Design of a 8-bits 10Ms/s parallel pipeline Analog to Digital converter with shared amplifiers intended for Micro Electrode Array interface

Master’s degree in Micro and Nano technologies for Integrated Systems

Final Project report submitted in partial fulfilment of the Master’s degree

Student: _______________Dario Albino CARNELLI
Supervisors: ___________Neil Joye
_________________________Thomas Liechti
Professor: _______________Yusuf Leblebici
_________________________Dr. Alexandre Schmid

July 2008
Acknowledgments

I have always desired to introduce innovative ideas and to work on a large project about electronic systems, with the freedom to organize it myself. The Microelectronic Systems Laboratory at EPFL gave me the opportunity to realize this particular desire in a comfortable working environment where I found all that things which are essential to develop brilliant ideas and to significantly increase our own cultural luggage not only from a professional point of view but also from human and social aspects.

I really thank Prof. Yusuf Leblebici and Prof. Alexandre Schmid to have given to me the possibility to work on an interesting project but also to have put me in contact with the wonderful team of the LSM laboratory.

I’m grateful to my supervisors Neil Joye for his availability and kindness and Thomas Liechti whom have given me the guidelines for the project and some really helpful hints to design a possible innovative architecture.

Special thanks to Alessandro Cervero to helped me a lot especially for the patience he put to explain to me how the different design kits works. I need to say that from my point of view he was like a “third supervisor”.

I cannot forget all my colleagues of the Nanotech-Master with whom I spent unbelievable and unforgettable days, especially Luigi di Lillo, Lattanzio Livio and Salvatore Fisichella.
Abstract

A parallel-pipelined A/D converter with an area efficient architecture is described. By sharing amplifiers along the pipeline, an 8-b pipeline is realized using just six amplifiers (instead of twelve amplifiers with a conventional pipeline architecture). By using two such pipelines in parallel, a 10 Msamples/s prototype A/D converter that is intended for a Multi Electrode Array (MEA) neurochip has been designed in a 0.18-µm CMOS technology.

Un convertisseur A/D parallèle-pipelined conçu en utilisant une architecture minimisant la surface est décrit dans ce rapport. En partageant les amplificateurs le long du pipeline, un pipeline de 8-bits est réalisé avec seulement six amplificateurs contrairement à douze amplificateurs utilisés dans une architecture pipeline conventionnelle. En utilisant deux de ces pipelines en parallèle, un 10’000 Echantillons/s prototype d’un convertisseur A/D a été élaboré en technologie CMOS UMC 0.18 µm. Ce prototype sera utilisé dans une Matrices de Micro Electrodes (MEAs).

Nel presente progetto di tesi sarà discusso il funzionamento di un convertitore parallelo A/D a 8 bit la cui tecnica di condivisione dell’amplificatore operazionale permette di ottenere un sistema dalle ridotte dimensioni. Utilizzando solamente sei amplificatori (al posto di dodici per una struttura pipeline convenzionale) è stato progettato un convertitore funzionante alla velocità di 10 Milioni di campioni al secondo. Tale sistema, progettato in tecnologia UMC 0.18 µm è impiegato per la conversione di segnali acquisiti tramite neurochip a matrice di microelettrodi (MEAs).
To my mother
# TABLE OF CONTENTS

1 **INTRODUCTION** .............................................................................................. 1
  1.1 **INTEREST IN NEURONAL SIGNAL ANALYSIS** ........................................... 1
  1.2 **NEURONAL SIGNALS ACQUISITION TECHNIQUES** ................................. 2
  1.3 **MULTI ELECTRODE ARRAY (MEA)** ........................................................ 3
  1.4 **STATE OF THE ART OF MEAS** ............................................................... 4

2 **HIGH RESOLUTION MEA FOR INTERFACING WITH NEURAL CELL CULTURES [4]** ................................................................................................... 7
  2.1 **PROJECT PRESENTATION** ........................................................................ 7
  2.2 **SPECIFICATIONS** ..................................................................................... 8
  2.3 **ARCHITECTURE CHOICE** ......................................................................... 9

