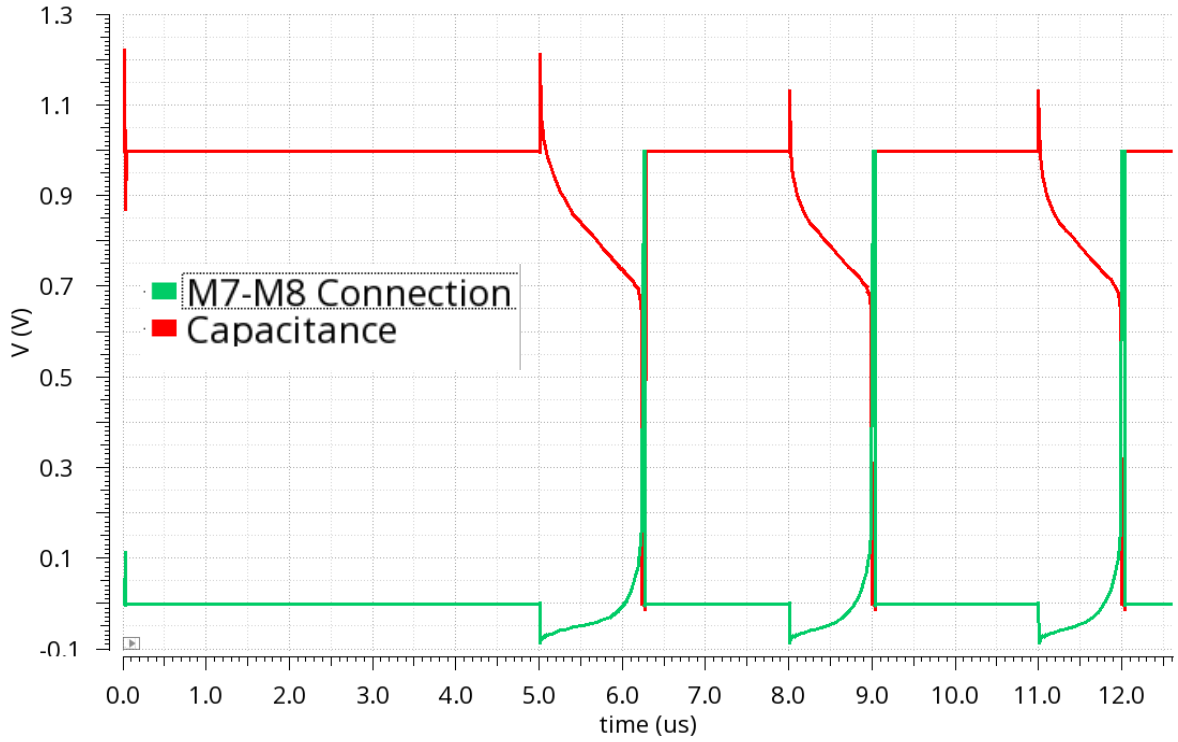


FIGURE 79 TENSION AT THE PMOS SWITCH NODE AFTER ADDING AN NMOS SWITCH TRANSISTOR

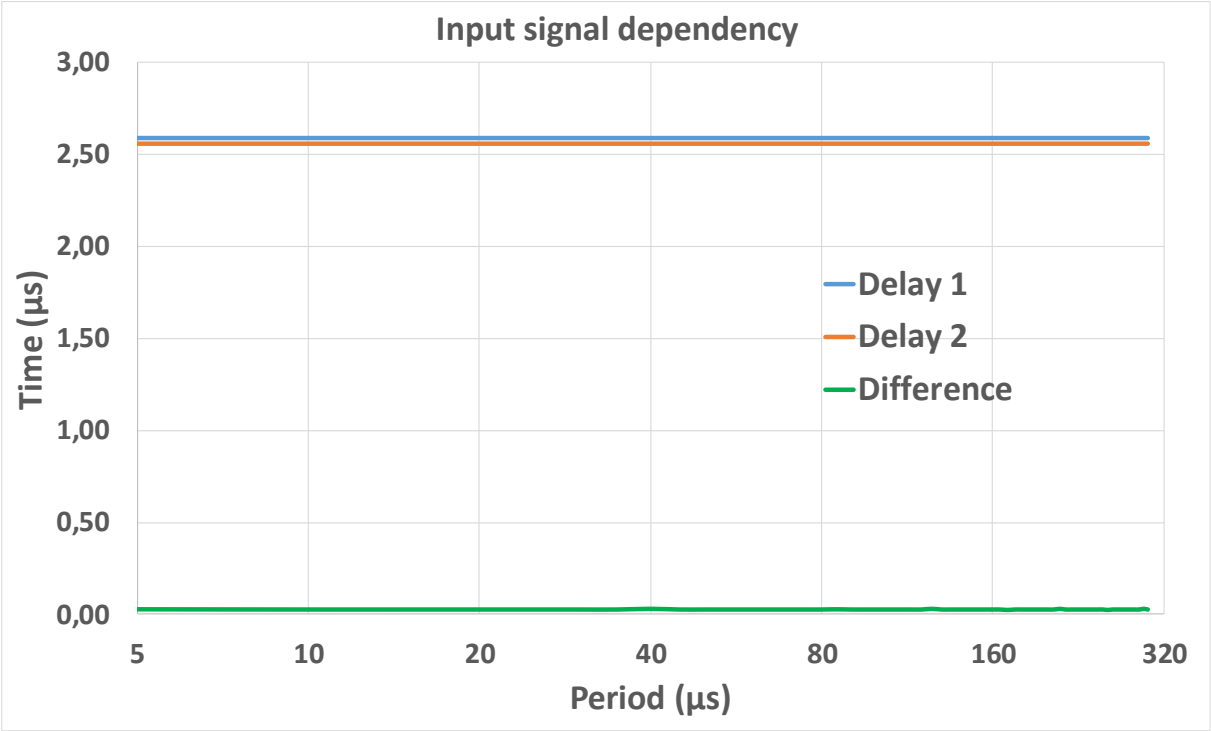


FIGURE 80 INPUT SIGNAL DEPENDENCY FOR THE SECOND DELAY ELEMENT VERSION

5.6 Elements for handshaking

As stated in the state-of-the-art analysis, a handshake mechanism is required, as well as proper signal synchronization.

Synchronization: Due to the asynchronous nature of signals, the UD signal must change at the falling edge of the Change pulse.

Handshaking: To avoid loss of information, a protocol is needed for the delay to keep the information available, until the next delay is capable to process it.

The protocol for handshaking used by [50] relies on a signal called acknowledge previous delay connecting two consecutive delays. This signal is at high level most of the time. This signal exhibits a negative pulse when two conditions are met.

- The previous delay has produced a rising edge on Change output
- The current delay is available to take new information.

The protocol is described in Figure 81. From this figure we can observe that the right timing to activate acknowledge is when input, reset and acknowledge from next delay are present simultaneously.

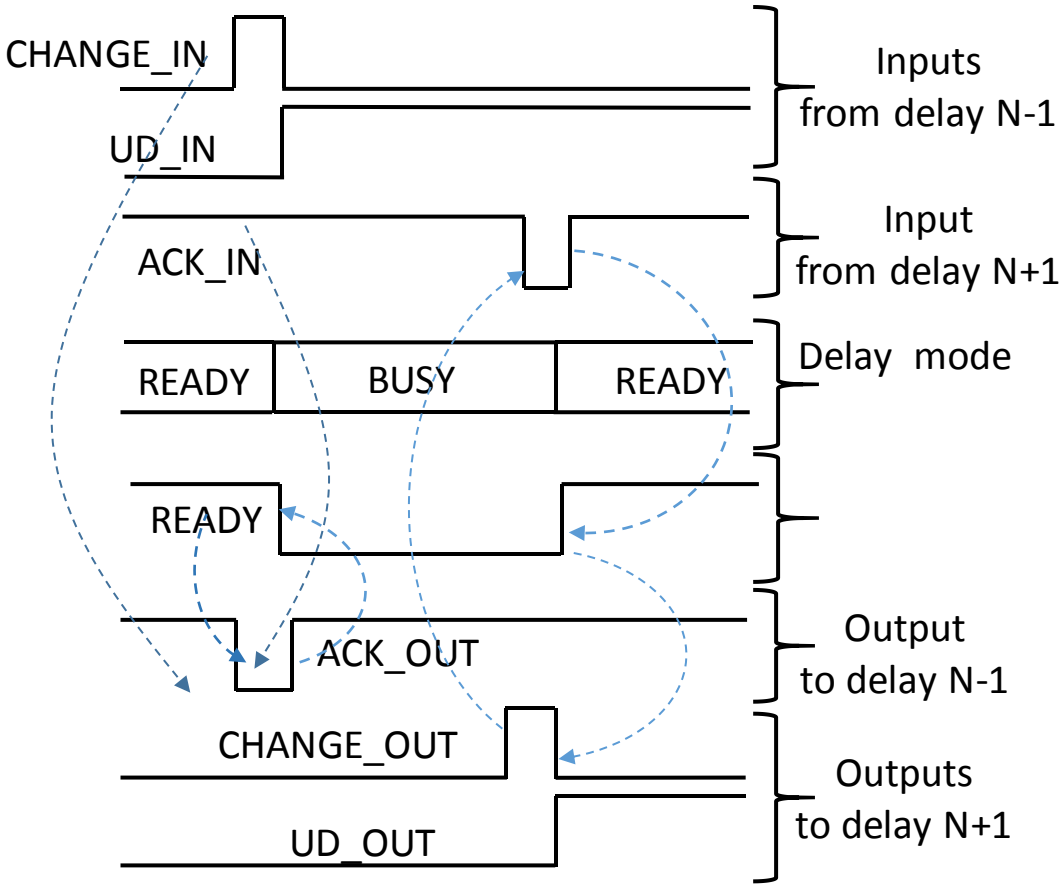


FIGURE 81 HANDSHAKING PROTOCOL

A simple implementation of the protocol using a three input NAND gate shown in Figure 82 is proposed by [50]. However, in this implementation a handshaking hazard may arise when the SR latch of delay N is faster than the SR of the delay N-1.

Figure 83 is an illustration of this situation. As we can see the acknowledge pulse (ACK_IN) that the N delay receives from the N+1 delay, is very short and doesn't give the N delay enough time to clear its output. As a consequence, the delay cell does not enter the "Ready" mode (Ready signal does not reach the expected high state value).

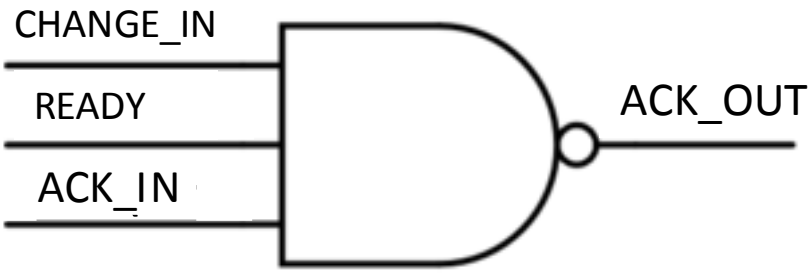


FIGURE 82 ACKNOWLEDGE SIGNAL GENERATION IN [50]

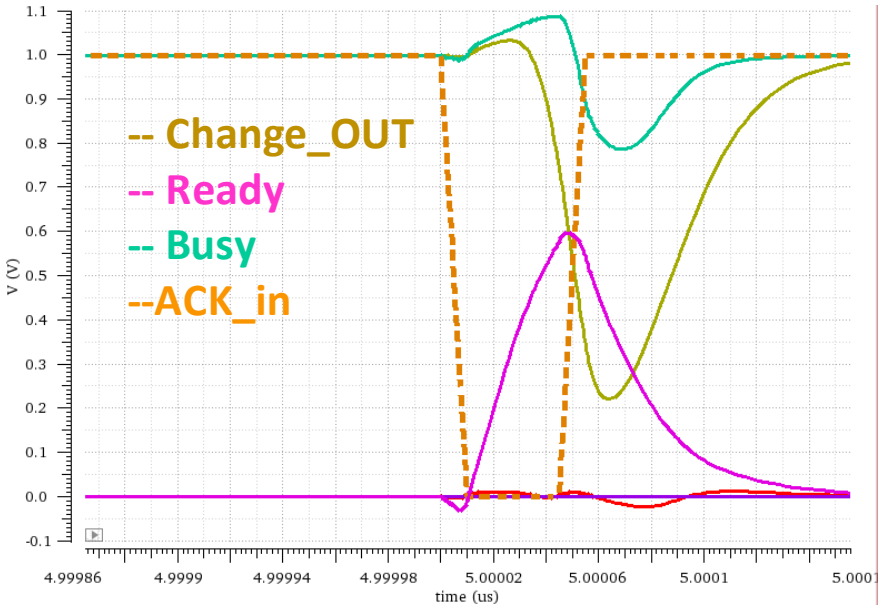
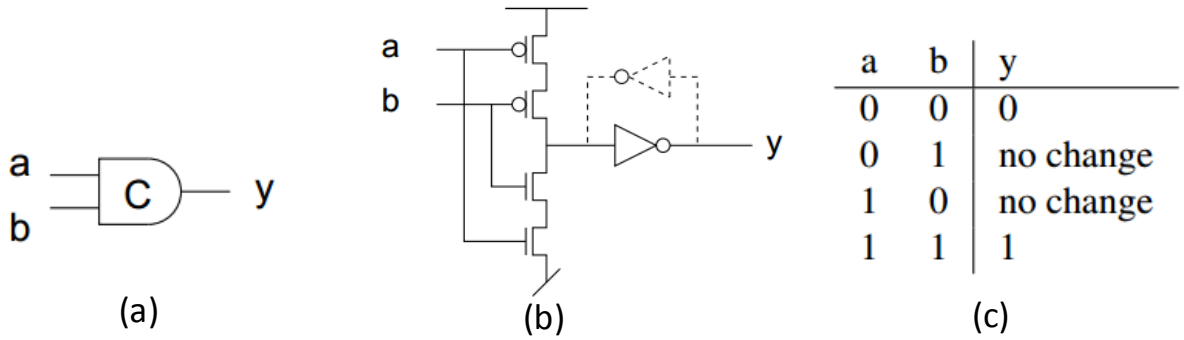


FIGURE 83 HANDSHAKING HAZARD

In [49] the handshaking issued is addressed by the use of a Muller gate also known as C-element. The Muller gate symbol, equivalent circuit and truth table are shown in Figure 84. The main trait of the Muller gate is that it waits for both inputs to be settle to the same value before changing its output.



a	b	y
0	0	0
0	1	no change
1	0	no change
1	1	1

(c)

FIGURE 84 CHARACTERISTICS OF A MULLER GATE. (A) SYMBOL, (B) COMMON IMPLEMENTATION (C) TRUTH TABLE [56]

We have decided to use the handshaking logic shown in Figure 85.Compared to the implementation presented in [49], we added a NMOS transistor (M15), to set the delay in Ready mode at circuit startup. Dimensions are W=500nm L=80nm. This transistor is particularly important because Monte-Carlo simulations reveal that in some cases due to transistor mismatch, the circuit is not properly set to the initial state. Thus, a relatively large transistor with a lower variation of its voltage threshold was chosen.

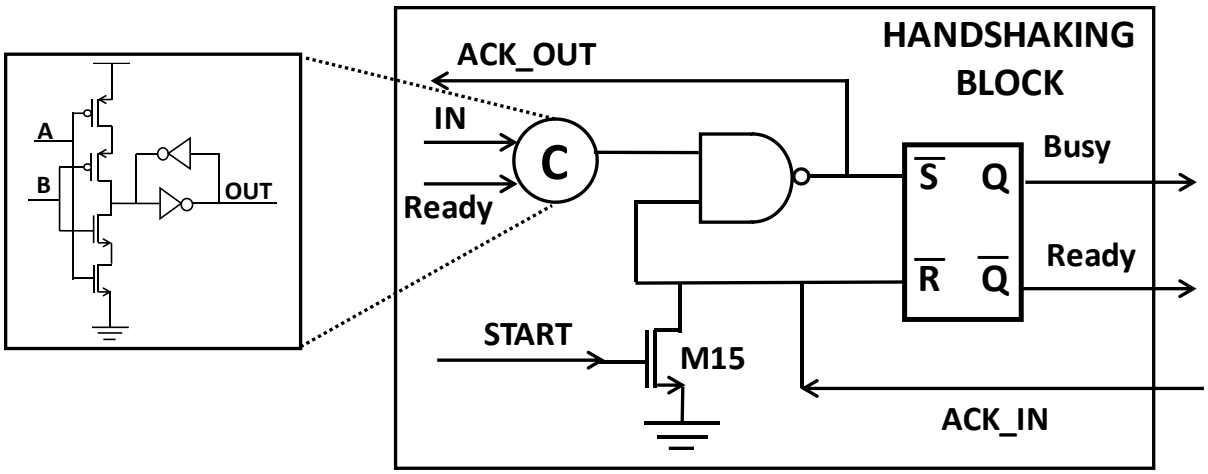


FIGURE 85 HAZARD FREE HANDSHAKING CIRCUIT

The robustness of the solution is demonstrated in Figure 86. We can see that the acknowledge pulse generated by the N+1 delay is not pulled up, until the N delay has changed to an idle state. The Muller-gate assures that the delay cell is set to the "Ready" state before accepting a new input pulse, and maintains the ACK_OUT signal low until the output of the previous delay cell is cleared.

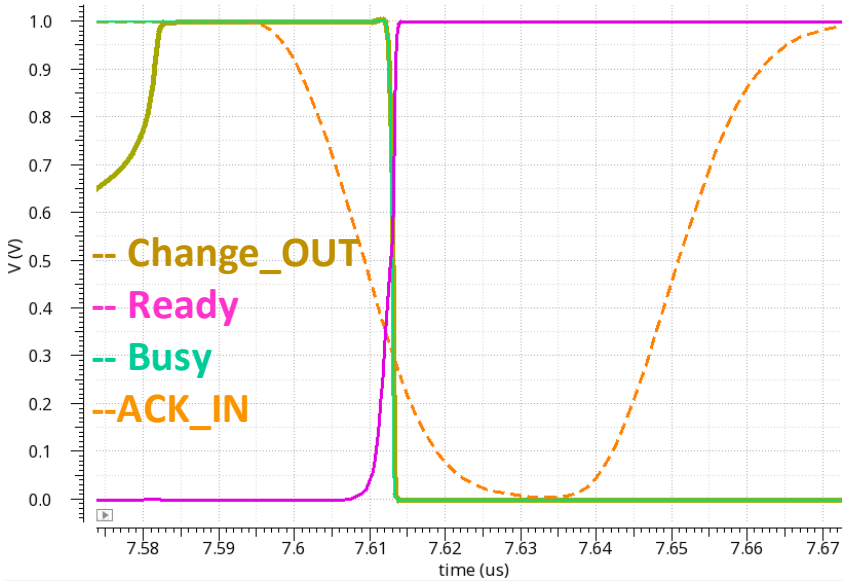


FIGURE 86 SIMULATION DEMONSTRATING THE FUNCTIONNING OF THE HANDSHAKE ELEMENTS

5.7 Reducing power consumption

Simulations were made to quantify the energy consumption of the delay element. Figure 87 shows the graphs that decompose the contribution of each element to the overall power consumption of the delay, both for static and dynamic power. The logic gates required for handshaking between delays contribute 60% to the dynamic power consumption.