3 **8-BIT PIPELINE ADC** ..................................................................................... 13
  3.1 **WORKING PRINCIPLE** ........................................................................... 13
  3.2 **ERROR CORRECTION WITH 1.5-BIT SUB-CELLS** ................................. 15
  3.3 **GLOBAL FLOORPLAN** ............................................................................. 16
  3.4 **1.5-BIT STAGE WORKING PRINCIPLE** .................................................. 18
    3.4.1 **Shared amplifier technique** ................................................................. 20
    3.4.2 **Coding scheme, input voltage and supply** .......................................... 21
    3.4.3 **Proposed Operational Transconductance Amplifier** ........................... 22
    3.4.4 **Proposed Comparator** ..................................................................... 28
    3.4.5 **Proposed Switches and Analog Multiplexers** .................................... 31
    3.4.6 **Proposed Delay lines** ..................................................................... 34
  3.5 **DIGITAL PROCESSING** ............................................................................. 37
    3.5.1 **Control signals** ............................................................................... 38
    3.5.2 **Delay chains and final adder** ............................................................ 41

4 **CONCLUSION** ................................................................................................. 45

REFERENCES .......................................................................................................... 47
Chapter 1

Introduction

1.1 Interest in neuronal signal analysis

Information coming from the outside world is processed in our brain by hundreds of billions of cells called neurons. These neurons are electrically excitable cells that process and transmit information in the brain and nervous system. Thus they play a key role for the whole organism. For example, results from neuroscience research confirm the stupefying computational power of our brain. Complex tasks like face recognition are carried out in real time using minimal power, thus beating any available digital computer. Brain and as consequence neurons have a considerable importance for life. The study of neurons is therefore a key factor for understanding the operation of our brain.

Furthermore, diseases based on malfunction of neural cells are among the most serious health problems. The actual knowledge is wide concerning the working principle of diseases at the single neuron level like sclerosis. However, diseases that involve the complexity of neuron communication such as Alzheimer’s disease, Parkinson’s disease, motor neurons disease, and Huntington’s disease, are still largely unknown. Currently only the symptoms of these diseases can be treated.

As a consequence a deeper study of the electrical properties of biological cells and tissues is needed. Using Multi Electrode Arrays (MEAs), it is possible for populations of neural cells to be examined simultaneously, thereby providing better insight into the functionality and interconnectivity of cellular networks. In this scope, a MEA with CMOS circuit which performs signal processing on-chip is being designed in the Laboratory of Microelectronic Systems (LSM) in EPFL. Thus, the scope of this thesis is to design an analog to digital converter used in the signal processing circuitry.

The first chapter of this thesis introduces different neural signal acquisition systems, MEAs, and the state of the Art of MEAs. The second chapter presents the project topic, its specification, and the architecture that has been selected for the analog to digital converter. The third chapter explains in more details the circuit that has been designed. The simulation results are described in the same chapter. Finally, conclusions are taken in the last chapter.
1.2 Neuronal signals acquisition techniques

To understand how information is processed in a neural network, researchers may need to understand how an action potential propagates from one neuron to another without external interference. In order to accomplish this task, tools and models which have been developed to study intact neurons are used.

Some of these tools are used to analyze distributed phenomenon on large areas covering the whole brain surface. Even if these techniques are reliable techniques, they are not suitable for the purpose of this work. For local analysis, classical electrophysiology techniques are based on the use of electrodes. These electrodes can be:

- Simple solid conductors (discs, needles)
- Tracks on Printed Circuit Boards (PCB)
- Hollow tubes filled with an electrolyte

All these techniques can be applied with different kind of preparations:

- Living cells;
- Excised tissues (acute or cultured);
- Cells dissociated from excised tissues (acute or cultured);
- Artificially grown cells or tissues.

Different possibilities are available in order to observe the electrical activity of cells, according to noise level, invasive or non invasive analysis, number of cells monitored for each electrode, etc….

The use of large electrode discs is the least invasive technique but the most affected by noise. Thus it is only possible to have indirect activity recording of several nearby cells simultaneously (multi unit recording) Hence small disc electrodes have been used to have single unit recording. The noise level is still too high and good bidirectional communication is nowadays a challenge.

Small sized sharp tip electrodes are used to successfully avoid noise and resolution problems with the drawback of reduced lifetime. Intracellular recording in these cases involves measuring a voltage and/or current across the membrane performing the so called “voltage clamp” (or “current clamp”) measurements that consist in holding the voltage (or current) level to a fixed value and measure the current (or voltage) variations.

Finally, the finest technique for neuron electrical activity recording is the patch-clamp one. Patch-clamp consists in a microscopic hollow pipette, filled with an electrolyte, which is pressed against the cell membrane.

In case of “whole cell recording”, fluid continuity is achieved by mean of a pressure pulse through the electrolyte across the membrane whereas in case of “patch recording”, the membrane may be left intact. In both cases a chlorided silver wire is inserted in the electrolyte and propagates the signal to the signal processing.
electronics. Drawback of “whole cell recording” technique is that the intracellular fluid mixes with the solution inside the pipette so that some components of the intracellular matrix can be diluted. This inconvenient could be avoided with the “perforated patch” technique in which the pipette is withdrawn from the cell pulling the “patch” (part of membrane in contact with the hollow tip of the pipette) away from the rest of the cell.

1.3 Multi Electrode Array (MEA)

The study of neural cellular functionality and electrical behaviour is considered as a fundament and prerequisite to better understand the operation of the brain. As technology has improved, neuroscientists have shown increased interest in the in-vitro experimentation methodology to support advanced studies of the brain and the operation of its constituting neurons. Using Multi Electrode Arrays (MEAs), it is now possible for populations of neural cells to be examined simultaneously, thereby providing better insight into the functionality and interconnectivity of cellular networks.

The use of CMOS electronics overcomes the drawbacks of traditional MEAs. Standard and commercial MEAs usually consist of a passive MEA chip with external signal processing circuitry and a system control that is realized by discrete off-chip components. However, each individual electrode needs to be wired to external electronics, which limits the size of the array. These MEAs usually contain approximately 60 electrodes with electrode sizes ranging from 10 to 30 µm, and inter-electrode spacing up to 100 µm. These dimensions are much larger than the 10 µm typical size of vertebrate neurons used during electrophysiological experiments. A photograph of a MEA setup during electrophysiological experiments is shown in Fig. 1

![Fig. 1 Multi Electrode Array setup](image)

However recently, a new generation of CMOS-based MEAs containing a high-density of sensors has emerged, as explained in Section 1.3. These new devices
INTRODUCTION

consist of an electrode array pos-processed on top of a CMOS circuit which performs data processing such as amplification, addressing, and analog to digital conversion. A conceptual schematic of a CMOS-based MEA is described in Fig. 2.

![Conceptual cross-section of a CMOS based MEA.](image)

1.4 State of the art of MEAs

Classical systems for electrophysiological experiments incorporate a MEA with external signal conditioning electronics and system control realized by discrete off-chip components. Each electrode requires a connection to external electronics, which attenuates the weak electric signals and limits the array size. Therefore, several groups have designed on-chip microelectronics MEAs with recording and stimulating capabilities. Two different approaches have been developed, the first one based on CMOS MEA with metal electrodes and the second one based on CMOS MEA with Electrolyte-Oxide-Semiconductor Field Effect Transistors (EOSFET).

CMOS MEA based on the EOSFET principle has been mostly developed by the group of P. Fromherz, MPI Munich [1]. The idea is to sense the extracellular potential of a neural cell by connecting it to the gate of an EOSFET. This principle has been implemented in a 128 x 128 CMOS MEA [2], which is at the moment the MEA with the largest number of electrodes. Each electrode has a size of 4.5 x 4.5 µm² and a pitch of 7.8 µm. Therefore this MEA is also the one with the highest resolution up to date. However, due to the very low capacitive current flowing through the electrodes, MEAs based on the EOSFET principle show some major drawbacks. First of all, stimulation of an electrical activity in neural cells is for the moment not feasible using EOS transistors. Secondly, in order to have a good recording quality of action potentials, only large cells from invertebrate animals such as the leech have been measured.