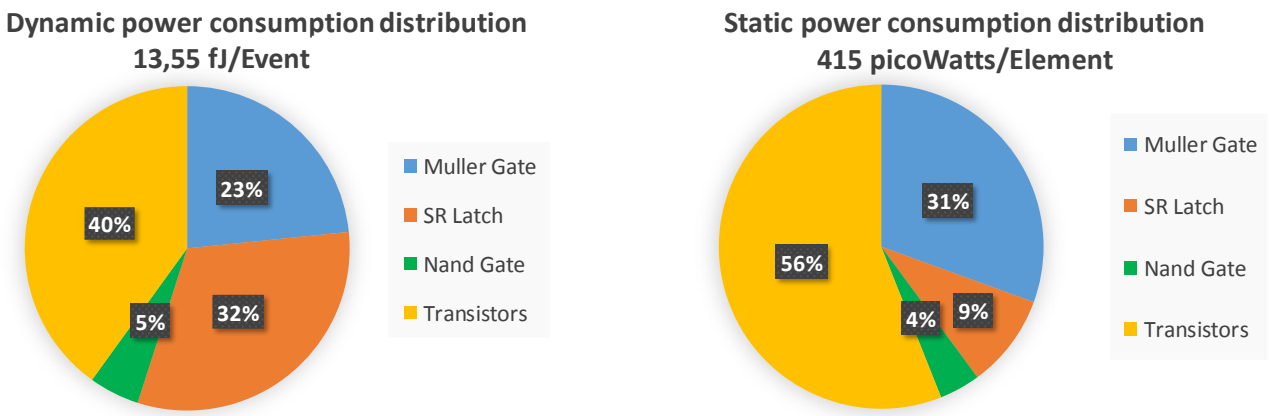


FIGURE 87 DYNAMIC AND STATIC POWER CONSUMPTION OF THE DELAY ELEMENT

As a reminder, our target is a dynamic consumption of 20fJ/Event and a static consumption of 40 pW/Element. We obviously have to improve static consumption.

5.7.1 EGVLVT transistors

A DC analysis of the delay element showed that leakage currents were contributing significantly to the static power consumption. In advanced technology nodes with reduced transistor gate oxide thickness significant gate leakage currents occur due to a phenomenon called thin-oxide gate tunneling, particularly for NMOS transistors. Due to thin-oxide gate tunneling RVT transistors have a gate leakage current value of 1.5 nA per µm². The only way to reduce gate leakage (on top of reducing the gate size, which was not possible because transistors in the digital part are close to minimum size) is to increase the gate oxide thickness. In the 28 FD-SOI technology, a second oxide is available (to implement 1.8V transistors), devices are EG transistors, with a penalty of increased Vt and increased minimum length (150 nm) which would impact significantly power consumption.

An intermediate transistor version is also available in the 28 FD-SOI technology, these devices are named EGV transistors. Although EGV transistors have the same gate oxide thickness than EG transistors, they operate at a lower voltage (1.5V) and have a lower minimum length (100nm). Additionally, there are two versions of the EGV transistor, an RVT with a voltage threshold around 700mV and a LVT transistor with a voltage threshold around 350mV, similar to the RVT transistors used in the first version of the delay. Thus, we have replaced wherever possible the RVT transistors with EGV LVT transistors, keeping the same W/L ratio. The expected impact is a gate leakage reduction without a significant impact on functionality.

The Muller gate and the NAND gates that were previously coming from the digital standard cell libraries, will have to be custom made using EGV LVT transistors.

To simplify the speech, standard gate transistors are called GO1 and the thick oxide transistors are called GO2.

On top of these modifications, a detailed analysis of the current consumption leads us to schematic modifications described below.

5.7.2 Inverter to control the cascode source

To avoid current flow in the current source in Ready mode, it has been decided to close the cascode transistor, with control of the cascode gate voltage. This is implemented using an inverter supplied by the cascode voltage and controlled by the Ready signal as shown in Figure 88. GO1 transistors with a minimum gate length have been used to limit the impact of overshoot voltage due to charge sharing during transitions.

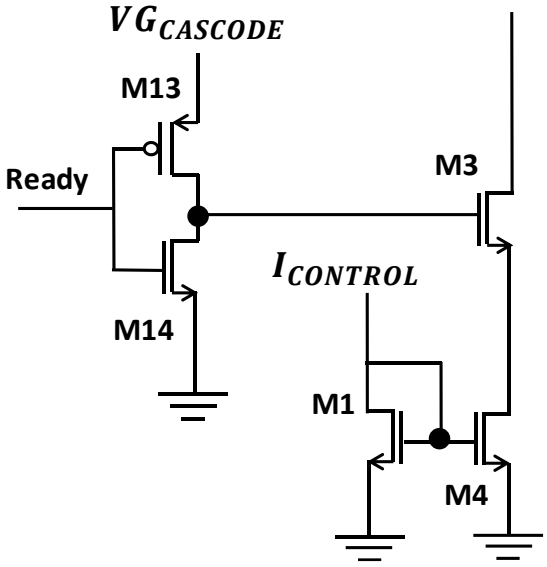


FIGURE 88 INVERTER USED TO CONTROL THE CASCODE TRANSISTOR

5.8 Other modifications

5.8.1 EGVLVT Transistors as a transmission gate

Another change implemented for the second version of the digital delay element is shown in Figure 89. A MOS pass gate switch constituted by EGVLVT transistors has replaced the M2 PMOS transistor, used as a switch to control the charge and discharge of the capacitance. The two main benefits are first the reduction of the parasitic current that limits the controllability of the delay element thanks to the replacement of the GO1 transistor. The second advantage is a significant decrease of the voltage overshoot seen in Figure 90. In simulations the parasitic current of the switch is now 1.2 pA allowing to operate with 25pA current source, while keeping leakage current below 5% of this value. Simultaneously the voltage overshoot has been reduced from approximately 300mV to less than 50mV as shown in Figure 91.

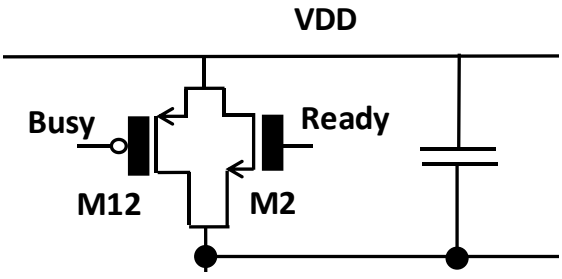


FIGURE 89 EGVLVT MOS PASSGATE TO CONTROL THE DISCHARGE CAPACITANCE

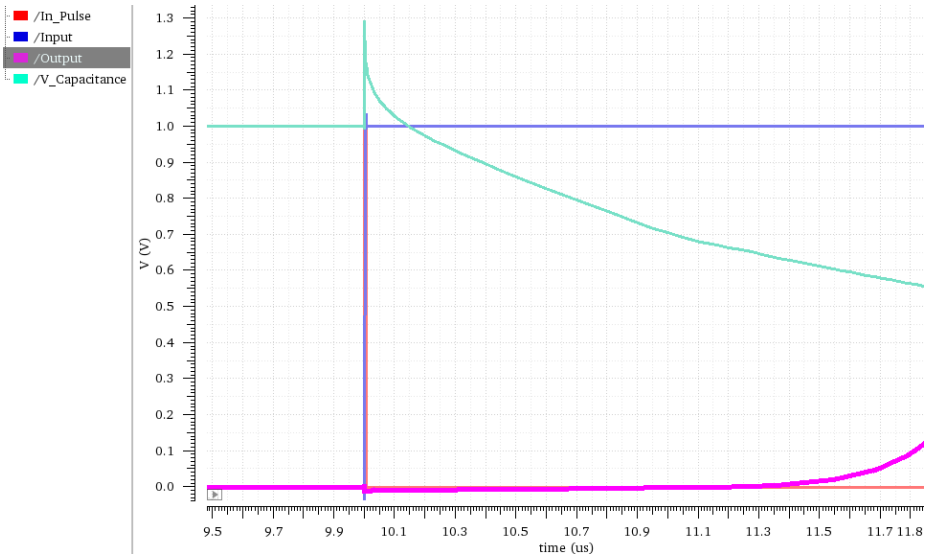


FIGURE 90 OVERSHOOT VOLTAGE BEFORE THE IMPLEMENTATION OF THE PMOS SWITCH

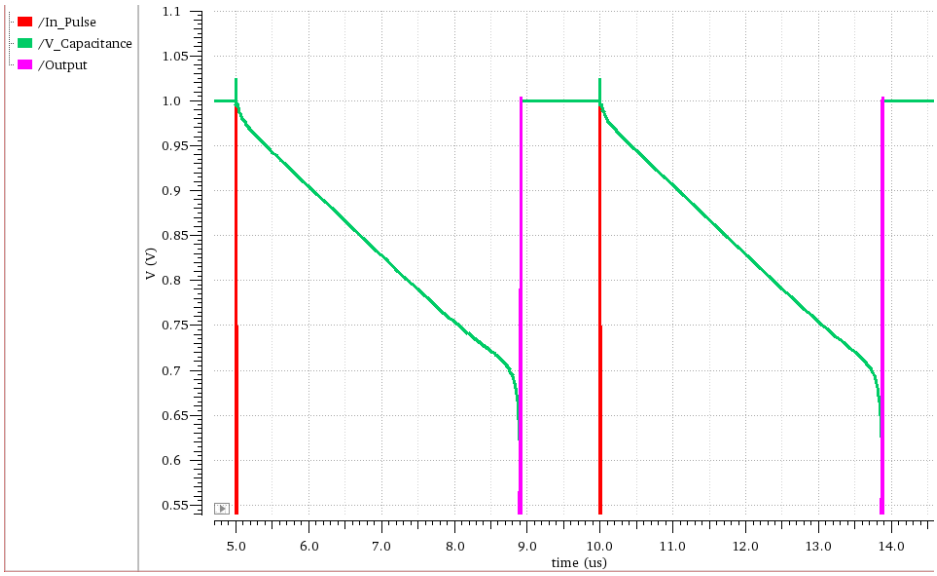


FIGURE 91 REDUCTION ON THE VOLTAGE OVERSHOOT DUE TO THE USE OF AN EGVLVT CMOS SWITCH

5.8.2 PMOS switch to control the fast discharge source

Another modification made to the circuit was to replace the M5 NMOS used as a switch to control the fast charge current source with an RVT PMOS transistor as shown in Figure 92. In the previous implementation there was a full voltage swing VDD-GND across the capacitor plates. This wastes energy, because the capacitor continued to be charged after the activation of the thyristor comparator. Thanks to the PMOS switch, the swing across the capacitor is limited from VDD to the threshold voltage of the PMOS transistor as shown in Figure 93. This PMOS transistor has been sized to have a low on-resistance, thus W= 400nm and L= 50 nm.

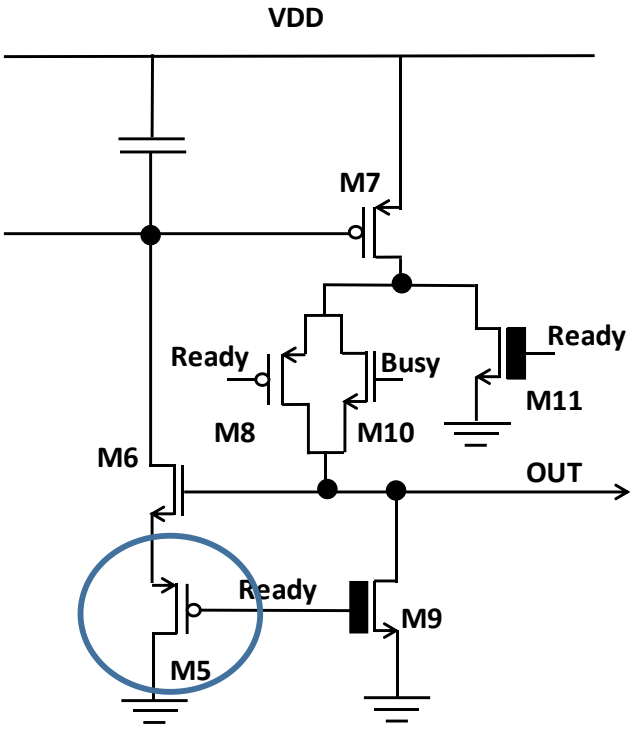


FIGURE 92 PMOS TRANSISTOR USED TO CONTROL THE FAST DISCHARGE SOURCE

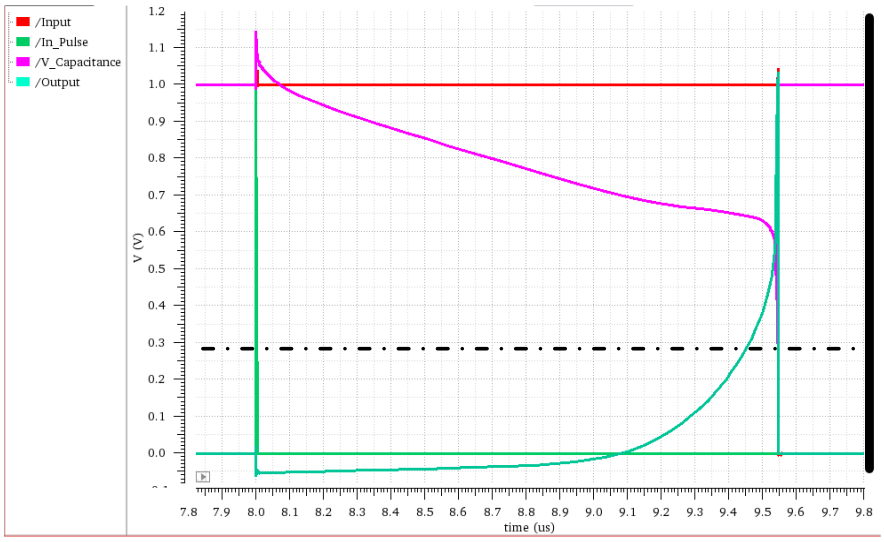


FIGURE 93 DELAY OPERATION WITH THE PMOS SWITCH CONTROLLING THE FAST CHARGE SOURCE

5.9 Delay final version schematic and performance

The final version of the delay-element that will be implemented into a layout is shown in Figure 94. The characteristics of each device are shown in Table 14.

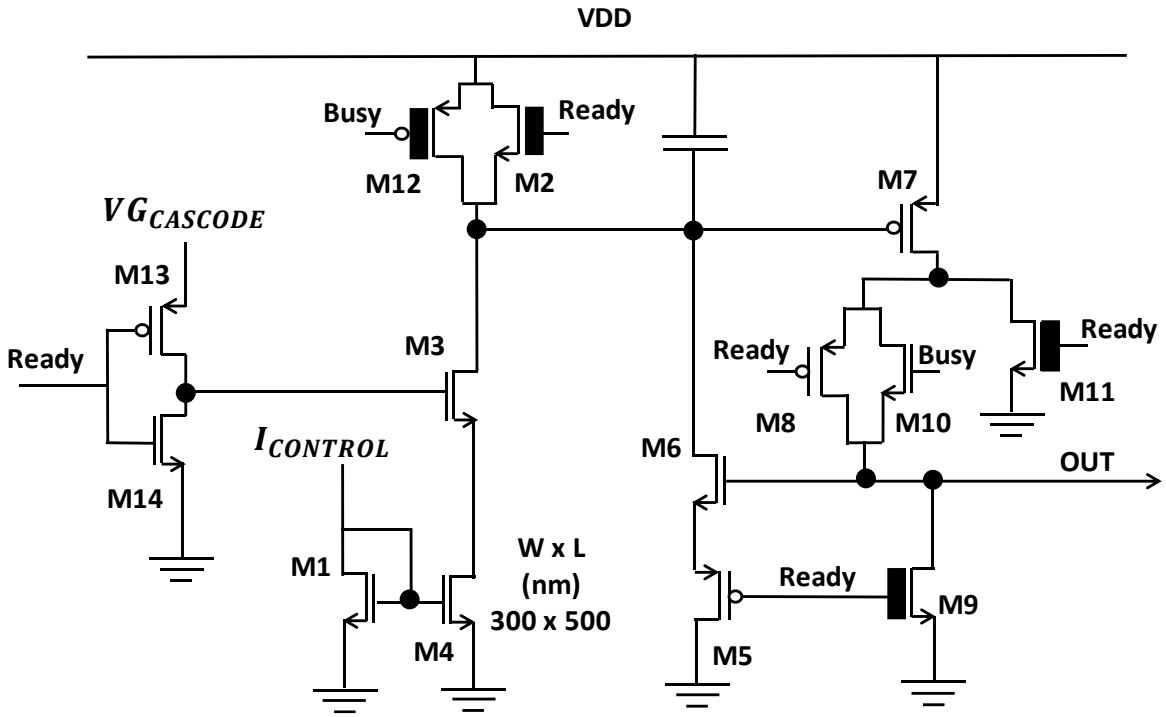


FIGURE 94 FINAL DELAY VERSION (HANDSHAKE ELEMENTS NOT INCLUDED)

TABLE 14 DEVICE SIZES OF THE DELAY SHOWN IN FIGURE 94

Device	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
W(nm)	200	300	300	400	300	300	300	160	300	160	200	164	108
L(nm)	80	100	500	50	40	100	40	400	40	100	100	30	30
Oxide	2	1	1	1	1	1	1	2	1	2	2	1	1

5.9.1 Power consumption

Table 15 shows a comparison between the static and the dynamic power consumption of the delay blocks in the final version and the first version. The static power consumption of the cell is reduced by a factor 20 with a minor impact on the dynamic power consumption.

TABLE 15 POWER CONSUMPTION COMPARISON BETWEEN THE FIRST AND FINAL VERSIONS OF THE DELAY ELEMENT

	Power	Muller Gate	SR Latch	Nand Gate	Transistors	Total
RVT	**Static**	127 pW	38,8 pW	16,9 pW	233 pW	**415 pW**
	Dynamic	3,17 fJ/Ev	4,27 fJ	0,68 fJ	5,42 fJ	**13,5 fJ**
EGVLVT	**Static**	2,30 pW	2,81 pW	2,80 pW	12,8 pW	**20,7 pW**
	Dynamic	4,53 fJ	4,97 fJ	0,76 fJ	5,15 fJ	**15,4 fJ**

5.9.2 Mismatch

Although this section appears late in the design description, mismatch aspects have been evaluated earlier when sizing the current source. Purpose is now to determine the contribution of each transistor, especially the ones forming the thyristor and to confirm initial choice of current source sizing.

A Monte-Carlo simulation with 1000 runs, was conducted for 4 different delay values. Simultaneously input-signal dependency is quantified using two 20µs spaced input pulses. For each run, the delay time for pulse 1, pulse 2, and the difference between them was quantified. The resulting distributions are shown in Figure 95. The results are compiled in Table 16.

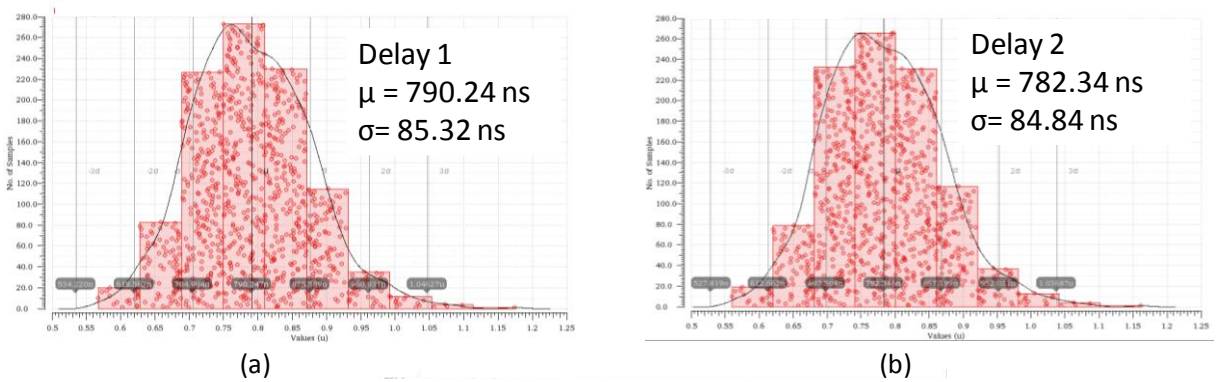


FIGURE 95 MONTE-CARLO SIMULATION RUN (A) DELAY TIME FOR THE FIRST PULSE, (B) DELAY TIME FOR THE SECOND PULSE

Control current	Delay 1	Delay 2	Mismatch 1 σ/µ	Mismatch 2 σ/µ	Difference **(Delay 1 – Delay 2)/Delay 1**
25 pA	15.51 µs	15.39 µs	9%	9.1%	0.77%
125 pA	4.74 µs	4.70 µs	10.5%	10.5%	0.84%
250 pA	2.64 µs	2.62 µs	10.8%	10.7%	0.75%
1 nA	0.79 µs	0.78 µs	10.8%	10.7%	1.26%

TABLE 16 MISMATCH RESULTS OF THE DELAY ELEMENT

The specification of σ/µ < 30% is respected for all delay values with a significant margin. In a future version the transistors size could be reduced to improve power consumption, while respecting the mismatch constraint.