Up to date, most of the other groups have developed CMOS MEAs with metal electrodes. The operating principle is based on sensing the extracellular potential with a metal electrode and to amplify the signal with the appropriate on-chip circuitry. Biocompatible materials such as gold (Au), platinum (Pt), palladium (Pd), titanium nitride (TiN), iridium oxide (IrO₂) or others have been used to produce the electrodes. Considering CMOS MEAs based on metal electrodes, the most complete system has
been developed by the group of A. Hierlemann in ETH [3]. An 11’000 electrode CMOS MEA with a pitch of 18 µm has been manufactured using CMOS 0.6 µm technology. This MEA is able to stimulate and record the electrical activity of neural cells. Therefore, it is a very attractive system for electrophysiological experiments.
Chapter 2

High resolution MEA for interfacing with neural cell cultures [4]

2.1 Project presentation

An innovative three-dimensional tip electrode array technology for in-vitro electrophysiological experiments has been developed in the Laboratory of Microelectronic Systems (LSM) at EPFL. The goal of this MEA is to record and stimulate the neural activity of cell cultures with high spatial resolution to demonstrate also the superiority of three-dimensional shaped electrodes in terms of electrical coupling.

A first generation of MEA (Fig. 3) has already been manufactured in the Center of Micro Nano Technology (CMI) of EPFL, it consist of tip electrodes made of a layer of platinum (Pt) deposited on top of silicon dioxide ($\text{SiO}_2$) cones.

Three-dimensional MEAs with an electrode diameter of 3-4 $\mu$m, a height of 1.75 $\mu$m, and a pitch dimension of 5-6 $\mu$m have been manufactured. Due to the high spatial resolution of these three-dimensional MEAs, the electrode size is very small.
compared to state of the Art MEAs (see Section 1.4). Thus the electrode impedance is increased and generates a higher input noise which affects the SNR (Signal to Noise Ratio). Therefore using an off-chip readout circuitry results in more challenging signal acquisition compare to on-chip readout circuitry. This disadvantage of off-chip signal acquisition is mainly due to the long wires that carry the signal from the electrode area to the instruments which are additional source of noise (antenna effect, crosstalk, increased thermal noise, etc…).

Therefore, the idea is to build an on-chip signal processing circuitry using 0.18 \( \mu \)m CMOS technology. The three-dimensional electrodes are then expected to be post-processed on top of the signal processing circuitry, as depicted in Fig. 4. The main blocks of the signal acquisition system are

- Low noise amplifiers directly and closely connected to the electrodes.
- Analog to Digital converters in order to converter the amplified recorded signals to digital signals which could be read by an external acquisition system such as a personal computer.
- Addressing circuitry in order to scan selected electrodes.

![Fig. 4 Conceptual cross-section of a CMOS based MEA with three-dimensional electrodes post-processed on top of a CMOS signal acquisition circuitry.](image)

2.2 Specifications

As explained in Section 2.1, a CMOS based MEA is being developed in the Laboratory of Microelectronic Systems (LSM). The goal of this MEA is to record and stimulate the neural activity of cell cultures with high spatial resolution. Here are some of the main specifications that have to be reached by the MEA:

- 100 x 100 electrodes.
- Electrode pitch of 4-6 \( \mu \)m.
- Full frame readout rate of 20 Kfps.

The goal of this project is to design 8-bit A/D converters that will convert the amplified measured signals to digital signals that can be used by the off-chip data acquisition system.

The main Idea is to place one or more ADCs just in the space adjacent to the electrode matrix. Different solutions can be adopted either to use different stripe
shaped ADC on both two opposite sides of the MEA structure or to place two or more bigger square ADCs that are working with an higher frequency clock (Fig. 5).

![Diagram](image)

Fig. 5 Number of ADC and clock frequency needed to read 20kfps

2.3 Architecture choice

In literature, different architectures have been successfully implemented in order to obtain an Analog to Digital conversion. Some ADCs are used for a very high precision and low sample rate such as: Dual slope ADC; Single slope ADC (or integrating ADC); Digital ramp ADC (works in a similar way of the Integrating ADC) and Tracking ADC. Other ADCs are used for high speed such as: Folded ADC; Pipelined ADC and Full-Flash ADC. In this paragraph these architectures are analyzed in order to evaluate the working frequency and highlight strong and weak points of each one for the required application.

**Dual slope:** for an 8-bit resolution, a maximum of \(2 \cdot 2^8 = 512\) clock cycles are needed. If 20 Ks/s have to be processed (from only one electrode), a working frequency of 10.240MHz (512 x 20kHz) is needed in the case where two adjacent samples have the highest opposite values, which is the worst case. In terms of area it requires two operational amplifier and digital controlling circuitry.

**Single Slope:** similar to dual slope but in this case the integrator doesn’t need to ramp up to get the signal value. It is a faster approach; however there are a lot of drawbacks such as:

- Requires a precise timing with a quartz oscillator
- Calibration drift
- Not equal time space for conversion (signal dependent)

**Digital ramp:** uses the same working principle of the single slope ADC with the use of a DAC. In terms of speed is the same as Dual slope. It is smaller than the Single Slope and it has less drawbacks.

**Tracking:** the working principle of this kind of ADC consists in ramping up until the signal level has been reached and then follow the signal time by time. The standard tracking architecture is slower than other circuits discussed above to produce the output. However it is less noisy, because if the signal doesn’t change, the circuit will stay in the same state. It is very small in terms of area and it is a simple and flexible architecture also because it allows the ADC to go faster by increasing the area consumption. Using the standard architecture, a 10.240MHz (256 x 2 x 20kHz) frequency is needed in order to follow a reference square wave of 20kHz in a
reasonable way. Generally it is better to use this architecture when the signal that has to be converted changes smoothly and rarely takes a constant value for long time.

**Successive Approximation:** The only change in this design with respect to the digital ramp ADC's is a special counter circuit that is called *successive-approximation register*. It doesn’t count in a binary sequence, this register counts by trying all values of bits starting with the most-significant bit and finishing at the least-significant bit. Throughout the count process, the register is connected to the comparator's output to see if the binary count is less than or greater than the analog signal input, adjusting the bit values accordingly. The advantage to this counting strategy is much faster results: only 8 clock cycles are needed to determine the output.

![Successive Approximation ADC (input and DAC output)](image)

Fig. 6 shows what happens to the output of the D/A converter output when a sinusoidal signal of $2MHz$ is processed and the clock cycle is $64MHz$, this is the case if 100 electrodes are read from only one of this ADCs. Thus it can be deduced that the use of such ADC is suitable when the input signal takes constant values for the entire period. This is the case if a sample and hold circuit is used in order to hold the input voltage value.

A good advantage of this architecture is that it could be arranged in a striped way [5] and so used as matrix frontend (Fig. 7). Documentation reports that SAR-ADC is used for high accuracy (10bit – 12bit) and low speed (100kS/s – 5MS/s) applications [5] [6] [7].

![Shape of a Successive Approximation A/D converter](image)

Fig. 7 shape of a Successive Approximation A/D converter [6]
All the architectures discussed above are mainly for high resolution and low speed. Before discussing about high speed ADC, some key points have to be highlighted:

a) It is not possible to fit an ADC within the space among electrodes because of the short space, routing problems, and induced noise from the controlling signals.

b) Assuming that is possible to efficiently route signals over the four sides and to use all the sides, it is required to place 396 ADCs with a width of 5µm, each of which is controlling 26 electrodes. Thus the limit is to place in a so small space an ADC working at \(20kHz \times 26 = 520kS/s\). The only possible solution could be shrinking a SAR-ADC (always using the smallest clock frequency that is 4.160MHz because to convert only one value 8 clock cycles are needed).

c) In the case this is possible, still a huge routing problem has to be solved. Therefore the power dissipation will increase and huge noise will affect the signals because of the large amount of wires which transit electrical signals,

**Flash:** This architecture is used for ultra fast conversion speed (up to 2GS/s). It is the easiest and most commonly used architecture for fast A/D converters. Unfortunately is the most area expensive because of the large amount of comparators that it needs, hence is never used for high accuracy conversion (at most 10 bits, but usually 8 or less).

**Folding:** with this technique a very fast conversion is achievable in a smaller space than full flash solution. It is possible to obtain an 8 bit resolution with only 30 comparators by just folding the signal before a Flash ADC.