We can observe that σ/µ increases with current and reaches a maximum value. This is probably linked to the subthreshold slope of the current source, moving from 73mV/Dec at 25pA to 65mV/Dec in the 100pA – 1nA range.

Looking at the difference between delay 1 and delay 2, we see that it is always below 1.5% of the delay value and will not disturb the overall filter operation. In most cases delay 1 is larger than delay 2. This can be explained by the handshaking mechanism arbitrations.

5.10 Delay taps layout implementation

In this stage we start the layout implementation of the delay cell. The first step is to consider the floorplan of the complete delay tap, to determine the constraints of the delay layout.

Care must be taken to propose an implementation that will not compromise the delay performance, in particular mismatch and power consumption.

The most straightforward implementation for a delay chain to optimize signal propagation distance would be a straight line. This would be optimum for the dynamic consumption (less parasitic capacitance in the connections), but the first delay will be far away from the last one, compromising the mismatch, additionally the shape factor wouldn't be optimal (a very long and narrow cell). Thus, a ladder-snail structure has been proposed as a tradeoff between consumption and mismatch. The ladder-snail structure is shown in Figure 96.

The connection strategy is as follows:

All propagating signals (Change, UD, and ACK) are propagated horizontally, together with the two bias voltages (cascode voltage and reference voltage).

All signals identical to every cell (GND, VDD, and Start) are propagated vertically.

To avoid overlap of interconnections from one line to next one, the cells of two consecutive lines are flipped across the x and the y-axis.

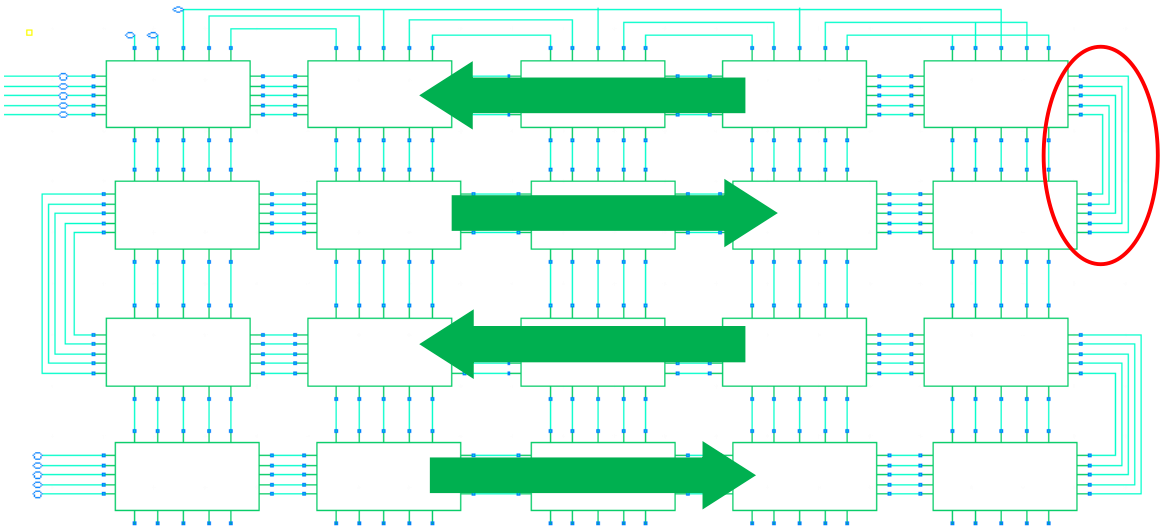


FIGURE 96 LADDER-SNAIL STRUCTURE, EMPLOYED FOR CASCADING GRANULAR DELAYS

For easy integration in a complete filter, it is convenient for the input and the output of the tap to be on the same side because both are connected to the same logical block. This constraint translates into having an even number of lines. Trying different values for the number of lines, while trying to keep close to a square aspect ratio, the chosen solution is 8*12. The number of granular delays per tap is thus

reduced from 100 to 96 resulting in a slight reduction of the maximum input slope that the system can process.

Our delay needs two biasing voltages namely cascode voltage and reference voltage, shown in Figure 94 that will be generated with current mirrors.

Since mismatch is negatively impacted by the distance between transistors, the distance between current sources and its associated current mirror must be small. There are two extreme implementation solutions. One mirror can bias all the delays or each delay can have its own mirror. The first solution is optimal for static consumption (one single reference current) but is the worst for mismatch (reference mirror cannot be close to all the delays). The second solution optimizes the matching between source and mirror but adds some drawbacks. It is needed to have one reference current per delay (96 per delay tap), bad for static consumption and increasing the complexity of routing, on top of that generating 96 currents will add another mismatch contribution. An intermediate solution has been chosen, which consists of dividing the tap into four sections, each one having a dedicated current mirror and reference current as shown in Figure 97. The current mirror is located at the center of the section, to have the smallest maximum distance possible.

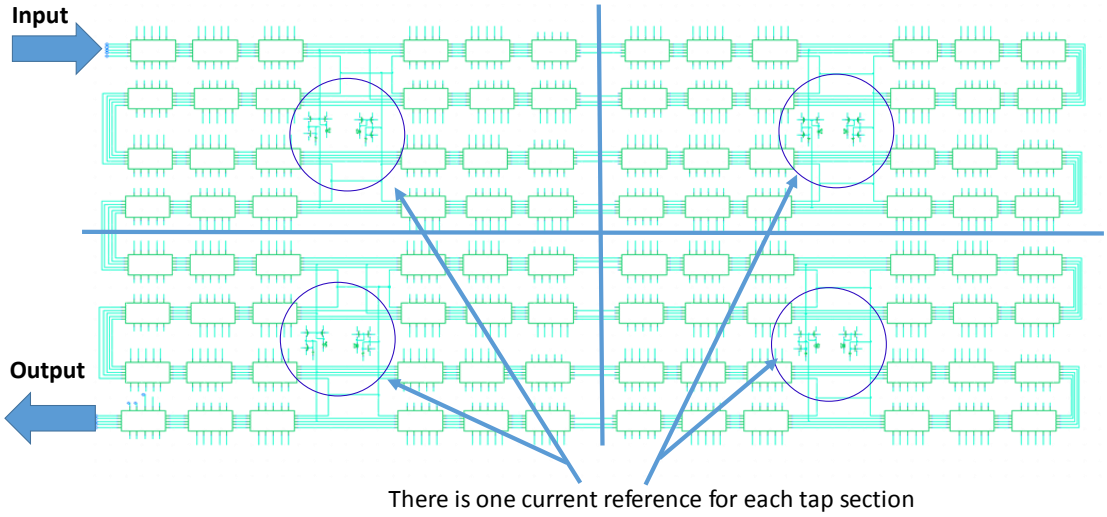


FIGURE 97 DELAY TAP IMPLEMENTATION

5.11 Single delay layout

Figure 98 shows the final layout of the delay cell. In order to implement the ladder-snail structures without any extra connection, the three vertically propagated signals (GND, VDD, Start) must be present both on the top and on the bottom (to propagate the signal). These three signals must as well be symmetric around the vertical center axis to be able to connect flipped cells.

The schematic is conform to Figure 94. With handshaking coming from Figure 85 and a D latch, similar to the one in signal CHANGE path, with UD_IN as input and UD_OUT as output. The latch is triggered by the “Ready” signal.

A custom-made D latch was created using the NAND gates previously created with GO2 transistors, to carry the UD signal.

As NAND gates and the C gate were custom-made, the layout has been optimized for this particular application. To speed up the layout, we made the decision to implement the capacitance using an available device from the design kit. Unfortunately, the smallest metal-on-metal capacitor, made with 6 metal layers, has a value of 4.5 fF, slightly larger than the one in the schematic (1.5fF). The impact of this change will be described later.

Since the design rules impose a minimum spacing between different oxide transistors, the total area of the circuit ends up being slightly larger than the original estimation at 75 µm².

In advanced CMOS technology nodes, it is mandatory to comply with density design rules, for any technological layer. As the automatic tiler, would have added poly and active dummies inside the cell, compromising mismatch, we added manually poly (red) and active silicon area (green) sections. To comply with density rules without needing the automatic tiler tool.

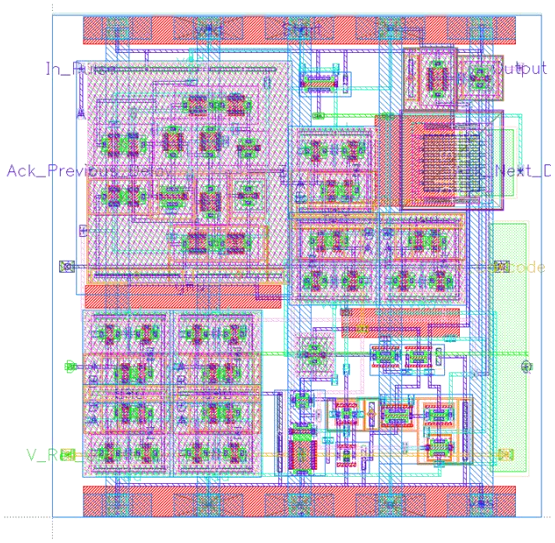


FIGURE 98 FINAL DELAY CELL LAYOUT MODIFY TO CORRECT DENSITY ERRORS

5.12 Post layout simulation

As capacitance value is key for the performance of our delay, it is mandatory to evaluate the impact of the layout parasitic capacitance. The impact of the integration capacitance value will be evaluated as shown in Figure 99.

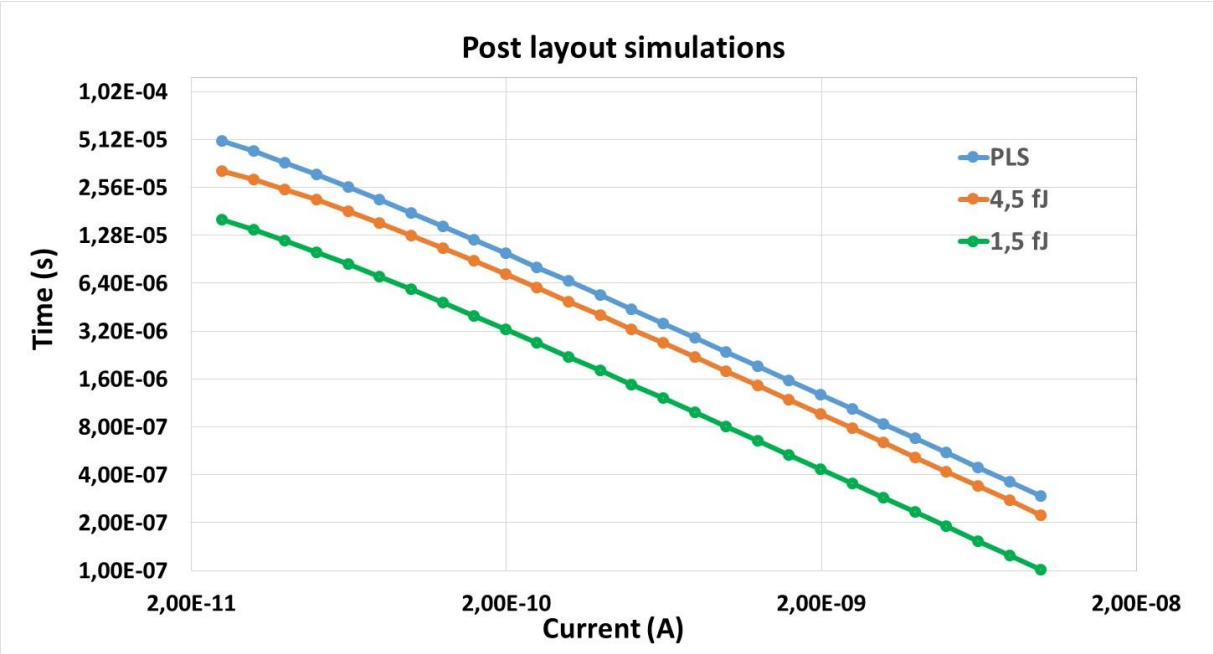


FIGURE 99 POST LAYOUT SIMULATIONS CONSIDERING ALL PARASITICS, THE IMPLEMENTED CAPACITANCE VALUE 4.5FJ AND THE ORIGINAL DESIGN WITH A 1.5FJ CAPACITANCE

The impact of the new capacitance value is an increase of delay for the same current. The impact of parasitic capacitances of the interconnections is an increase of approximately 50% in the delay time with respect to the time simulated when considering only the 4.5fJ capacitance.

5.13 Delay tap and test circuit layout

Figure 100 shows the layout implementation from the granular element, from a single element to a delay tap, to the complete structure composed by eight delay taps. One can recognize the ladder-snail structure.

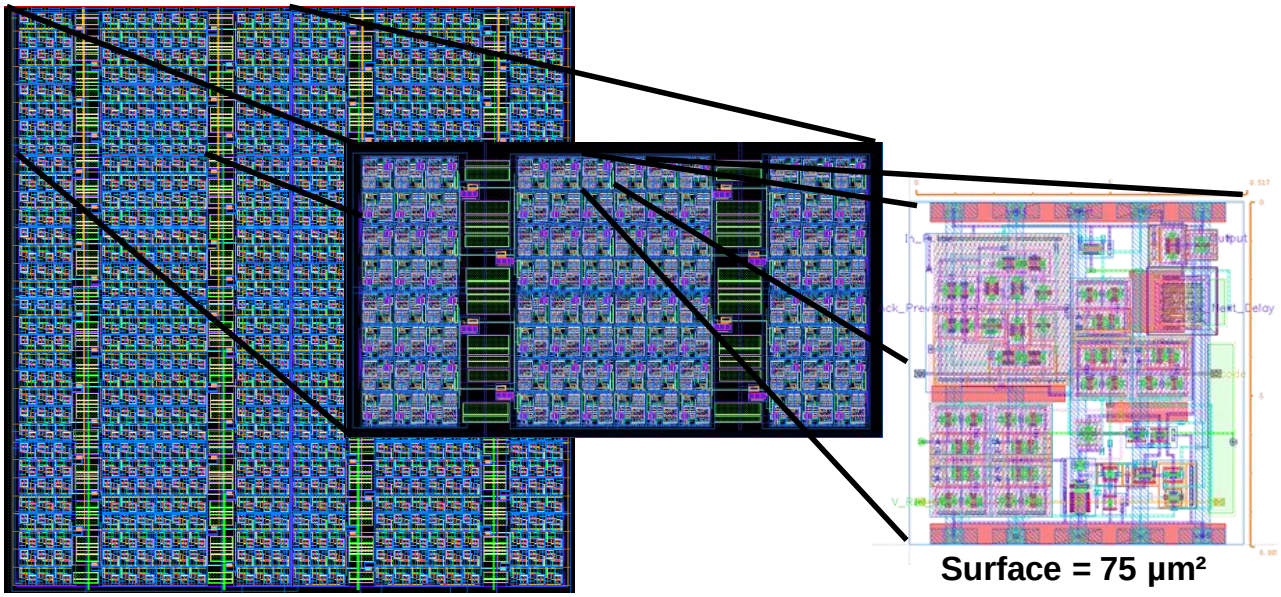


FIGURE 100 LAYOUT IMPLEMENTATION OF THE DELAY TAPS

Figure 101 shows the layout of the integrated circuit send for fabrication. .

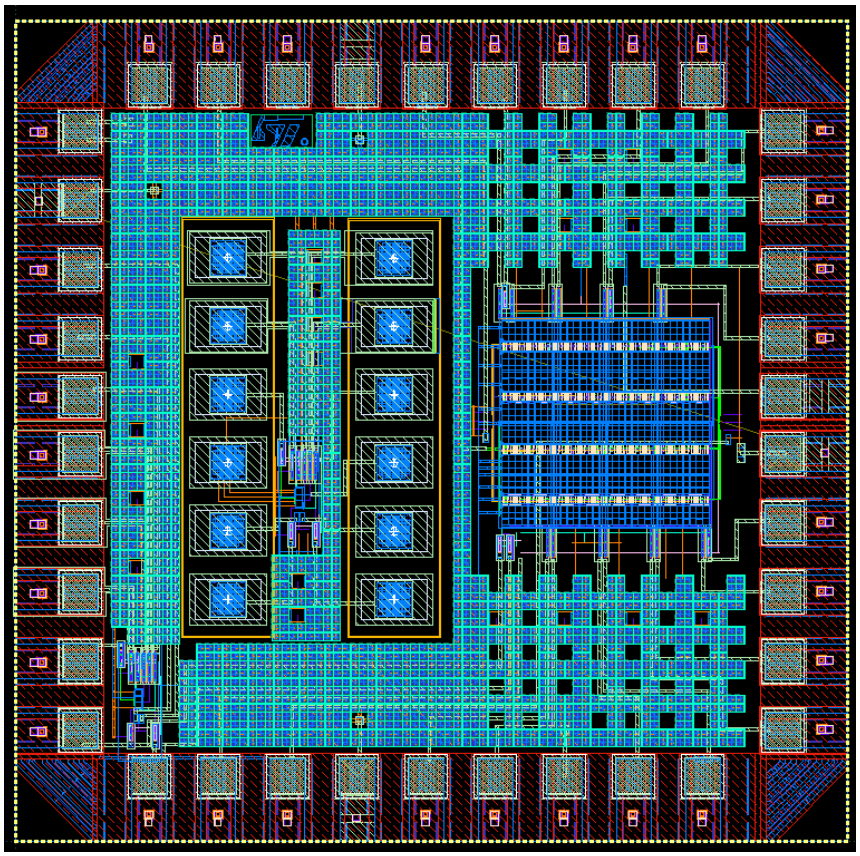


FIGURE 101 INTEGRATED CIRCUIT LAYOUT

The circuit has been drawn inside an electrostatic-sensitive device (ESD) ring providing protection for the circuit as well as capability to package the circuit using wire bonding techniques.

It features eight delay taps composed each by 96 granular delay elements, on the right half of the circuit.

To be able to evaluate the performance of a single delay, groups of three chained delay elements were created, with terminals on all signals. To be functional an inverter driven by the CHANGE_OUT of the third delay, was connected to ACK_IN of the same delay.

These groups have been repeated twice. A first occurrence on the bottom left corner connected to wire-bonded pads, to be measured inside a package. The second occurrence on the left half, connected to probing pads for on wafer measurements.

Small blocks have been added as well as:

- Output buffer banks enabling the circuit to drive measurement equipment (sized to drive 35pF). Please note that buffers are supplied with an independent supply pad, not to interfere with delay consumption measurements.
- Input buffer banks to additionally protect the circuit inputs.
- Biasing functions providing the biasing currents for all current mirrors, using a single reference current, coming from the external world.

Finally, the empty spaces have been filled with decoupling capacitances, to improve the robustness of the circuit to digital commutation.

5.14 Final performance assessment

A simulation of a complete delay tap can be seen in Figure 102 and Figure 103 . Figure 88 shows the propagation of the CHANGE signal and Figure 103 shows the propagation of the UD signal. The outputs of the two first delay tap elements and the last element are shown. A particular property due to the use of asynchronous logic is the waveform of the output signal of the delay elements. The handshake mechanism is triggered by the rising edge of the output signal, since we use asynchronous logic this process is done at maximum silicon speed, not limited by any clock. The acknowledge signal that clears the previous delay is so fast that the output of the delay does not have time to reach the VDD value. The output only has time to reach 0.5 V for a 1V VDD. Since the voltage dynmic of the output node is reduced the dynamic power consumption is also reduced. The output of the last element does have the time to reach VDD because is connected to an inverter that generates the acknowledge signal, which is not the case for other elements, as detailed in 5.13.