Each incoming sample is processed in one clock cycle at the expense of power consumption that is needed to extend the bandwidth of the comparators (anyway it is possible to reduce the power when sampling rate and analog input bandwidth are fixed). With a matrix of 10'000 electrodes, two or more of these ADCs could be adequate, according to the area availability, in order to be sure that all the signals will be processed in time at 100MS/s as explained in [8].

Some drawbacks are anyway present:

- Higher repetition rate of folding results in a rounding-off of the tips of the folded signal.
- It may be needed to put the flash ADC system far away from the electrodes area to avoid excessive noise due to the high frequency clock cycle.
- A folding ADC working at 100Ms/s with 8 bit output is the state of the Art for folding ADCs and may result in a not so easy design. A way to avoid this problem is to find a good compromise between area and clock speed to determine how many parallel system have to be placed in order to lower the clock, use narrower band comparators and subsystems, ease the system design.
*Subranging or Pipelined:* This architecture is the one that can best fit all the specifications and offer a lot of advantages with respect to all the other architectures.

First of all it has a natural striped structure that allows a good routing. It has also a modular structure such that it is possible to increase resolution only by adding blocks (but for pipelined structure with more than 8 bits, temperature and area [9] could be a problem). It is also possible to build a reconfigurable ADC structure as reported in [10] and obtain more precise/less precise, fast or slow acquisition according to user preferences.

This structure is well documented, easy to design and analyze, and with the SAR structure [11] is the most commonly used with matrix sensor structures [12]. There are already documented papers about pipelined ADC used with a width of 40-50µm. For the purpose of this project, it is needed to control 10 rows of electrodes from each side hence 500 electrodes. This means that a speed of 10MS/s must be achievable for this structures [12, 13].

With respect to the Folded ADC, this architecture is more area expensive1, but can run with a drastically reduced speed with a clock frequency of 10MHz. To have a better idea, a table of developed architecture is presented:

<table>
<thead>
<tr>
<th>Ref/Year</th>
<th>Feature</th>
<th>Technology</th>
<th>Figure of merit</th>
<th>Area(mm²)</th>
<th>Power/Voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td>JSSC 12/1999</td>
<td>Pipelined delta - sigma</td>
<td>1,2µm CMOS</td>
<td>SNR 74dB - 18Ms/s</td>
<td>48</td>
<td>324mW/3.3V</td>
</tr>
<tr>
<td>JSSC 12/1997</td>
<td>Pipelined delta - sigma</td>
<td>0,6µm CMOS</td>
<td>16 bits – 20Ms/s</td>
<td>35</td>
<td>550mW/5V</td>
</tr>
<tr>
<td>JSSC 12/1998</td>
<td>Self calibrating</td>
<td>0,5µm CMOS</td>
<td>12 bits – 10Ms/s</td>
<td>15</td>
<td>335mW/3.3V</td>
</tr>
<tr>
<td>JSSC 12/1996</td>
<td>Microcontroller calibrating</td>
<td>1µm BiCMOS</td>
<td>16bits – 1Ms/s</td>
<td>35</td>
<td>200mW/5V</td>
</tr>
<tr>
<td>ISSCC98</td>
<td>Analog calibrating</td>
<td>1µm CMOS</td>
<td>10bits – 40Ms/s</td>
<td>47</td>
<td>650mW/5V</td>
</tr>
<tr>
<td>ISSCC98</td>
<td>2 channels</td>
<td>0,5µm CMOS</td>
<td>8bits – 75Ms/s</td>
<td>5.5</td>
<td>70mW/3.3V</td>
</tr>
<tr>
<td>ISSCC98</td>
<td>2 channels</td>
<td>1µm CMOS</td>
<td>10bits – 40Ms/s</td>
<td>42</td>
<td>565mW/5V</td>
</tr>
<tr>
<td>ISCAS98</td>
<td>Current mode</td>
<td>0,5µm CMOS</td>
<td>8bits – 20Ms/s</td>
<td>N/A</td>
<td>22mW/2.4V</td>
</tr>
<tr>
<td>CAS II99</td>
<td>Good linearity</td>
<td>0,5µm CMOS</td>
<td>12bits – 3.3Ms/s</td>
<td>N/A</td>
<td>300mW/5V</td>
</tr>
</tbody>
</table>

Tab. 1 already developed architectures and figures of merit

---

1 A pipelined ADC could have a length from 1mm to 1.5mm [12] resulting in an area consumption on one edge of 0.495mm² to 0.742mm² while for a folded ADC an area consumption of 0.28mm² is reported [8].
Chapter 3

8-bit Pipeline ADC

3.1 Working principle

A general architecture for a pipeline ADC consists of a cascade of identical stages [14] and of a digital correction and aligner processor (Fig. 8) that produces the final output.