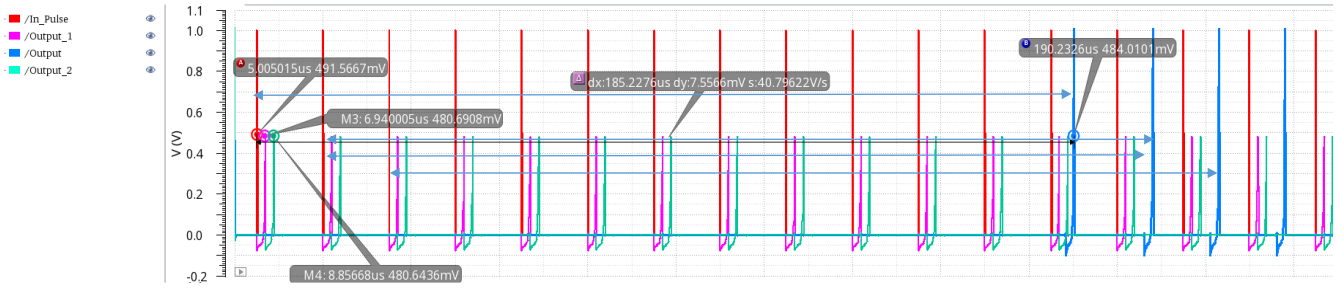


FIGURE 102 SIMULATION OF A COEFFICIENT DELAY TAP

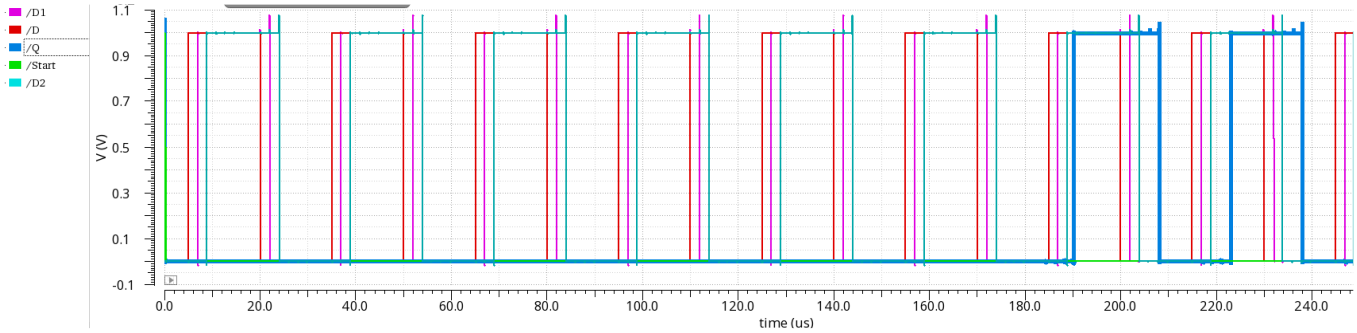


FIGURE 103 PROPAGATION OF THE BINARY INFORMATION THROUGH THE DELAY TAP

Figure 104 shows the simulation of the three-delay structure with output buffers. Since the capacitance of an oscilloscope probe is 20pF buffers have been designed to handle a capacitance around 35pF.

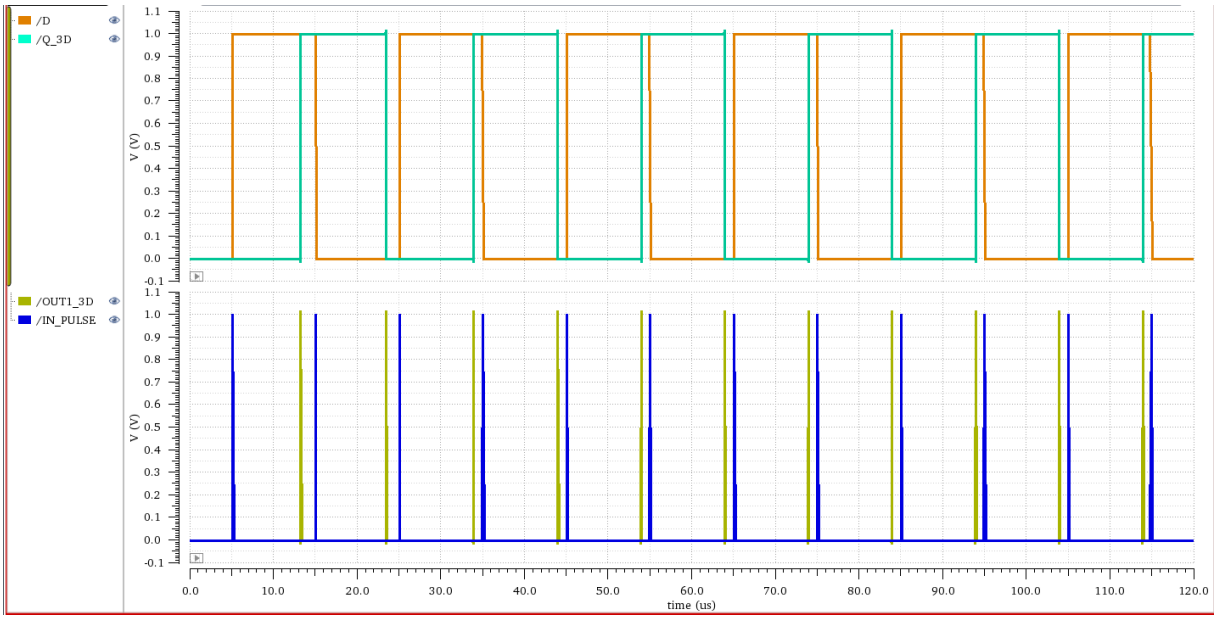


FIGURE 104 3 DELAY SIMULATION WITH OUTPUT BUFFERS

Figure 105 shows the simulation of the eight delay tap structure. Without any apparent reason some pulses present longer delays particularly for taps 6, 7 and 8, however this validates the robustness of the handshake mechanism, because the length of the associated UD value is also increased.

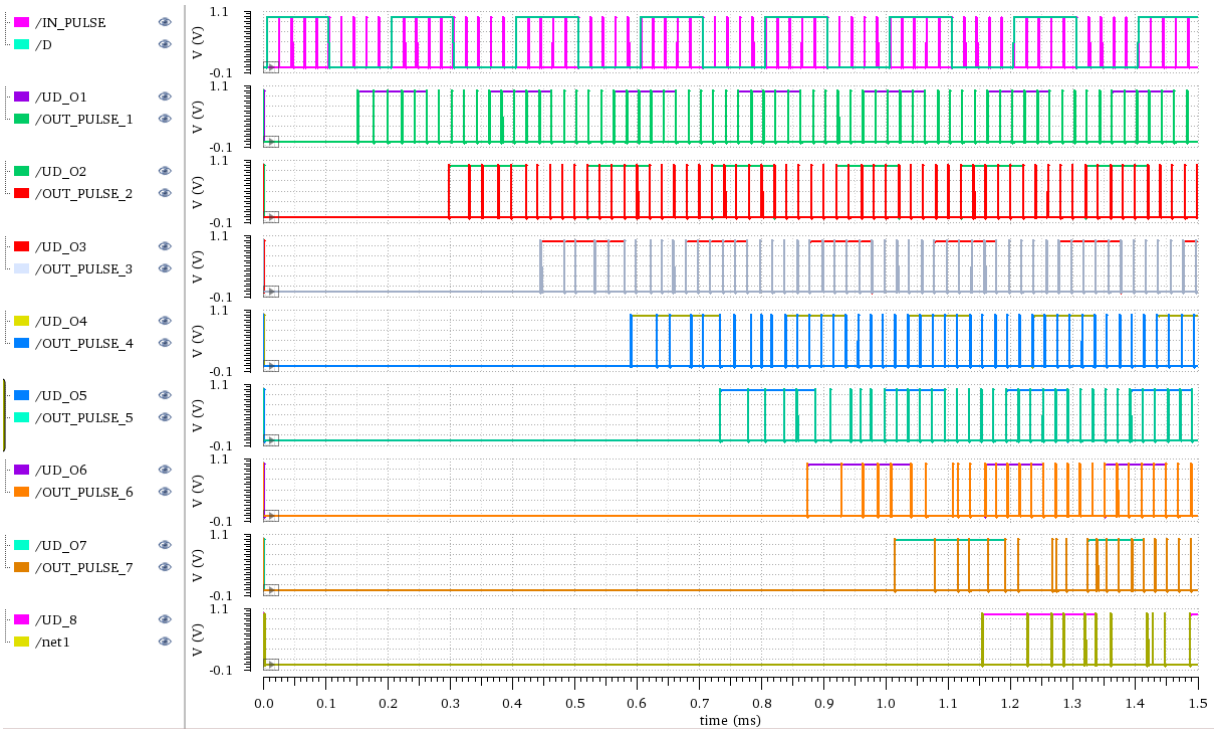


FIGURE 105 EIGHT DELAY TAP STRUCTURE SIMULATION

Table 17 shows a state-of-the-art comparison of different delay elements, including our digital delay cell.

TABLE 17 PERFORMANCE COMPARISON BETWEEN THE STATE-OF-THE-ART AND OUR DESIGNED DELAY ELEMENT

Parameter	Vezyrtzis, 2014 [51]	Schell, 2008 [50]	Sourikopoulos, 2016 [53]	Kurchuk, 2010 [54]	Chen, 2018 [49]	**This work**
V_{DD}	1 V	1V	1V	1.2V	1.2V	**1V**
Tuning range	15-500 ns	5 ns – 1 µs	110ps - 4.5 ns	0.3 - 300 ns	5 ns - 10µs	**5 ns-51.25 µs***
Matching	-	-	-	12.4%	2.3%	**9-11%**
Energy/delay	50 fJ	50 fJ	12.5fJ	20 fJ	83fJ	**•15.4 fJ**
Area	-	36 µm²	21 µm²	-	97µm²	**74 µm²**
Input-independent Delay	Yes	Yes	Yes	Yes	Yes	**Yes**
Robustness	Potential Hazard	Potential Hazard	No handshaking	No handshaking	Hazard-free handshaking	**Hazard-free handshaking**

***Maximum delay value obtained for a biasing current of 25pA •Does not includes the consumption of the D latch**

- Compared to the state-of-the-art, our delay offers the largest tuning range, 5 times larger than the previous best.
- Among the implementation including handshaking our delay exhibits the best consumption.
- For the mismatch, only [49] shows a better performance. As explained in chapter 5.3, mismatch can be traded with consumption explaining why consumption is higher than any other implementation.

5.15 Overall system performance and comparison with the State-of-the-art

To estimate the overall system performance, we also have to consider the associated D latch used to carry the UD signal. The addition of D latch raises the dynamic power consumption of the delay element from 15.4fJ/event up to 36.7fJ/event. The static power consumption only suffers a slight increase of 2pW. It is to note that since the D latch is driven by the SR latch, the simulated contribution of the SR latch to the overall power consumption contribution is increased significantly.

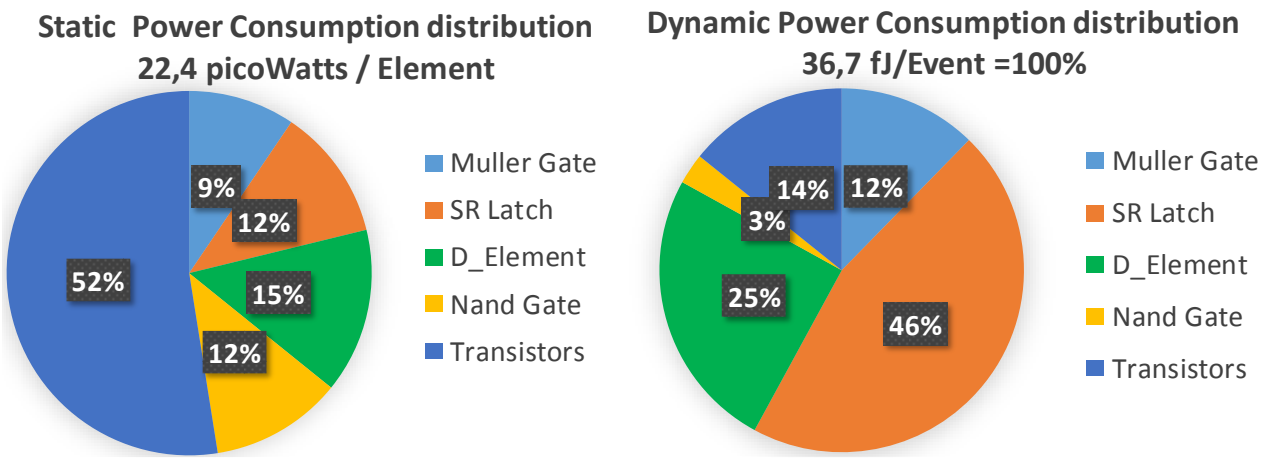


FIGURE 106 POWER CONSUMPTION BREAKDOWN OF THE DELAY ELEMENT WITH D LATCH

Table 18 shows a comparison between the performance of the state-of-the-art feature extractors and our proposed system based on the delay element performance. The methodology employed is the same used in chapter 4.7 with the scenario of female speech in a noiseless environment, using the final consumption figures.

TABLE 18 COMPARISON OF SYSTEM PERFORMANCE WITH STATE-OF-THE-ART IMPLEMENTATIONS

Feature Extractor	Yang, 2018 [12]	Yang, 2016 [44]	Badami, 2016 [11]	Yip, 2015 [10]	Wu, 2017 [9]	**This Work**
Technology	180nm	180nm	90nm	180nm	65nm	**28nm FDSOI**
Feature Type	Analog	Analog	Analog Filters	Digital Filters	Digital Filters	**CT-Filters**
Channel Number	16	64x2	16	8	18x4	**12**
Frequency Range (Hz)	100 – 5k	8 – 20k	75 – 5k	300 – 5.5k	160 – 8k	**60 – 6.3k**
Power (µW)	0.38	55	6	1.59[a]	13.8	***♪•2.85**
Power/Channel (nW)	24	430	380	199	190	**240**
Area/Channel (mm^2)	0.1	0.26	0.13	0.19	0.00934	**< 0,1 mm**

*FOR TYPICAL FEMALE SPEECH IN A NOISELESS ENVIRONMENT ♪AT A TEMPERATURE OF 23°C
•DELAY LINES BIASED AT 1V

Among the analog solutions, [12] outperforms the competitors. Nevertheless, when analyzing his paper, we can observe that out of band filter rejection is very poor.

For example, a full-scale 200Hz tone (corresponding to the central frequency of filter 3), produces a signal on filter 1 output, 9dB below full scale. This is obviously way below the 20dB needed to fulfill our system specification and translates into a poor hit-rate (82%). Except for [12], our proposed solution has the lowest power consumption and area.

Compared with digital implementations our system has a slightly higher power consumption per channel. However this is compensated by the fact that the consumption of our system is event-driven. In terms of area only the solution proposed by [10] has a smaller silicon area.

Our original goals of creating a system with event-driven power consumption, in the range of existing analog implementations, while occupying a silicon area in the range of existing digital implementations have been achieved.

5.16 Conclusion

An optimized thyristor-based digital delay element has been designed to meet the stringent constraints defined in our analysis of the proposed CT filter bank. (Chapter 4).

The key enablers are:

- The use of thick-oxide transistors to drastically reduce the gate leakage current, and thus the static power consumption.
- UTBB technology providing a near to ideal, down to the pico-ampere range, subthreshold slope allowing a large tuning range, low power consumption and reduced area.
- Small leakage thanks to SOI technology compatible with pico-ampere range operation.
- Non-doped channel improving the mismatch (even if the mismatch is degraded in subthreshold operation).

Silicon measurements will be fundamental to confirm the simulation promises and will be presented in the following chapter.

Chapter 6
Measurements and validations

6.1 Description

In the previous chapter, a detailed description of the design of a digital delay element was presented. In this chapter, we present the experimental setups and the obtained measurement results.

6.2 Integrated circuit and packaging

The die of our circuit is shown in Figure 107. It has 38 pads for inputs and outputs. The left-hand side pads are used for the three delays used for single delay characterization, while the rest of the pads are used by the delay tap structure.

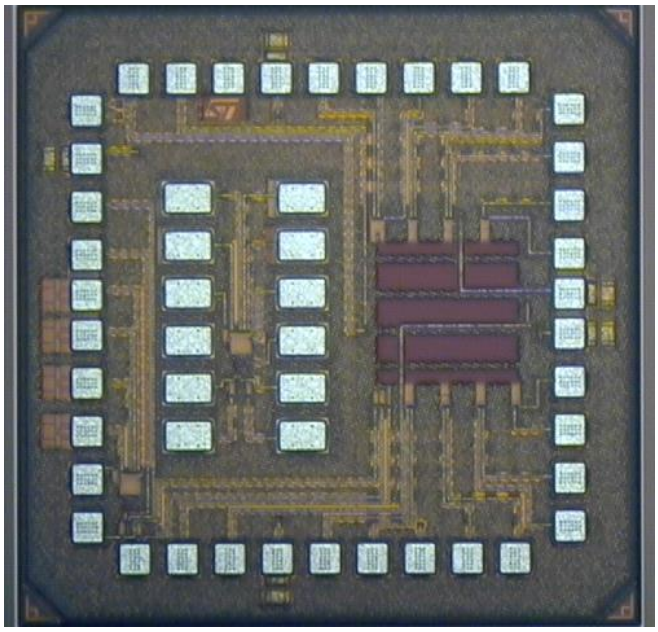


FIGURE 107 CHIP PHOTOGRAPH

The circuit is assembled in a QFN48 package shown in Figure 108, the package contains 48 pins, witch a pitch of 0.5mm. Its length and width are 7mm. Among the advantages of this package is its availability of sockets. The use of socket is not impacting the performance at such low-frequency and enables to test multiple dies without duplicating the test board. This is particularly important in our case where it is desired to test several dies, in order to evaluate die to die variations.

FIGURE 108 PACKAGED INTEGRATED CIRCUIT

Figure 109 shows the wire bonding connections between the die and the package. In Table 19 the wire bonding connections are detailed, also specifying the purpose of each signal.

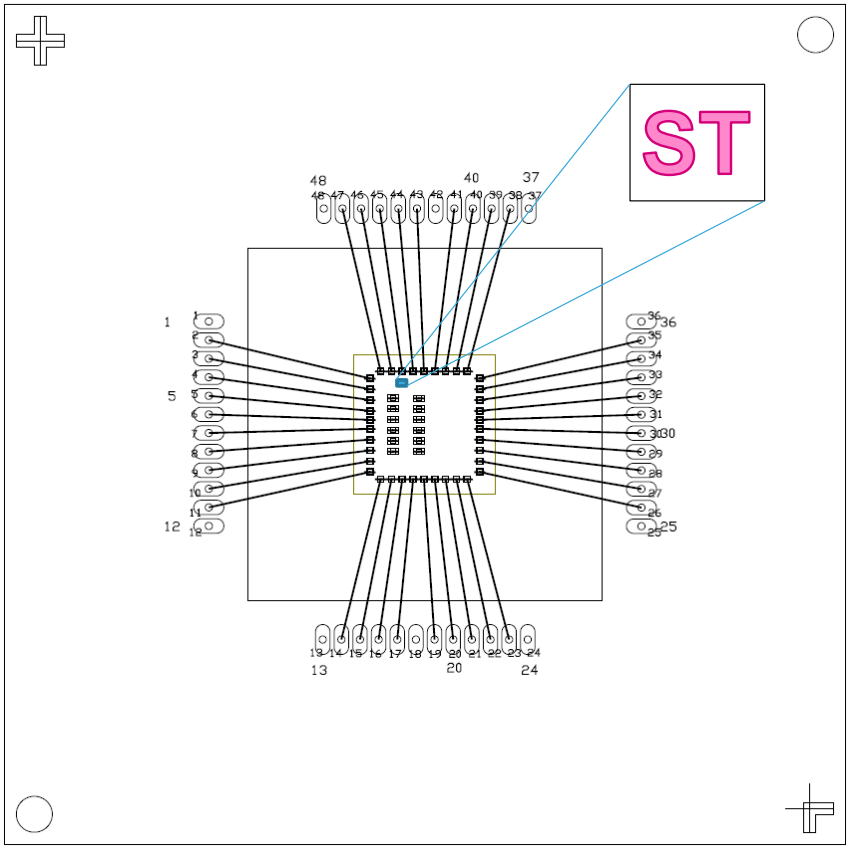


FIGURE 109 WIRE BONDING CONNECTIONS BETWEEN THE DIE AND THE PACKAGE

Signal	Die Pad	Package	Use	Signal	Die Pad	Package	Use
----	NC	1	----	-----	NC	25	-----
VDD	1	2	3D	UD_OUT3	20	26	TAPS
GND	2	3	3D/TAPS	CHANGE_OUT3	21	27	TAPS
Cascode	3	4	3D	UD_OUT4	22	28	TAPS
Current Control	4	5	3D	CHANGE_OUT4	23	29	TAPS
ACK_OUT	5	6	3D	GND	24	30	3D/TAPS
Change_OUT1	6	7	3D	VDD Buffers	25	31	3D/TAPS
Change_OUT2	7	8	3D	UD_OUT5	26	32	TAPS
UD_OUT	8	9	3D	CHANGE_OUT5	27	33	TAPS
Start	9	10	3D	UD_OUT6	28	34	TAPS
UD_IN	10	11	3D/TAPS	CHANGE_OUT6	29	35	TAPS
----	NC	12	----	----	NC	36	----
----	NC	13	----	----	NC	37	----
CHANGE_IN	11	14	3D/TAPS	UD_OUT7	30	38	TAPS
ACK_OUT	12	15	TAPS	CHANGE_OUT7	31	39	TAPS
Control Current	13	16	TAPS	CHANGE_OUT8	32	40	TAPS
GND	14	17	3D/TAPS	ACK_IN	33	41	TAPS
----	NC	18	----	-----	NC	42	-----
UD_OUT1	15	19	TAPS	UD_OUT8	34	43	TAPS
CHANGE_OUT1	16	20	TAPS	VDD Buffers	35	44	3D/TAPS
VDD	17	21	TAPS	NC	36	45	----
UD_OUT2	18	22	TAPS	Cascode	37	46	TAPS
CHANGE_OUT2	19	23	TAPS	Start	38	47	Taps
------	NC	24	-----	----	NC	48	----

TABLE 19 PINOUT CONNECTIONS BETWEEN THE DIE AND THE PACKAGE

6.3 Experimental setup

The socket and the board employed are shown in Figure 110. For this setup, only the signals noted 3D in the use column of Table 19 are connected. The board only contains the connectors required to bias the circuit and to feed the input signals, no external parts are needed for the circuit operation.

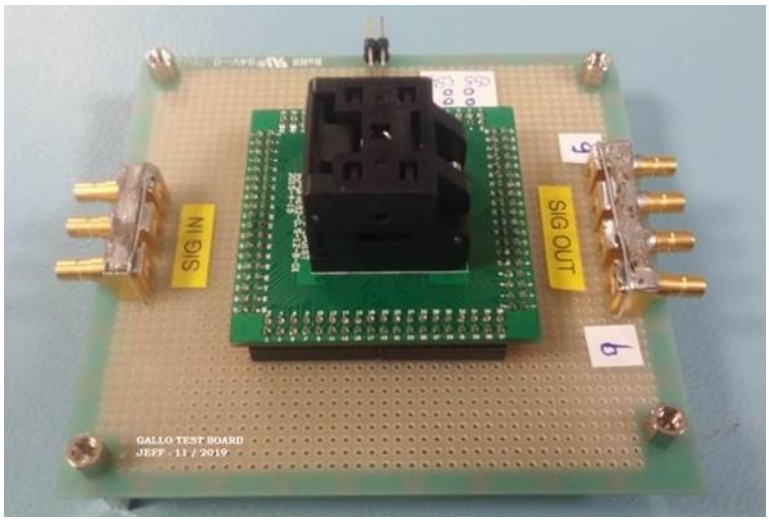


FIGURE 110 SOCKET AND BOARD EMPLOYED FOR MEASUREMENTS

To validate the delay, we need to generate three digital signals: CHANGE_IN and UD_IN emulating the CT ADC, and a Start acting as a reset. For this purpose, a data pattern generator is used.

The test bench setup schematic is shown in Figure 111. The equipment employed in the setup is listed below:

- 2 System source meter Keithley 2602 used to generate the voltages for VDD and the cascode and control current references, to measure the power consumption
- 2 Digital multimeter Agilent 3458A used as ammeters
- A digital signal generator DG2020A
- An Oscilloscope TDS3034B

Figure 112 shows the implementation of the test bench.

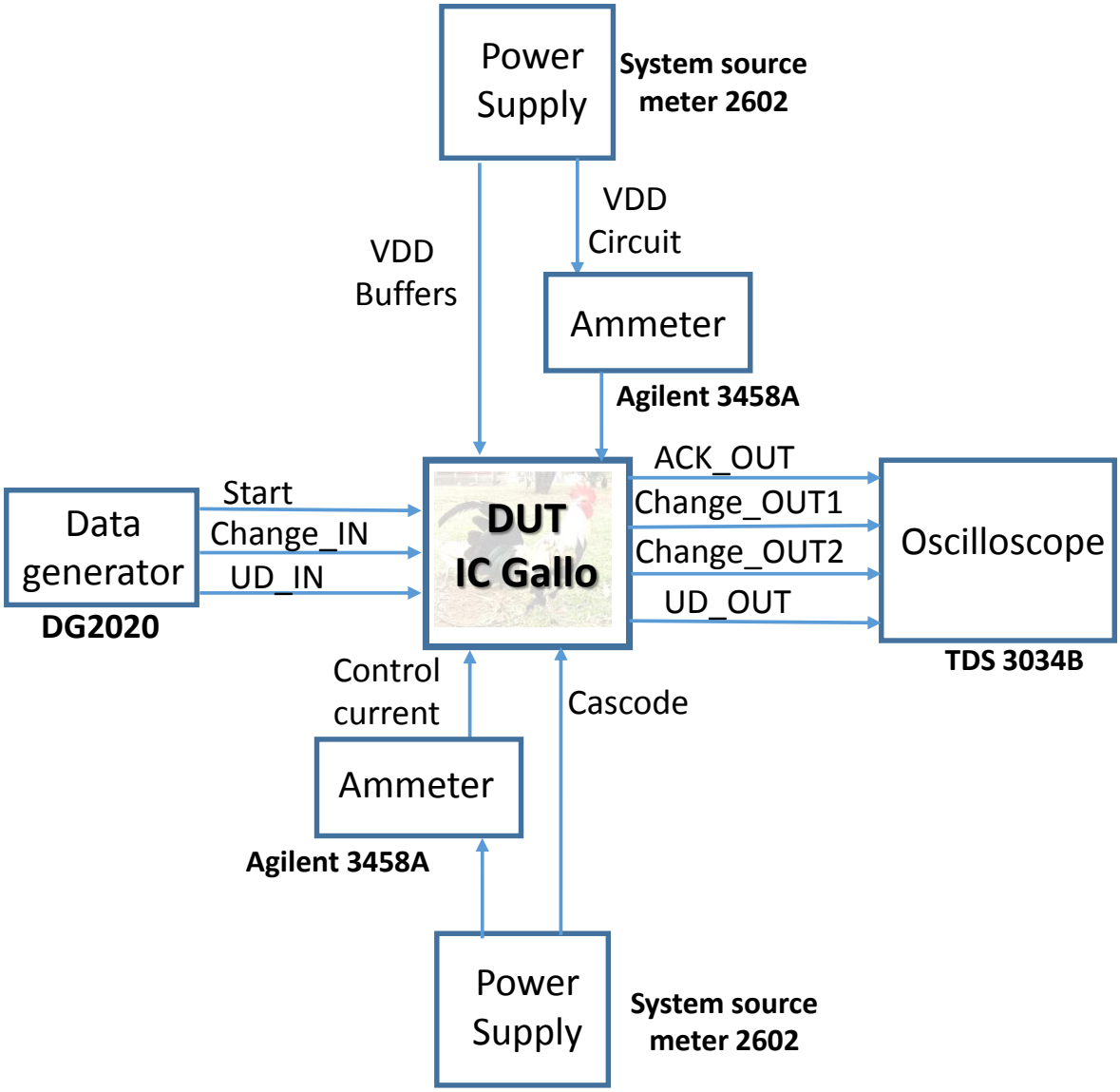


FIGURE 111 TEST BENCH SCHEMATIC

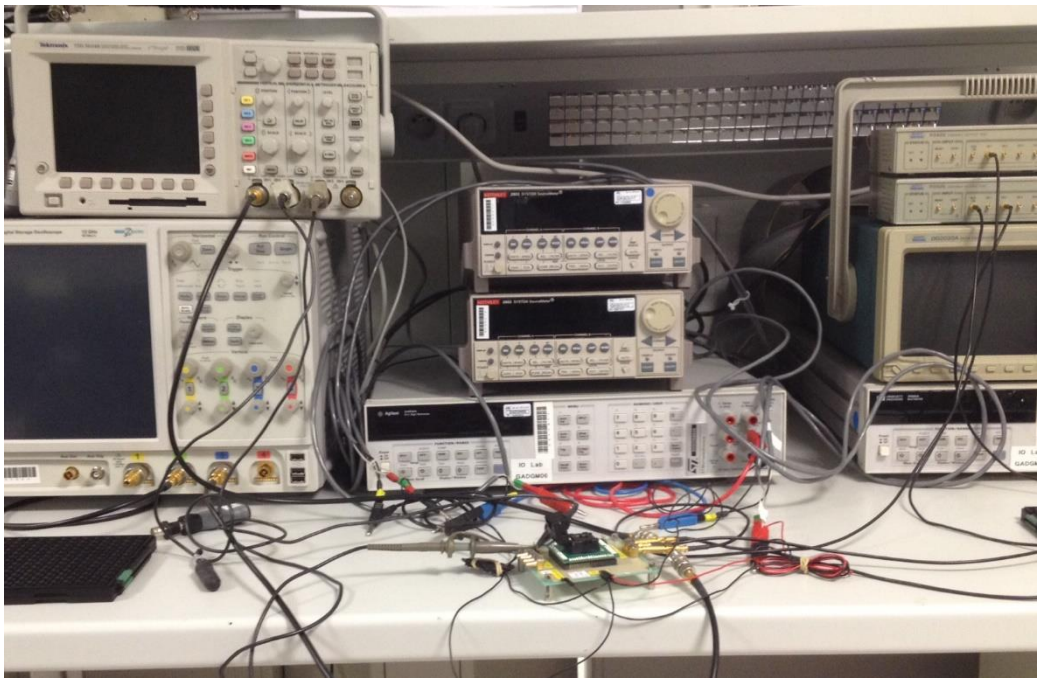


FIGURE 112 TEST BENCH IMPLEMENTATION

6.4 Experimental results

6.4.1 Biasing of the delay element

The first step is to make sure that we are capable of biasing the delay, in the same conditions as used in simulation (100pA – 5nA reference current). The biasing circuit is shown in Figure 113 and is composed of two symmetrical structures to generate both references for the control current and the cascode biasing.

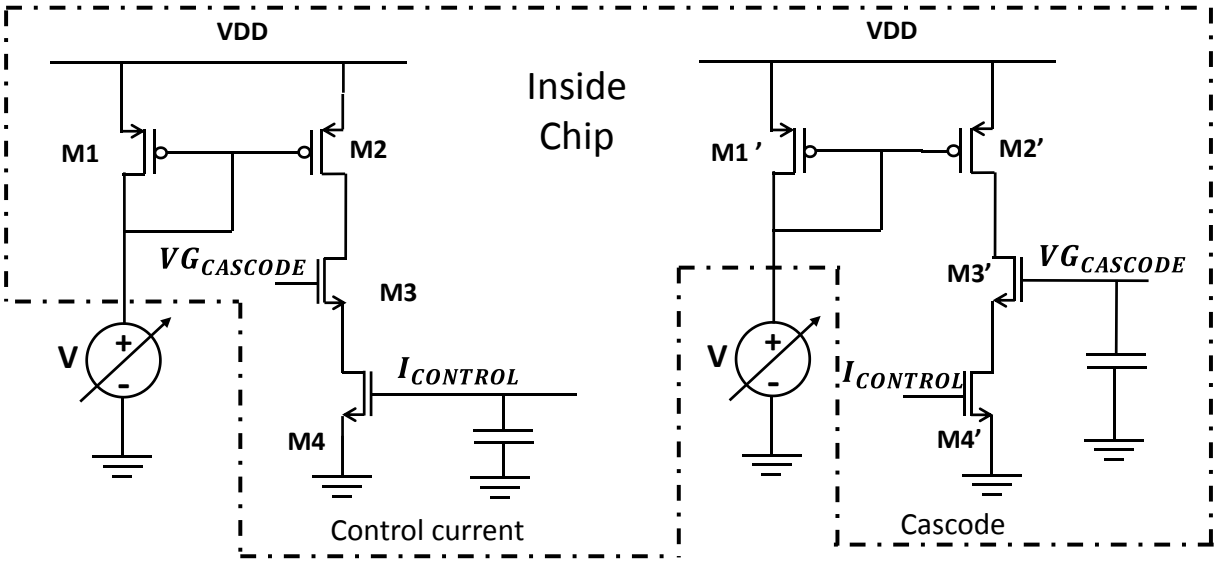


FIGURE 113 BIASING CIRCUIT OF THE DELAY STRUCTURE

Figure 114 shows the characterization curve of the PMOS transistor (M1). VDD voltage was kept constant at 1V and the current control pin voltage was varied from 0 to Some small discrepencies appear for VDD – VGD voltages between 0.3V and 0.5 V, at 0.3V this is due to the noise added by the parasitic components of the biasing

sources, that make measures below 100pA particularly difficult. For 0.4 V and 0.5 V this could be attributed to small variations between the measured biasing tension and the real tension at the node. For instance a small variation of 0.02 V is sufficient to increase or decrease the current by a factor 2. Finally for VDD – VGD voltages below 0.2 V the measured value correspond to the noise fluctuations (some pA).

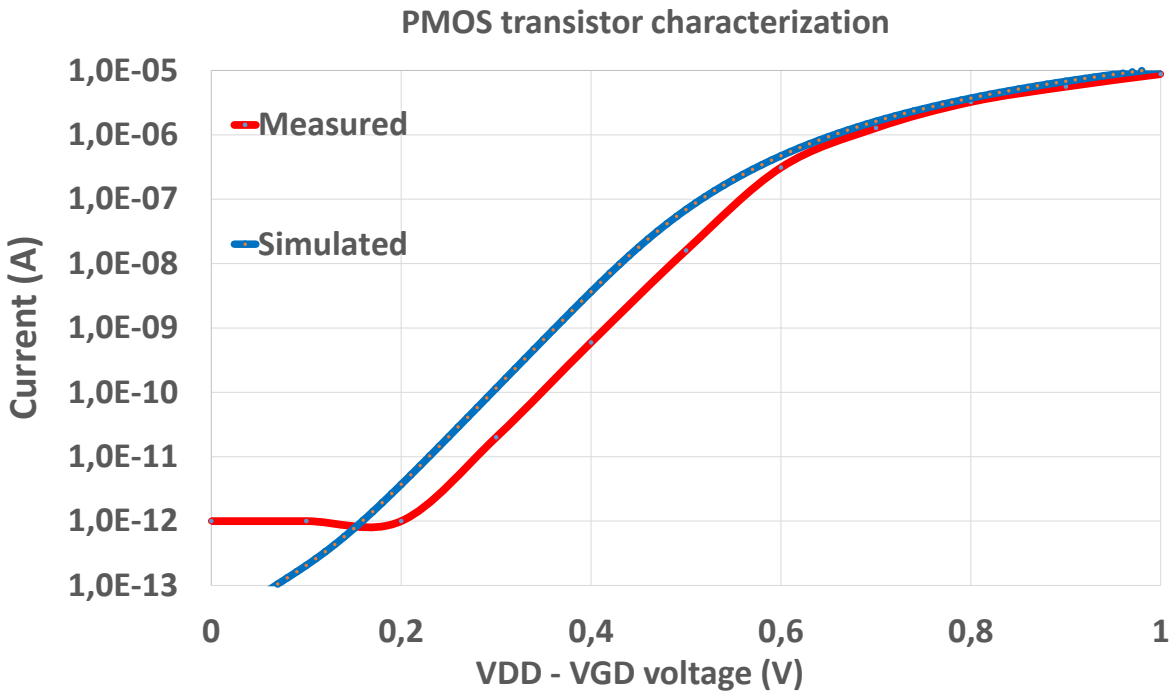


FIGURE 114 CHARACTERIZATION OF THE PMOS TRANSISTOR EMPLOYED AS CURRENT REFERENCE

6.4.2 Delay element characteristics

Knowing that we are capable of biasing the delay, let's check that its behavior conforms with simulations.

The first pattern consists of a Start signal, followed by a sequence of regular 200ns pulses with a 100ms period on CHANGE_IN. The UD_IN signal is kept low. The results are shown in Figure 115. We can see the CHANGE_IN signal (blue), the ACK_OUT (cyan), and CHANGE_OUT1 (pink) signals generated by the delay in response to the stimuli.

When an input pulse arrives, the acknowledge signal is pulled down indicating that the pulse has been received. We can notice the impact of the Muller gate employed to implement the handshake logic: the acknowledge signal is kept at its low value until the input pulse has ended. Thus, both pulses have approximately the same duration of 200ns.

The impact of asynchronous logic is also seen in the output signal. As soon as the output voltage begins to increase, the Muller gate of the next delay detects the rising edge, thus, the handshake logic generates after two NAND gates delays an acknowledge pulse resets the delay. Therefore, the output pulse is narrow, approximately 40ns. We notice that the pulse width fluctuates, without any explanation so far.

The asynchronous nature of the acknowledge signal is not suited for triggering the oscilloscope and measuring the delay time.

A better candidate to measure the delay is the UD_IN signal.

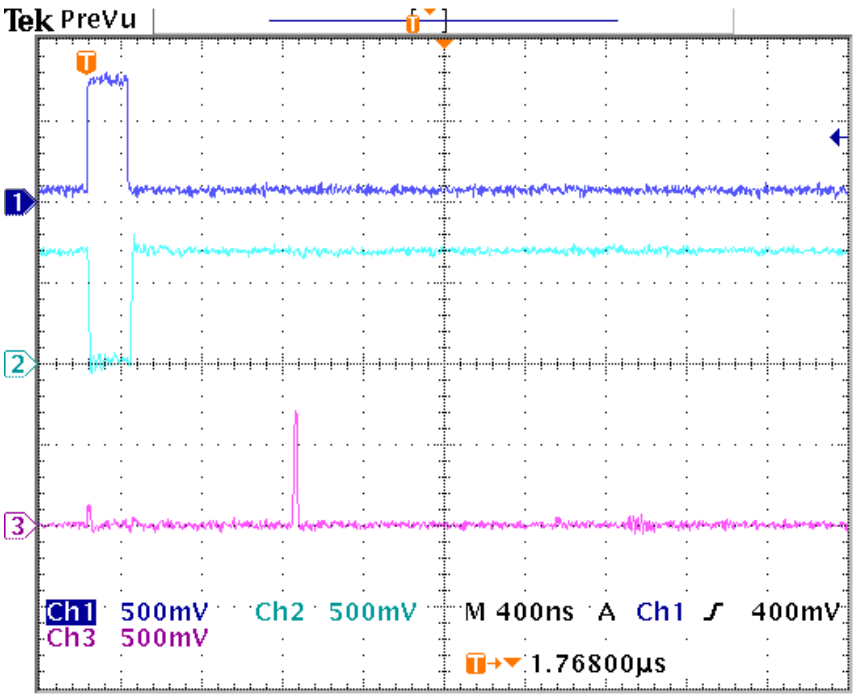


FIGURE 115 DELAY SIGNALS: INPUT SIGNAL (BLUE), ACKNOWLEDGE SIGNAL (CYAN), OUTPUT (PINK)

A second data pattern is created. Start and CHANGE_IN signals are kept the same, the UD_IN signal is now toggling between 0 and 1 in the middle of change pulses. The pattern and the outputs are shown in Figure 116.

Opposite to ACK_OUT signal, the UD_OUT signal exhibits clean edges and a 50% duty cycle and thus is suitable for oscilloscope triggering. Delay value will be measured between UD_IN and UD_OUT.

Figure 117 and Figure 118 complement Figure 116 with different reference currents. For the three figures we can see the CHANGE_IN (blue) and UD_IN (pink) input signals, the ACK_OUT (cyan) and UD_OUT (green) outputs generated by the delay in response to the stimuli. We can visualize the large tuning range of the delay element. Going from 5.43µs (Figure 116) to 97µs (Figure 118). It is to notice that the delayed output can be also be considered as a phase-shifted version of the input.