In each sub-converter stage there is an n-bits flash ADC unit, that used to convert directly the incoming signal; an n-bits DAC unit, that is used to reconvert the output produced by the flash ADC in a quantized analog signal, and an amplifier which purpose is to amplify and send to the next stage the residue signal produced by the subtraction of the input signal with the quantized one produced by the DAC (in Fig. 9 is shown a block diagram of a single stage ).

![Fig. 8 Pipeline or Subranging converter architecture](image)

![Fig. 9 sub-converter architecture](image)
Basically each stage produces a coarse approximation of the input signal and sends the quantization error to the next stage for further processing. Assuming that a voltage $v_{in}$ is sent to this cell, and that a 1 bit cell is used, the DAC converter will output either the highest possible analog value ($V_{ref}$) or the lowest possible ($-V_{ref}/2$ in case of a symmetric range) according to the input signal if it crosses or not the zero voltage (hence the MSB will be produced by the first stage), this reference voltage is then subtracted to the incoming signal and amplified by a factor of 2:

$$v_{out} = 2v_{in} + V_{ref} \text{ if } v_{in} < 0V,$$
$$v_{out} = 2v_{in} - V_{ref} \text{ if } v_{in} > 0V,$$

The next stage takes this signal as input and performs the same operations; the result is shown in Fig. 10:

![Diagram](image)

Fig. 10 Residue output of the first 1-bit stage (left) and the second 1-bit stage (right)

Each time an abrupt transition occurs on the output waveform there will be a change in the logic value produced by the stage. Observing in Fig. 10 it is possible to clearly notice that the first cell decides whether the signal is positive or negative (MSB) then the second stage splits each of those big regions (positive and negative) in two smaller sub-region and decides whether the signal is on the left sub-region or the right one and so on, producing directly the entire conversion starting by the MSB.

The problem with these 1 bit cells is linked to the precision of the zero crossing comparator, if an offset is present and the value of the signal is in between 0 and the offset value, all the next stages will saturate producing an erroneous conversion. A better approach is to use 1.5-bit sub-converters.
3.2 Error correction with 1.5-bit sub-cells

A 1.5-bit cell is used mainly to have the great advantage of the digital correction of errors due to not precise decision thresholds or offset error, avoiding the design of a high precision, area consuming comparators, more space is hence left for the delay optimization.

To explain how the error correction is carried out, a 2-bit sub cell will be used.

![Fig. 11 output of a 2-bit stage](image1)
![Fig. 12 reduced interstage gain](image2)

In Fig. 11 the input-output characteristic of a 2-bit stage is presented, the gray square represent the limits of the input range (vertical sides) and the limit of the next stage input range (horizontal ones). In this case, if an error on the threshold occurs (red line) the second stage will interpret this result as an over-range value that makes the successive stages saturating. This means that no room is left for errors in sub-ADC or sub-DAC.

Even if an offset is present, the output will exceed the range bounded by $\pm V_{ref}$ (red line in Fig. 11), this will saturate the second stage and cause missing information. To avoid this problems two solution can be proposed:

1. Increase the range of the second stage
2. Reduce the inter-stage gain of the first ADC to tolerate sub-ADC errors.

In case a reduced inter-stage gain is used, the transfer function is illustrated in Fig. 13. In this particular case the gain has been reduced by two but it can be reduced less. In case the gain is reduced less a more complex digital processing is needed. If an error occurs, due to a threshold shifting in the first stage, this error will be recognized from the second stage since he output produced by the first one is still in the input range of the next one.