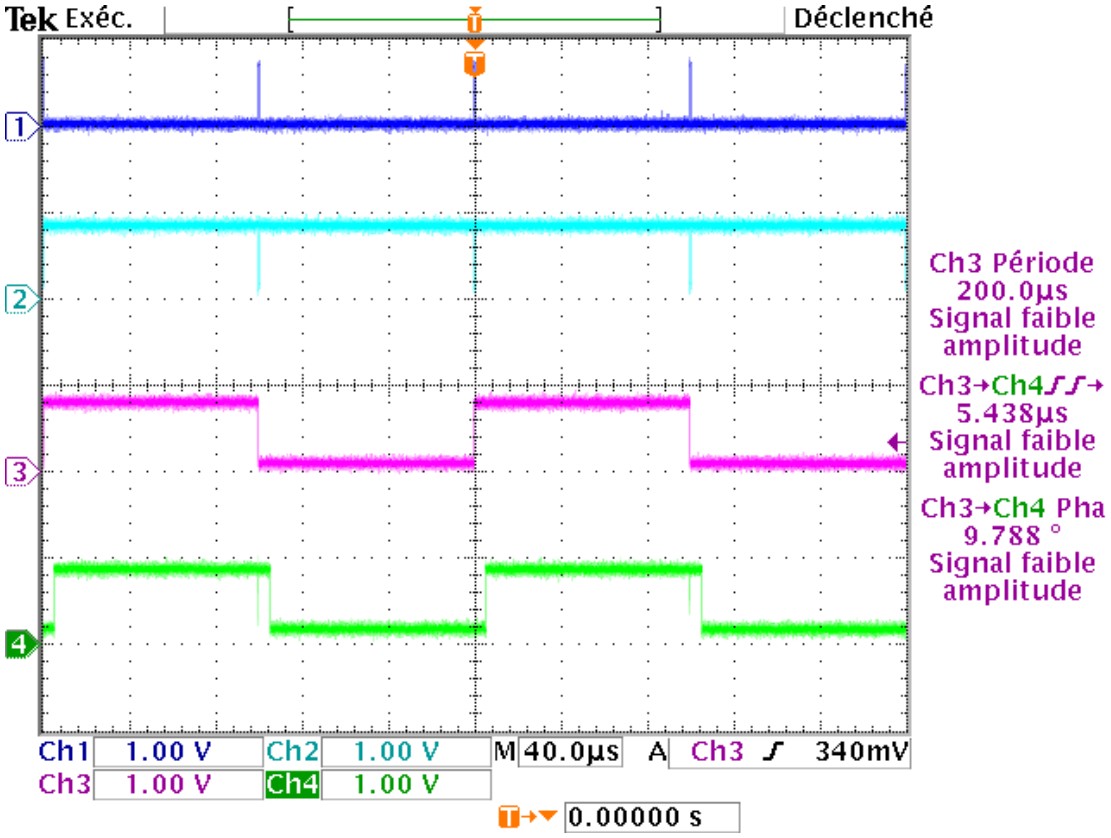


FIGURE 116 OSCILLOSCOPE SCREENSHOT DISPLAYING A 5.43 µS DELAY TIME

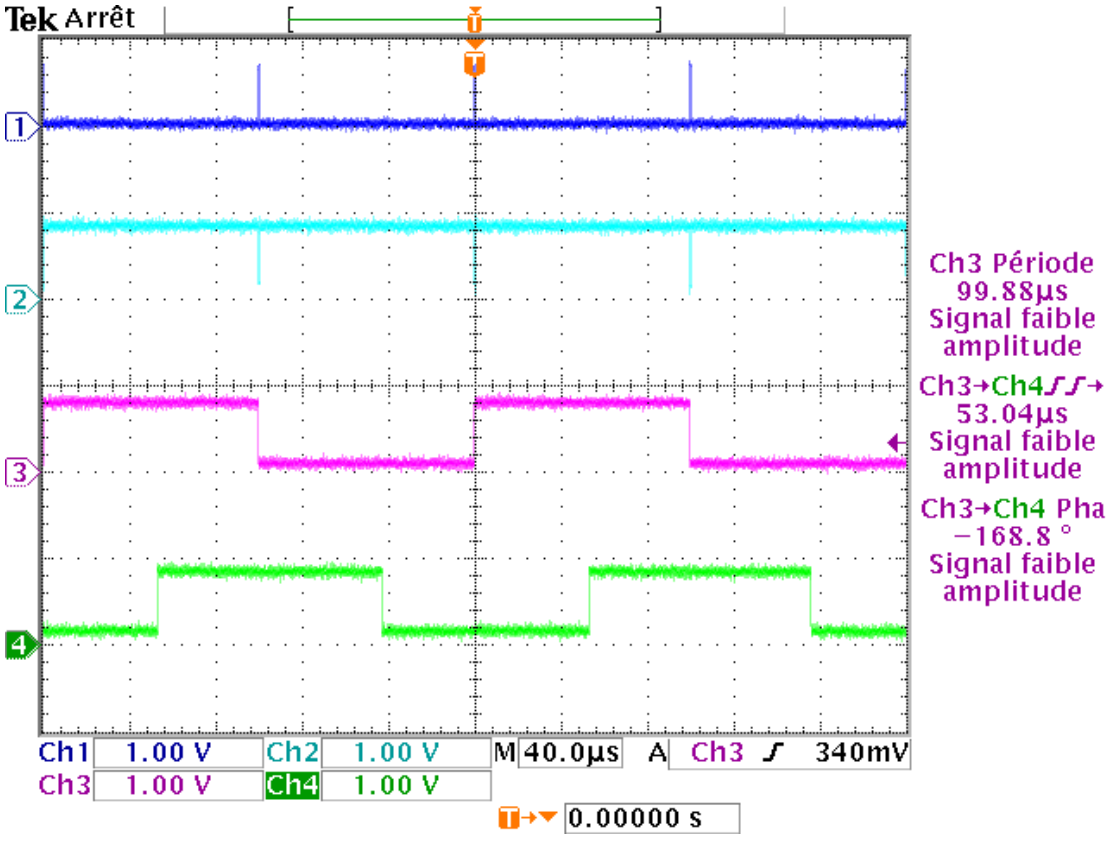


FIGURE 117 OSCILLOSCOPE SCREENSHOT DISPLAYING A 53 µS DELAY TIME

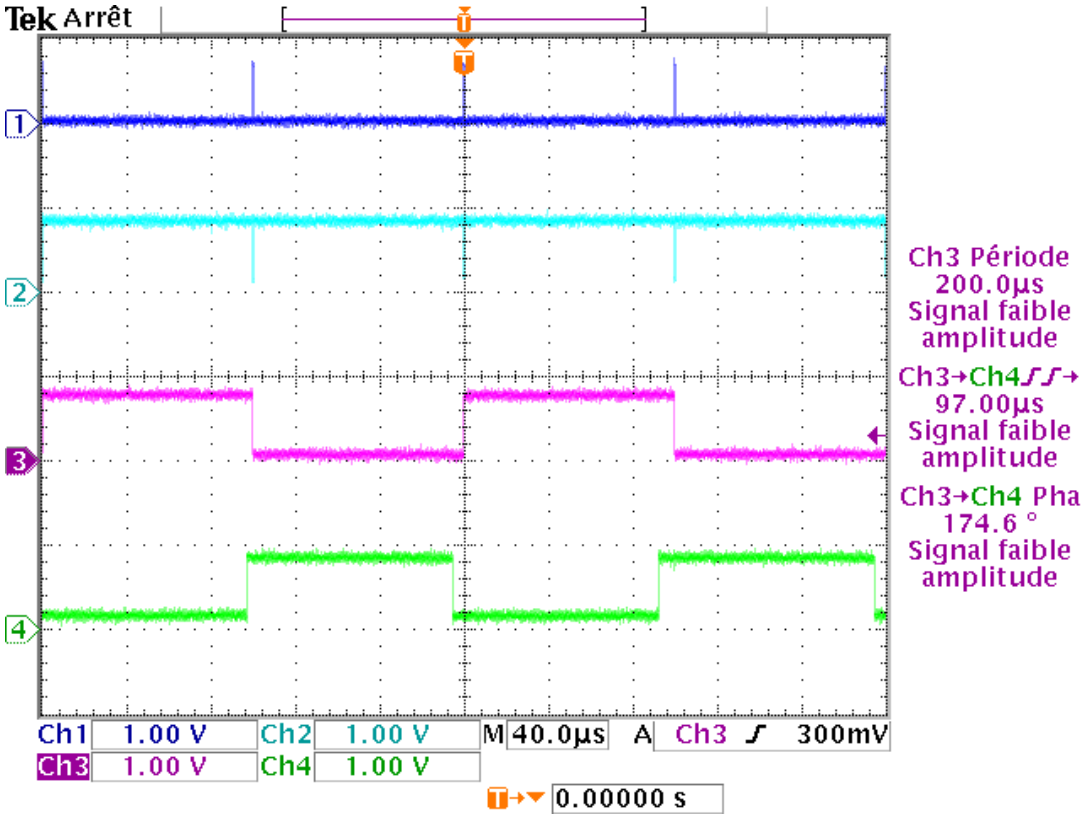


FIGURE 118 OSCILLOSCOPE SCREENSHOT DISPLAYING A 97 µS DELAY TIME

In this section, we have validated that the operation of the delay was conform to the simulation and we can proceed with quantitative evaluation.

6.4.3 Measurement variability

The circuit operates with very low currents well below the 1 nA, making it very sensitive to noise and supply voltage fluctuations. Thus, before proceeding with the characterization is important to evaluate the impact of variability.

The measured control current fluctuations depends significantly on the current meter caliber. Indeed, for the noise perspective, the current meter behaves as a resistance in series with the voltage source, fluctuating from 1Ω to 545kΩ depending on the selected measurement range. The voltage noise of the resistance is multiplied with the Gm of the transistor to produce current noise.

To minimize the impact, the current meter in the reference path will be tuned to the lowest resistance (1Ω) and the current meter in the VDD path will be removed for sensitive measurements.

The impact of the current meter resistance on delay has been characterized using single-shot measurements. The results can be seen in Figure 119. Even if the resistance does not impact the average delay (18.66 µs and 18.72 µs for the lowest and highest resistance), dispersion increases with resistance value.

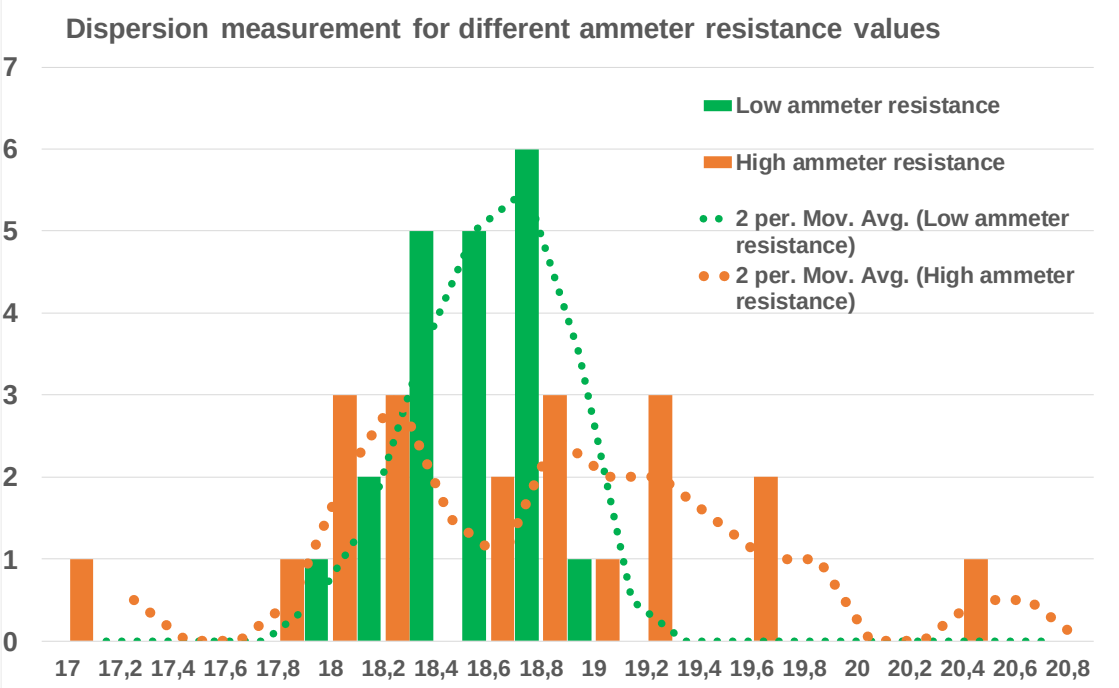


FIGURE 119 EFFECTS OF THE INTERNAL AMMETER RESISTANCE ON DISPERSION

6.4.4 Delay tuning range

We proceed to measure the delay tuning range. For this purpose, we quantify the delay time from currents going from 75pA up to 30nA, for a VDD voltage of 1V for different circuits as shown in Figure 120.

Although there is some expected variability due to mismatch, the chip shows a proper functioning and an average tuning range from 200ns to 24µs, for the measured current values.

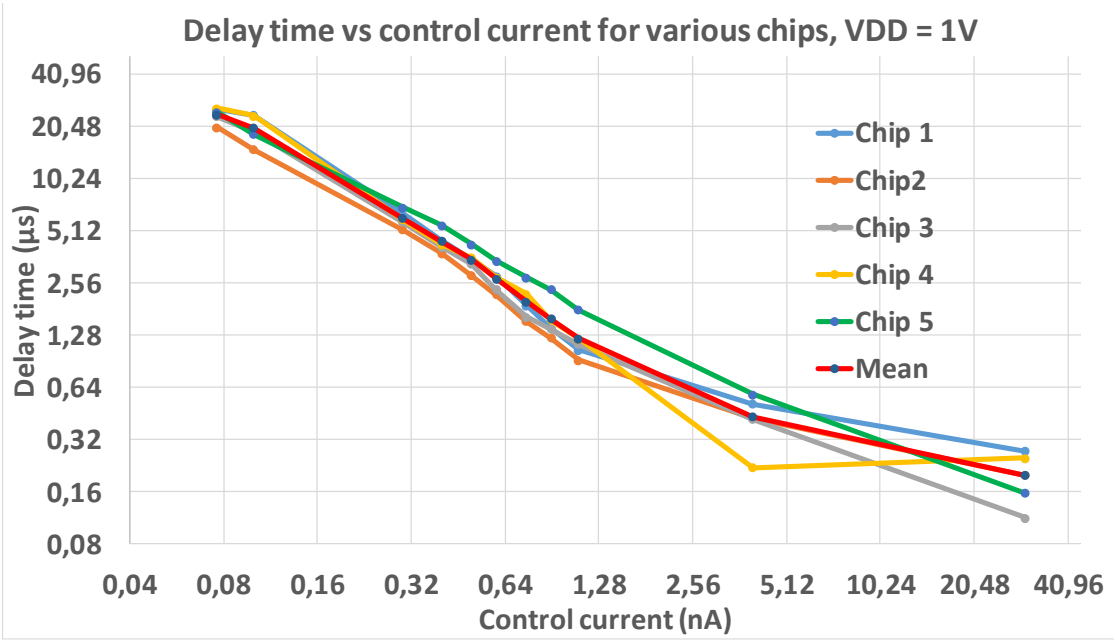


FIGURE 120 DELAY TIME VS CONTROL CURRENT FOR VARIOUS CHIPS

We now proceed to measure the delay tuning range for different VDD values. The simulated results are shown in Figure 121 while the measured values area shown in Figure 122. The measurement dispersion is also reduced for lower VDD voltages.

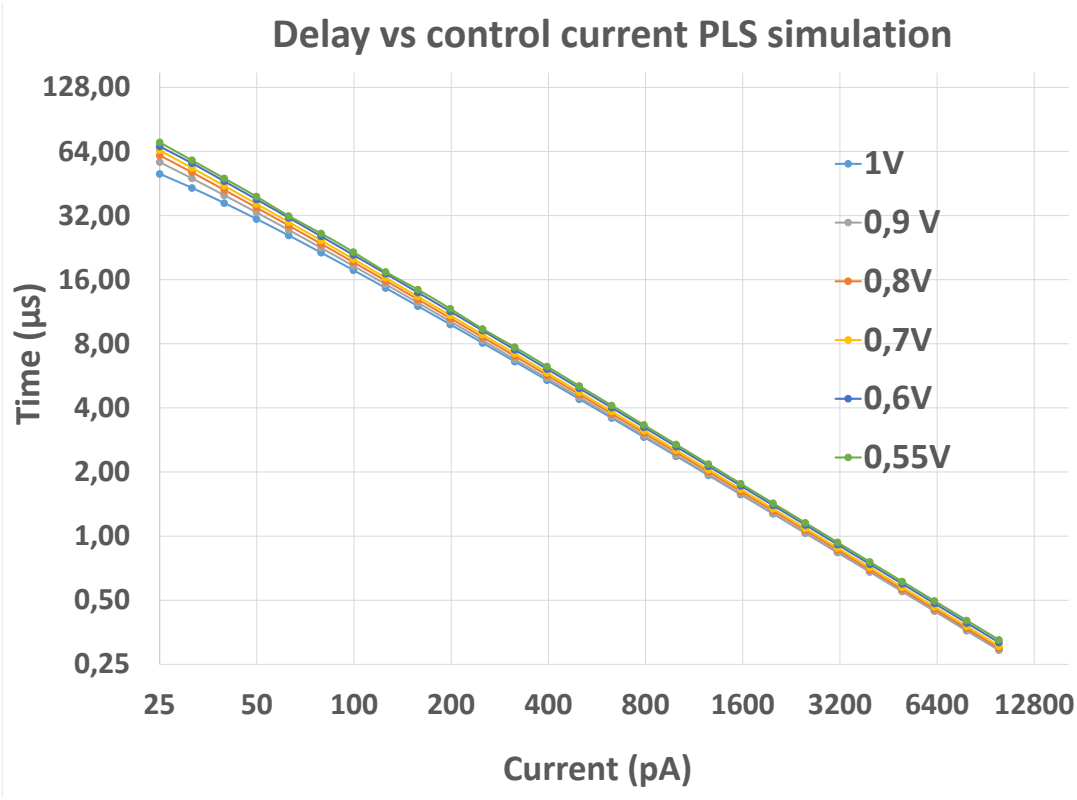


FIGURE 121 SIMULATED DELAY TIME FOR DIFFERENT VDD VALUES

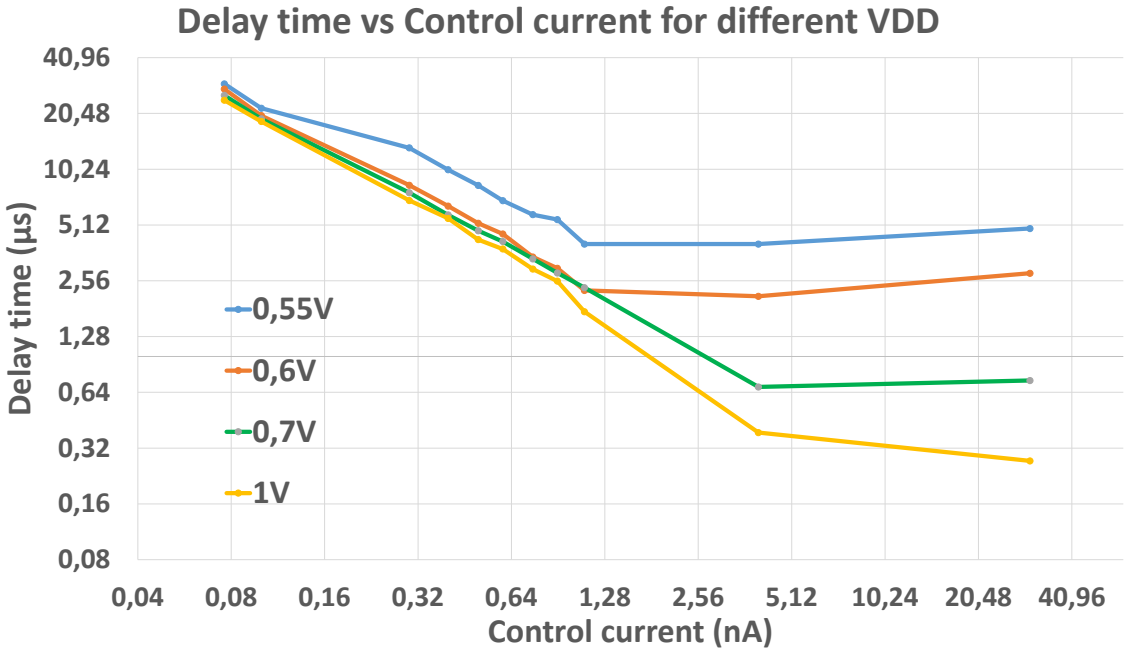


FIGURE 122 DELAY TIME FOR DIFFERENT VDD VALUES

The lowest VDD voltage achieving the desired tuning range for our application (825ns to 16µs), is 0.7V.

The delay value increases as the VDD voltage decreases. In the case of larger currents above 1.1 nA, the delay “saturates” for lower VDD values. Since the VDD voltage is decreased, the cascode transistor used to protect the control current transistor enters in ohmic mode and thus does not operate properly. To overcome this problem, the cascode voltage should be increased. As an example, the tuning range with VDD = 0.7V (Green curve in Figure 122) limited to approximately 750ns has been extended to 30ns by increasing cascode reference current from 1nA to 100nA. Figure 123 is an illustration of this operating mode, we can see the CHANGE_IN (blue) and UD_IN (pink) input signals, the ACK_OUT (cyan) and UD_OUT (green) outputs generated by the delay in response to the stimuli.

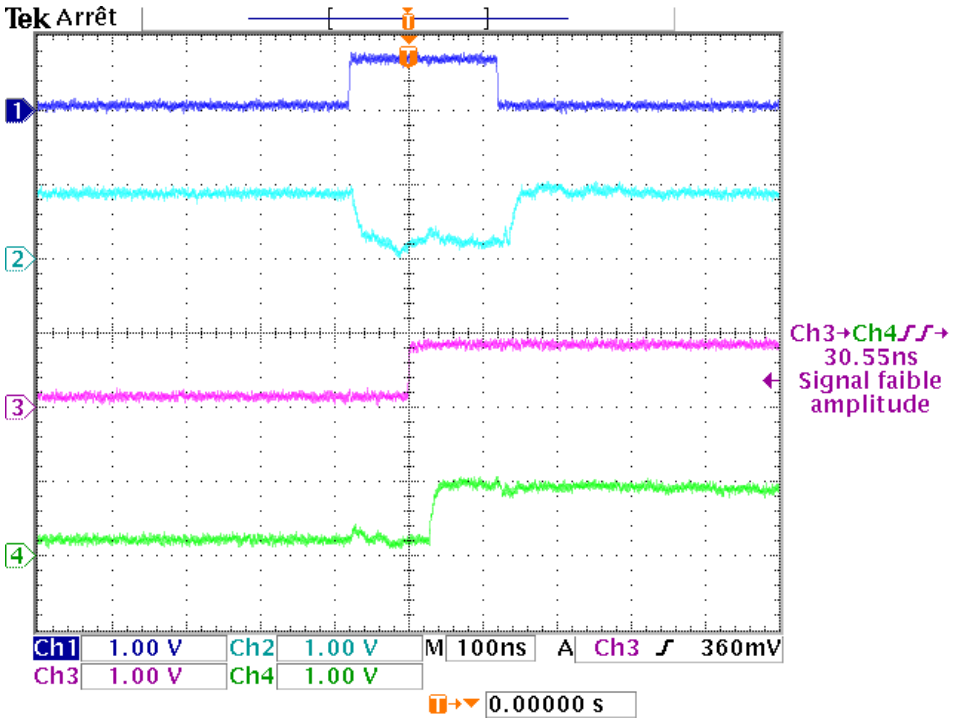


FIGURE 123 SHORTEST DELAY OBTAINED IN MEASUREMENTS FOR A VDD OF 0.7V

The lowest VDD voltage at which the circuit remains functional is 0.53V. For lower voltages handshaking elements generate false acknowledge signals, causing oscillations in the outputs. We recall that the logical gates implemented for the handshaking use EGVLVT transistors with a threshold voltage of around 0.35V.

6.4.5 Delay power consumption

First, the static power consumption is computed as the product of the VDD value and the current consumption, without applying any stimuli.

Then a periodic signal at 10 kHz is applied to the CHANGE_IN input, and the dynamic power consumption is calculated. The energy per event is thus the difference between dynamic and static consumption, divided by repetition frequency and the number of delay cells supplied by VDD (3). The energy value includes the energy required to control the latch that will be used to carry the UD_IN signal.

The static power consumption is measured at 80pA per delay cell. The dynamic energy per event is shown in Figure 124 for different VDD and control current values, among with the corresponding delay tuning range.

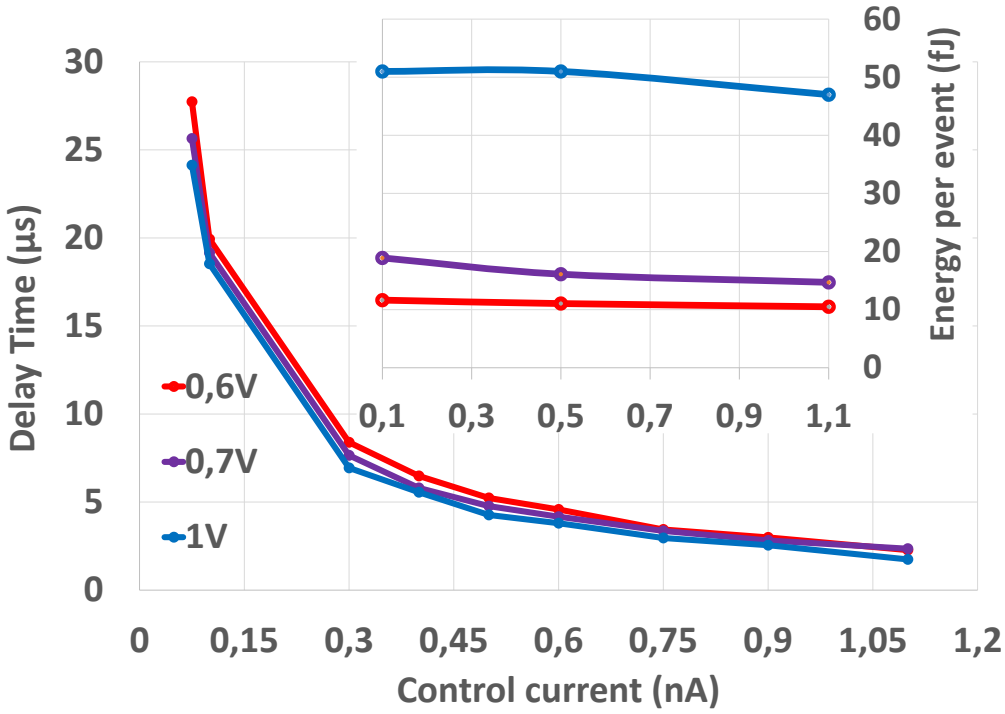


FIGURE 124 DYNAMIC POWER CONSUMPTION FOR DIFFERENT VDD VOLTAGES, THE CORRESPONDING DELAY TIME IS ALSO INDICATED

Measurement results are in line with simulation, we can see that reducing VDD, at the price of a lower tuning range, improves drastically the power consumption.

6.4.6 Input signal delay dependency

The input signal delay dependency was a major challenge of the design. Measurements have been performed to check this issue was properly tackled.

Input signal dependency has been evaluated using a repetitive input pulse, whose period is spanned from 5 µs to 200 µs. Figure 125 shows the measured input signal dependency on the input pulse period. Despite certain variability of the biasing source, it remains below 1.5% in the worst-case scenario, remaining steady at approximately 0.5% for pulses spaced more than 10µs. This performace is vastly sufficient to satisfy the system requirements.

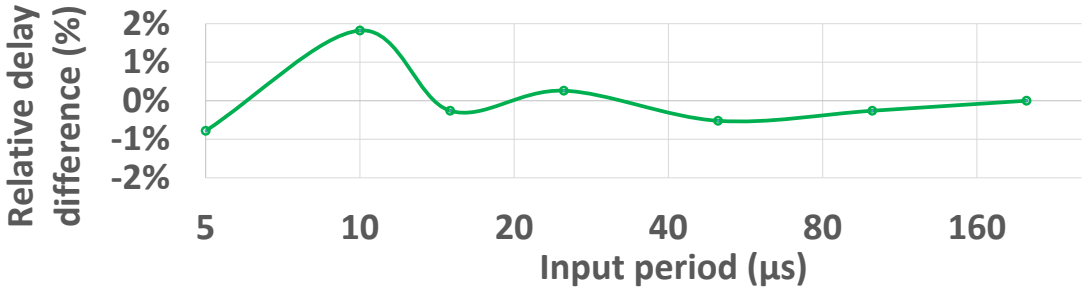


FIGURE 125 MEASURED DELAY TIME VERSUS THE PERIOD OF THE REPETITIVE INPUT PULSES FOR A DELAY CELL CONFIGURED AT 3.84 µS AND A SUPPLY VOLTAGE OF 0.7 V

6.4.7 Delay error

To evaluate the delay variations over multiple dies, the delay time of a unit delay cell was measured for two control currents (100 pA and 800 pA) across 22 different dies from the same wafer. Measurement results are shown in Table 20. The measured average delay is very close to the simulated value, while the standard deviation is measured to around 20-25%. It is to note that no calibration of any kind is applied here.

To ensure that the delay calibration can be performed for a single reference point, the relation between the relative error on the delay times for currents of 100 pA and 800 pA for each die can be analyzed. Figure 126 shows the relation between these errors for 22 dies. We can observe that the relative delay errors are correlated for the two current settings. If the calibration is performed for a 100 pA current, then the resulting standard deviation when the current is 800 pA is reduced to 5.01%. This property could be particularly beneficial when adding configurability at a system level.

TABLE 20 STATISTICAL ANALYSIS OF DELAY VARIATIONS DUE TO MISMATCH

Reference current	100 pA	800 pA
Simulated delay time	19.95 µs	3.09 µs
Measured Average delay µ	20.60 µs	3.06 µs
Measured Std. Deviation σ	4.62 µs	0.71 µs
σ/µ %	22.43%	23.20%

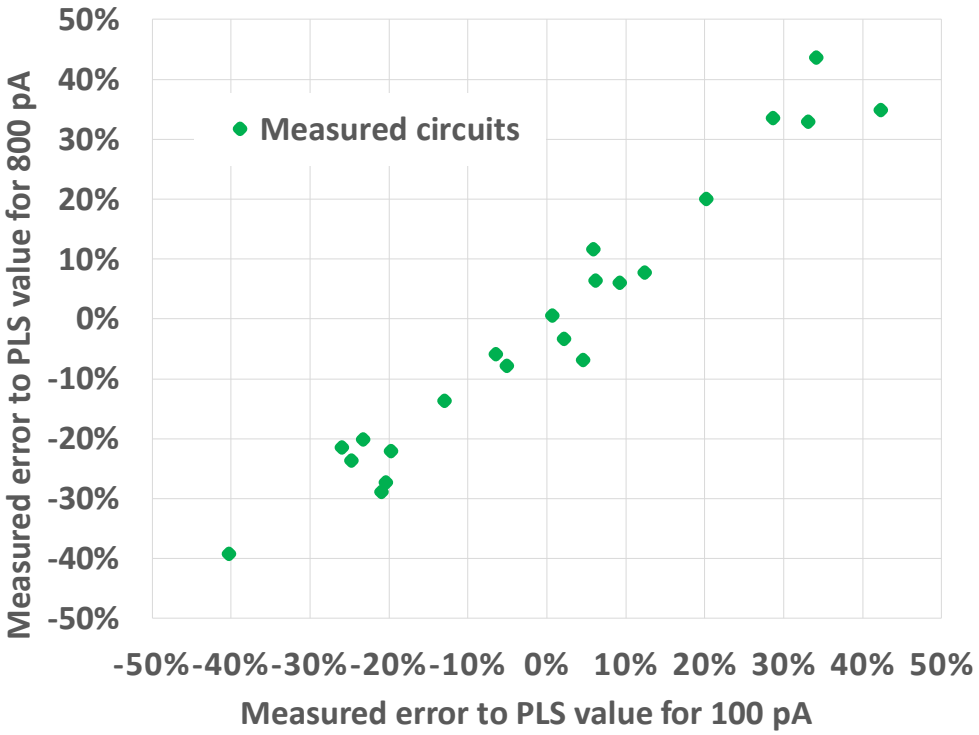


FIGURE 126 STATISTICAL DISPERSION IN RELATIVE PERCENTAGE WITH RESPECT TO THE SIMULATED VALUES, FOR 2 CONTROL CURRENTS MEASURED FOR 22 CIRCUITS

6.4.8 Unit cells and delay tap mismatches

The local mismatch of delay cells on the same die has been evaluated by measuring the delay value of the taps of the delay line (composed of 8 sections of 96 unit delay cells). The control current is fixed at 5 nA. This measurement has been performed over 3 different circuits. For each circuit, Table 21 indicates the mean of the 8 measured delays and the standard deviation, as well as the combination of the 3 obtained values.

To calculate the mismatch associated with the unit delay cell, we need to introduce the biasing scheme of the complete delay line. The current mirror (transistor N5) is common for a section of 24 unit delay cells correlating the current deviation of the corresponding delays. The Icontrol for each section are mirror copies of the reference current. Assuming from the simulations of a unit cell that the control current deviation contributes about P = 50% of the total delay variations, the intrinsic mismatch of the delay cell can then be calculated in Table II using the following formula:

$$\frac{\sigma_{unit\_delay}}{\mu_{unit\_delay}} = \frac{\sigma_{tap}}{\mu_{tap}}\sqrt{\frac{96}{1+24k^2}}, k = \frac{P}{1-P} \quad (21)$$

The calculated value is consistent with simulation results obtained through 1000 Monte Carlo simulations, where mismatch varied between 9% for a 25pA current to 11% for currents above 100pA. The system requires a delay tap mismatch of 11%, and thus the measured mismatch performance of the delay is more than sufficient to comply with the system requirements.

TABLE 21 MEASURED MISMATCH FOR DIFFERENT DELAY TAPS

	Mean (µs)	Std. Deviation/Mean (%)
Chip 1	63.17	5.85
Chip 2	63.87	3.61
Chip 3	66.11	4.91
Total (24 points)	---	4.20%
Delay Cell	0.67	8.23%

6.4.9 Static power consumption

Table 22 shows the static power consumption of the delay element as a function of the temperature. The measured value was obtained at a temperature of 23°C by measuring the static consumption of the 8 delay taps and dividing by the number of individual delays (768 cells). The static power for the bias circuit is not included. As we can see the temperature plays an important role in raising the static power consumption almost by factor 10 when the temperature raises to 60°C.

TABLE 22 STATIC POWER CONSUMPTION AS A FUNCTION OF THE TEMPERATURE

Temperature (°C)	-20	0	23	40	60
Static power (pW)	7.84•	10.57•	14.32★	50.26•	130•

★MEASURED VALUE •SIMULATION RESULTS

6.5 Comparison with other works

The comparison of this work with the state-of-the-art is shown in Table 23. This work presents the largest tuning range, the largest absolute delay value and the highest energy efficiency for a delay element with handshaking. This work is also the only one that reports measured mismatch values.

TABLE 23 COMPARISON WITH STATE-OF-THE-ART DELAY CELLS

Parameter	Vezyrtzis, 2014 [51]	Schell, 2008 [50]	Sourikopoulos, 2016 [53]	Kurchuk, 2010 [54]	Chen, 2018 [49]	**This work**
V_{DD}	1 V	1V	1V	1.2V•	1.2V	**0.7V**
Tuning range	15-500 ns	5 ns – 1 µs	110ps - 4.5 ns	0.3 - 300 ns	5 ns - 10µs	**30 ns - 97 µs**
Matching	-	-	-	12.4%	2.3%★•	**8.23%**
Energy/delay	50 fJ	50 fJ	12.5fJ	20 fJ	83fJ	**15 fJ**
Area	-	36 µm²	21 µm²	-	97µm²	**74 µm²**
Input-independent Delay	Yes	Yes	Yes	Yes	Yes	**Yes**
Robustness	Potential Hazard	Potential Hazard	No handshaking	No handshaking	Hazard-free handshaking	**Hazard-free handshaking**

•Simulation results ★100 Monte Carlo simulations

The technology node 28 nm FDSOI and its improved channel control are the key enablers that have allow to increase the tuning range while maintaining a low-power consumption. The overall surface of the delay element is not reduced, since many of the transistors employed in the design have thick-oxide gate and have a larger surface, while the main current source has a significant gate length to prevent the effects of channel length modulation.

6.6 Delay design improvement suggestions

Some modifications to the structure could be done to ease the test of the delay cell.

- The use of weighted current mirrors as references with a division ratio of 10 to 20 would make the generation of the biasing current (and its measurement) much easier.

- As low duty-cycle signals (such as CHANGE_OUT) are very difficult to measure, all signals used out of the chip should be conditioned inside the chip. For example, the use of a toggle flip-flop could drastically improve the measurement accuracy.

6.7 System performance estimation with measured results

Table 24 shows a comparison of the estimated system performance based on the measured power consumption of the delay element operating at a voltage of 0.7 V at a temperature of 23°C. By biasing the delay elements with a voltage of 0.7 V, an important reduction in energy consumption is obtained. The detailed analysis is already presented in chapter 5.15.

TABLE 24 COMPARISON OF SYSTEM PERFORMANCE WITH STATE-OF-THE-ART IMPLEMENTATIONS BASED ON MEASURED DELAY PERFORMANCE

Feature Extractor	Yang, 2018 [12]	Yang, 2016 [44]	Badami, 2016 [11]	Yip, 2015 [10]	Wu, 2017 [9]	**This Work**
Technology	180nm	180nm	90nm	180nm	65nm	**28nm FDSOI**
Feature Type	Analog Filters	Analog Filters	Analog Filters	Digital Filters	Digital Filters	**CT-Filters**
Channel Number	16	64x2	16	8	18x4	**12**
Frequency Range (Hz)	100 – 5k	8 – 20k	75 – 5k	300 – 5.5k	160 – 8k	**60 – 6.3k**
Power (µW)	0.38	55	6	1.59[a]	13.8	**♪•*0.97**
Power/Channel (nW)	24	430	380	199	190	**♪•*81**
Area/Channel (mm^2)	0.1	0.26	0.13	0.19	0.00934	**< 0,1**

*FOR TYPICAL FEMALE SPEECH IN A NOISELESS ENVIRONMENT ♪AT A TEMPERATURE OF 23°C
•DELAY LINES BIASED AT 0.7 V

The proposed delay element biased at 0.7 V allows to improve the performance of the proposed system, by reducing the dynamic power consumption by approximately a factor 3. The delay line is thus a proper solution to implement the proposed CT filter bank, while maintaining a low-power consumption.

6.8 Conclusion

The performance of the digital delay cell designed and fabricated on 28nm FD-SOI has been measured.

The experimental results validate the designed digital delay element. The results obtained during measurements matched those expected during the simulation phase, corroborating the proper behavior in weak inversion of 28nm FD-SOI technology.

The temperature drastically affects the static power consumption of the system. This is an important factor to consider in the case of a sensing application where the environment can vary drastically.

Experiments show that we can control a delay as large as 100µs.

Future measurements could be the characterization of the delay taps and on-wafer measurements to trace charts of mismatch depending on the location on the wafer.

Chapter 7
Conclusion and perspectives

7.1 Conclusion

In this work, we studied the implementation of a CT DSP FIR filter bank applied to generate spectrograms, that will be used for audio recognition applications. The main premise of finding an alternative to Nyquist rate analog-to-digital conversion, to efficiently process sparse signals with a varying spectral rate has been achieved.

The proposed system architecture is inspired by the functioning of the human cochlea. It comprises two novel systems blocks to cascade CT FIR filters, and estimate the energy of a signal by calculating the instantaneous power, exploiting the properties that CT signals can only vary by one LSB at the time. Energy estimation is also used to create bandpass filters from the difference of two lowpass filter outputs.

The main components of a CT system are digital-delay taps. Since none of the existing delay elements in the state-of-the-art complies with our system specifications, a novel digital-delay element has been designed, taking advantage of the technological properties of the FD-SOI technology.

Measurement results show the digital delay element achieves a tuning range between 30ns – 95µs with a power consumption below 15fJ for a VDD voltage of 0.7 volts. Extrapolating the measured performance to system-level we obtain a dynamic power consumption inferior to 2µW for typical woman speech, and a silicon area below 1.5mm².

7.2 Research Perspectives

Since this work is the result of a large conjunction of topics, there are several perspectives open both for the subject as a whole and for each individual topic.

7.2.1 Short term perspectives

At a circuit level on wafer measurements to study mismatch. The objective is to create a wafer map to trace the evolution of mismatch according to wafer position.This measure will provide valuable design knowledge, for future implementations operating in weak inversion.

At a system level in a very short term the integration of the on-chip delay taps with a FPGA containing the DSP signal processing functions to demonstrate the functionality of the proposed concepts and the system as a whole. The proposed demonstrator scheme of the complete system is shown in Figure 127. The FPGA contains the digital functions corresponding to the ADC blocks, the cascaded filters and the energy estimation.

We propose a three step analysis in which fist a single filter function is demonstrated to show functionality of the blocks. In a second step the implementation of a single cascaded filter function, and finally the implementation of the complete system.

Additionally, taking advantage of the FPGA configurability an in-depth study of the alternative circuit to estimate the energy of a signal in CT, shown in Figure 51, and a comparison with the method currently proposed.

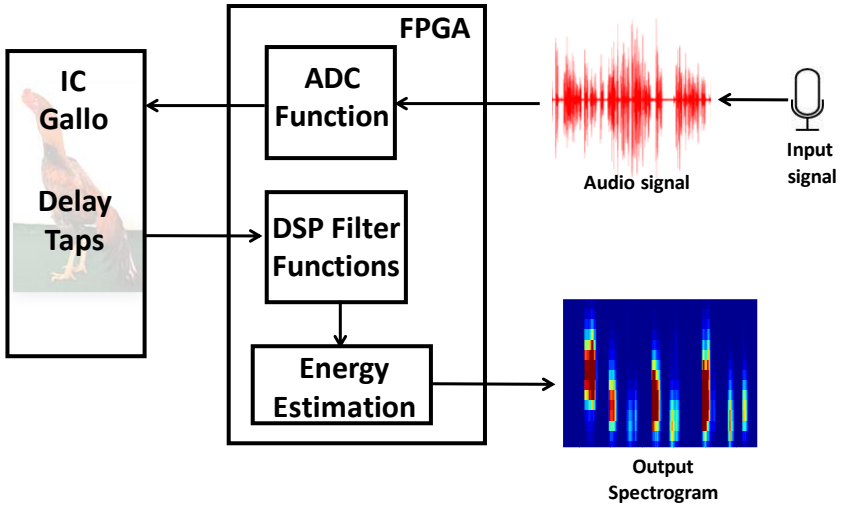


FIGURE 127 PROPOSED DEMONSTRATOR SCHEME

7.2.2 System architecture improvements

At the system level, one of the main benefits of the architecture to be exploited is the configurability offered by FIR filters. A recognition algorithm can be combined with the system, in order to modify or activate different filters for different scenarios. As described in [45], the system configurability can serve as knobs for the system, in order to always find the best compromise between power consumption and accuracy detection rate.

An example of a configurable FIR filter bank can be seen in [10], where a filter bank operates in three different modes. The number of filters varies in each mode 4, 6 or 8; the filters coefficient vary according to each mode and additional taps are activated or deactivated, to further adapt the frequency response of each filter, to each scenario.

Thus, we can envision the integration of our system with a detection algorithm configured to detect the operating environment. This detection system will then generate a signal to adjust the filter response accordingly. In our proposed system the frequency response of the filter bank can be modified by changing the coefficients, by modifying the control currents of the digital delay elements. Additional changes can also be made to add the option of activating or deactivating additional delay taps if required.

Another important block that must be studied and incorporated into the system is the circuit required to precisely tune the delays inside each glitch eliminator, and maintain their value constant despite the process voltage and temperature (PVT) variations.

7.2.3 Delay element improvements

The major issue concerning the delay elements is the generation of the very low biasing currents. Whose fluctuations add variability impacting the delay time. In the current implementation the biasing currents are externally generated and copied into the circuit trough current mirror schemes.

As shown during the measurement phase, the equipment employed for external generation of the biasing currents suffers important fluctuations and is impacted by parasitic components of the measurement equipment. Thus, two paths can be taken:

The first is to generate the biasing current references externally, using circuits like weighted current mirrors with a large division ratio (10-20) that would simplify the generation of the biasing currents. The second is the use of circuits like a bandgap reference to internally generate the bias currents for the delays, eliminating the disturbance added by external biasing sources.

An important consideration is that the circuits generating the biasing currents must have a large tuning range while operating in weak inversion. This poses an interesting challenge because most of the existing biasing circuits are built for transistors operating in strong inversion.

7.2.4 Alternative paths

In this work we proposed a filter bank that mimics the functioning of the human cochlea, for voice recognition applications using CT-DSP techniques. To respond to the challenge of implementing narrow transition bandwidth filters, we proposed a combination of the glitch eliminator and the delta encoder blocks, that allows efficient cascade of CT DSP FIR filters.

Since the proposed system mimics the functioning of the human cochlea, it can potentially serve as the digital processing unit for a cochlear implant. Thus it will be interesting to study the interfaces required to implement the proposed filter bank in a cochlear implant, like the one proposed in [57].

On the other hand, we recall that in previous implementations, asynchronous handshake protocols using tap delays are used to eliminate glitches in CT [24]. Thus, an interesting alternative is the use of dual-rail asynchronous protocols (in which every bit is encoded in to two bits) to implement the cascade of CT DSP FIR filters. Although dual-rail logic requires the use of larger circuits because each bit is now encoded into two bits [56], they eliminate the need for matched delays and the circuitry required to tune them. Another important benefit of dual-rail protocols is that signal processing speed is not limited by any clock or delay. This property could be particularly beneficial for CT systems targeting GHz range operation like [26].

Finally, the study of the CT system and the blocks proposed to cascade filters as part of the ultra-low power (ULP) CT receiver presented in [23]. Several analyses could be envisioned:

- The impact of delay tap mismatch used in the DSP unit on the BER (Bit-Error-Rate) and SNR, particularly for low ADC resolutions.
- The impact of glitch elimination techniques on the error vector magnitude (EVM).
- A comparison in terms of SNR, EVM between our proposed implementation and those using different asynchronous logic protocols.

7.3 A personal note on the future of CT systems

At the beginning of this work, one of the main objectives was to identify the bottlenecks to be overcome to efficiently implement continuous-time (CT) systems.

As described throughout my manuscript, CT systems offer great advantages, among which the potential of a larger SNR for the same ADC resolution than DT systems and event-driven adaptable power consumption. The question is then: how to motivate designers to learn and to use CT techniques?

The absence of a clock can be considered the main obstacle in the implementation of CT systems. The use of complex delay elements, not always easy to design or implement, becomes mandatory to process the samples generated by the ADC. Additionally, the matched delays required for the digital signal processing blocks, make it difficult for the implementation of the digital logic.

A major reason for a designer to choose a DT DSP implementation rather than its CT counterpart is automation. The DT circuit will be implemented using a high-level description and a bench of available CAD tools, when CT will necessitate a full-custom design, from the delay design and layout to the final system assembly.

To incentivize designers to the benefits, an important step will be to transform the delay elements and the delay taps, like those presented in this work, into IPs, available in common design libraries. Designers could then automatically create delay taps, optimized for a given frequency and ADC resolution, drastically reducing the burden of implementing a CT system.

For a long time, the analog-to-digital conversion race has been dominated by the race for the extra-decibel of SNR as described in [4]. Nevertheless, with the emergence of new IoT applications where the race is not dominated by SNR or maximum operating frequency, but by adaptability and reduced power consumption, CT systems could maybe find a niche and cement their place in the industry.

Bibliography

[1] C. Loius, “2018 Roundup Of Internet Of Things Forecasts And Market Estimates,” *Forbes Website*, 2018. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2018/12/13/2018-roundup-of-internet-of-things-forecasts-and-market-estimates/#23ccd43b7d83.

[2] C. Loius, “10 Charts That Will Challenge Your Perspective Of IoT’s Growth,” *Forbes Website*, 2018. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2018/06/06/10-charts-that-will-challenge-your-perspective-of-iots-growth/#613c511c3ecc%0Ahttps://www.forbes.com/sites/louiscolumbus/2018/06/06/10-charts-that-will-challenge-your-perspective-of-iots-growth/#280d69.

[3] Yahoo Finance, “NBA Players to receive coronavirus ‘smart ring.’” .

[4] B. Murmann, “The Race for the Extra Decibel: A Brief Review of Current ADC Performance Trajectories,” *IEEE Solid-State Circuits Mag.*, vol. 7, no. 3, pp. 58–66, Jun. 2015.

[5] M. Kocakulak and I. Butun, “An overview of Wireless Sensor Networks towards internet of things,” *2017 IEEE 7th Annu. Comput. Commun. Work. Conf. CCWC 2017*, pp. 1–6, 2017.

[6] H. Kuwabara, “Acoustic properties of phonemes in continuous speech for different speaking rate,” in *International Conference on Spoken Language Processing, ICSLP, Proceedings*, 1996, vol. 4, pp. 2435–2438.

[7] M. Verhelst and A. Bahai, “Where Analog Meets Digital: Analog?to?Information Conversion and beyond,” *IEEE Solid-State Circuits Mag.*, vol. 7, no. 3, pp. 67–80, Jun. 2015.

[8] J. Yoo, L. Yan, D. El-Damak, M. A. Bin Altaf, A. H. Shoeb, and A. P. Chandrakasan, “An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor,” *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 214–228, 2013.

[9] H. S. Wu, Z. Zhang, and M. C. Papaefthymiou, “A 13.8μW binaural dual-microphone digital ANSI S1.11 filter bank for hearing AIDS with zero-short-circuit-current logic in 65nm CMOS,” in *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, 2017, vol. 60, pp. 348–349.

[10] M. Yip, R. Jin, H. H. Nakajima, K. M. Stankovic, and A. P. Chandrakasan, “A fully-implantable cochlear implant SoC with piezoelectric middle-ear sensor and arbitrary waveform neural stimulation,” *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 214–229, 2015.

[11] K. M. H. Badami, S. Lauwereins, W. Meert, and M. Verhelst, “A 90 nm CMOS, 6 μW power-proportional acoustic sensing frontend for voice activity detection,” *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 291–302, 2016.

[12] M. Yang, C. Yeh, Y. Zhou, J. P. Cerqueira, A. A. Lazar, and M. Seok, “ISSCC 2018 / SESSION 21 / EXTENDING SILICON AND ITS APPLICATIONS / 21 . 2 A 1μW Voice Activity Detector Using Analog Feature,” *2018 IEEE Int. Solid -*

State Circuits Conf. -, vol. 3, pp. 346–348, 2018.

[13] S. Lecoq, J. Le Bellego, A. Gonzalez, B. Larras, and A. Frappe, “Low-complexity feature extraction unit for ‘Wake-on-Feature’ speech processing,” in *2018 25th IEEE International Conference on Electronics Circuits and Systems, ICECS 2018*, 2019, pp. 677–680.

[14] Gonzalez, Angel; Frappé, Antoine; Larras, Benoit; Kaiser Andreas; Cathelin, Philippe, “A Wide Tuning Range Delay Element for Event-Driven Processing of Low-Frequency Signals in 28nm FD-SOI CMOS,” IEEE *Solid-State Circuits Letters*, vol. Accepted, 2020.

[15] R. Sarpeshkar, “Analog versus digital: extrapolating from electronics to neurobiology.,” *Neural Comput.*, vol. 10, no. 7, pp. 1601–1638, 1998.

[16] Y. W. Li, K. L. Shepard, and Y. P. Tsividis, “A continuous-time programmable digital FIR filter,” *IEEE J. Solid-State Circuits*, vol. 41, no. 11, pp. 2512–2520, 2006.

[17] Y. Chen, X. Zhang, Y. Lian, R. Manohar, and Y. Tsividis, “A Continuous-Time Digital IIR Filter With Signal-Derived Timing and Fully Agile Power Consumption,” *IEEE J. Solid-State Circuits*, vol. 53, no. 2, pp. 418–430, 2018.

[18] Y. Tsividis, “Event-Driven Data Acquisition and Digital Signal Processing — A Tutorial,” *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 57, no. 8, pp. 577–581, 2010.

[19] H. Pan and A. A. Abidi, “Spectral spurs due to quantization in Nyquist ADCs,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, 2004.

[20] M. Kurchuk and Y. Tsividis, “Signal-dependent variable-resolution clockless A/D conversion with application to continuous-time digital signal processing,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 57, no. 5, pp. 982–991, 2010.

[21] H. Pan and A. A. Abidi, “Spectral spurs due to quantization in Nyquist ADCs,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, 2004.

[22] C. Weltin-Wu and Y. Tsividis, “An event-driven clockless level-crossing ADC with signal-dependent adaptive resolution,” *IEEE J. Solid-State Circuits*, vol. 48, no. 9, pp. 2180–2190, 2013.

[23] S. Patil, A. Ratiu, D. Morche, and Y. Tsividis, “A 3–10 fJ/conv-step Error-Shaping Alias-Free Continuous-Time ADC,” *IEEE J. Solid-State Circuits*, vol. 51, no. 4, pp. 908–918, 2016.

[24] B. Schell and Y. Tsividis, “A Continuous-Time ADC / DSP / DAC System Power Dissipation,” *IEEE J. Solid-State Circuits*, vol. 43, no. 11, pp. 2472–2481, 2008.

[25] J. B. O’Neal, “Delta Modulation Quantizing Noise Analytical and Computer Simulation Results for Gaussian and Television Input Signals,” *Bell Syst. Tech. J.*, vol. 45, no. 1, 1966.

[26] M. Kurchuk, C. Weltin-Wu, D. Morche, and Y. Tsividis, “Event-driven GHz-range continuous-time digital signal processor with activity-dependent power dissipation,” *IEEE J. Solid-State Circuits*, vol. 47, no. 9, pp. 2164–2173, 2012.

[27] Y. Li, D. Zhao, and W. A. Serdijn, “A sub-microwatt asynchronous level-crossing

ADC for biomedical applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 2, pp. 149–157, 2013.

[28] C. Weltin-Wu and Y. Tsividis, "An event-driven, alias-free ADC with signal-dependent resolution," *IEEE Symp. VLSI Circuits, Dig. Tech. Pap.*, pp. 28–29, 2012.

[29] Y. Tsividis, "Digital signal processing in continuous time: A possibility for avoiding aliasing and reducing quantization error," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2004, vol. 2, no. 3, pp. ii-589–92.

[30] A. Patki and G. Thiagarajan, "Low complexity, low latency resampling of asynchronously sampled signals," in *2016 International Conference on Signal Processing and Communications, SPCOM 2016*, 2016.

[31] "Applications of Speech Recognition," *Getsmarter*, 2019. [Online]. Available: https://www.getsmarter.com/blog/market-trends/applications-of-speech-recognition/.

[32] D. J. White, P. E. William, M. W. Hoffman, S. Balkir, and N. Schemm, "Analog sensing front-end system for harmonic signal classification," *ISCAS 2012 - 2012 IEEE Int. Symp. Circuits Syst.*, pp. 1155–1158, 2012.

[33] S. Lauwereins, W. Meert, J. Gemmeke, and M. Verhelst, "Ultra-low-power voice-activity-detector through context- and resource-cost-aware feature selection in decision trees," *IEEE Int. Work. Mach. Learn. Signal Process. MLSP*, pp. 0–5, 2014.

[34] P. C. Loizou, "Mimiking the Human Ear," *IEEE Signal Process. Mag.*, pp. 101–130, 1998.

[35] "Human Speech Spectrum, Frequency Range, Formants." [Online]. Available: http://www.bnoack.com/index.html?http&&&www.bnoack.com/audio/speech-level.html.

[36] John Ohala, "John Ohala, Emeritus Lab Director." [Online]. Available: http://linguistics.berkeley.edu/~ohala/.

[37] K. L. Anderson, "Speech Spectrum Audiogram for Consonants.Based on the work of Ross, M., Brackett, D., & Maxon, A.B (1991). Assessment and management of mainstreamed hearing-impaired children: Principles and practices." 2012.

[38] L. Goldstein, "General Phonetics Home Page," 2019. [Online]. Available: https://sail.usc.edu/~lgoldste/General_Phonetics/.

[39] E. Bernhardsson, "Language pitch," 2017. [Online]. Available: https://erikbern.com/2017/02/01/language-pitch.htmll.

[40] D. Byrne *et al.*, "An international comparison of long-term average speech spectra," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 2108–2120, 1994.

[41] National Institute on Deafness and Other Communication Disorders (NIDCD), "Journey of Sound to the Brain (video)," 2017. [Online]. Available: https://www.nidcd.nih.gov/health/journey-of-sound-video.

Bibliography

[42] A. Raychowdhury *et al.*, “A 2.3 nJ / Frame Voice Activity Detector-Based Audio Applications in 32-nm CMOS,” *IEEE J. Solid-State Circuits*, vol. 48, no. 8, pp. 1963–1969, 2013.

[43] M. Price, J. Glass, and A. P. Chandrakasan, “A Low-Power Speech Recognizer and Voice Activity Detector Using Deep Neural Networks,” *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 66–75, 2018.

[44] M. Yang, C. Chien, and S. Member, “A 0 . 5 V 55 μ W 64 × 2 Channel Binaural Silicon,” *IEEE J. Solid-State Circuits*, vol. 51, no. 11, pp. 2554–2569, 2016.

[45] J. H. Teo, S. Cheng, and M. Alioto, “Low-Energy Voice Activity Detection via Energy-Quality Scaling From Data Conversion to Machine Learning,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. PP, pp. 1–11, 2020.

[46] P. C. Loizou, M. Dorman, and Z. Tu, “On the number of channels needed to understand speech,” *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 2097–2103, 1999.

[47] “Transition band.” [Online]. Available: https://en.wikipedia.org/wiki/Transition_band.

[48] I. Sourikopoulos, “Continuous-time digital processing techniques applied to channel equalization for low-power millimiter wave communications,” Université de Lille, 2015.

[49] Y. Chen, R. Manohar, and Y. Tsividis, “Design of tunable digital delay cells,” *Proc. Cust. Integr. Circuits Conf.*, vol. 2017-April, pp. 1–4, 2017.

[50] B. Schell and Y. Tsividis, “A low power tunable delay element suitable for asynchronous delays of burst information,” in *IEEE Journal of Solid-State Circuits*, 2008, vol. 43, no. 5, pp. 1227–1234.

[51] C. Vezyrtzis, W. Jiang, S. M. Nowick, and Y. Tsividis, “A flexible, event-driven digital filter with frequency response independent of input sample rate,” *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2292–2304, 2014.

[52] G. Kim, M. K. Kim, B. S. Chang, and W. Kim, “A low-voltage, low-power CMOS delay element,” *IEEE J. Solid-State Circuits*, vol. 31, no. 7, pp. 966–971, Jul. 1996.

[53] I. Sourikopoulos, A. Frappé, A. Cathelin, L. Clavier, and A. Kaiser, “A digital delay line with coarse/fine tuning through gate/body biasing in 28nm FDSOI,” *Eur. Solid-State Circuits Conf.*, vol. 2016-Octob, pp. 145–148, 2016.

[54] M. Kurchuk and Y. Tsividis, “Energy-efficient asynchronous delay element with wide controllability,” in *ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems*, 2010, pp. 3837–3840.

[55] A. Cathelin, “RF/Analog and Mixed-Signal Design Techniques in FD-SOI Technology.” CICC Custom Integrated Circuits Conference Educational Session, 2017.

[56] J. Sparsø and S. Furber, “Principles of asynchronous circuit design – A systems perspective,” *Kluwer Acad. Publ. …*, p. 337, 2007.

[57] R. Sarpeshkar *et al.*, “An ultra-low-power programmable analog bionic ear

Bibliography

processor," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 711–727, 2005.