Moreover, whatever input is higher than $V_{ref}/2$ is immediately recognized from the stage like a right threshold shifting of the previous stage and will produce a logic
value that will cause the increase of one LSB to the first stage output during the
digital correction and, whatever input is lower than $-V_{\text{ref}}/2$ is recognized from the
next stage like a left threshold shifting of the previous stage and will produce a logic
value that will cause the subtraction of one LSB during the error correction.

Both addition and subtraction operation are needed in this case, increasing the
complexity of the correction logic and making more complicate the code assignment
for the pipeline stages.

This problem can be avoided by intentionally add a $-V_{\text{ref}}/4$ offset to the input of
the sub ADC and to the output of the sub DAC, this way the transfer function will be
modified like in Fig. 13.

Since overranging can be dete cted from the sec ond stage, the architecture of the
stages can be simplified avoiding the offset introduction before the ADC and after the
DAC just eliminating the comparator with threshold $3V_{\text{ref}}/4$ hence obtaining the
already discussed 1.5-bit cell.

Hence only two comparators are needed: one of them with input reference $V_{\text{ref}}/4$
and the other one with input reference $-V_{\text{ref}}/4$.

### 3.3 Global floorplan

In this paragraph a complete global schematic view of the ADC will be presented
focusing in a detailed way on the choice that have been made to make the entire
system as compact and reliable as possible.
The whole circuit consists of two parallel pipelined ADCs whose analog part is represented by the chain of six 1.5-bit doubled sub converters with shared amplifier, each one providing only one effective bit to each of the two ADCs; and of a shared 2-bits full-flash ADC tail. The resolution of the entire structure is hence 8 bits per ADC.

The parallel pipeline works like it is two independent ADCs that are working with opposite phase in a non-overlapping mode. Within the analog part is included a delay line in each stage, in order to produce locally three delayed copies of the clock, avoiding problems of huge and complex routing of signals, hence avoiding possible delays due to parasitic capacitances among wires and parasitic resistance due to long path metals necessary for the clock distribution. An entire metal layer (metal 4) is supposed to be used to uniformly distribute the clock all over the ADCs area.

The analog part occupies the central area of the two parallel pipeline ADCs because this way a shortest routing is obtained, and the digital part is designed to occupy a very narrow strip on both sides of the analog part.

A smart distribution of the clock feeds twelve delay lines (two for each sub converter): six of them with an opposite phase clock. Then, the three delayed copies of the clock with the clock itself are sent to the control logic of the digital part that synthesizes directly the switches driving signals and the latching signals to send to the comparators. The output of each stage is sent back to the digital part and it is used to compute the final result and also to produce the driving signals for the analog multiplexers (in Fig. 16 a mixed mode simulation shows the clock signals and the driving signals).

Each adjacent stage is working with opposite phases in a checkerboard fashion, so, instead to distribute the clock in a checkerboard way each successive stage has its own control logic that uses the incoming clock signal to produce either the “In phase” driving signals or the “opposite phase” driving signals, therefore the opposite phase clock will be sent only to one of the two ADCs.

Because clock regeneration is often needed some buffers will be inserted where the clock signal is split in two path, a smart way to distribute the opposite phase clock is hence to substitute the buffer with an inverter where it is needed.
3.4 1.5-bit stage working principle

As previously mentioned, a 1.5 bit sub-converter has to perform basically four operations:

1. Sample the incoming signal at the desired frequency and hold the value for the entire period
2. Make a 1.5 bit conversion of the sampled signal
3. Reconvert the produced digital output in a quantized signal
4. Subtract the input signal with the quantized one and amplify to obtain at the output the residue signal.

In Fig. 18 a classic implementation of this sub-system is shown. This system uses a 2 level analog to digital converter with reference levels $+V_{ref}/4$ and $-V_{ref}/4$. The
digital to analog converter is represented simply by a three inputs analog MUX selecting from three different reference voltages: \(-V_{ref}, 0\) and \(+V_{ref}\).

While the circuit is in the sampling mode the switches \(S_1\) and \(S_2\) are connected to the input node while the switch \(S_3\) is closed to ground allowing the incoming signal to charge the capacitors \(C_f\) and \(C_s\).

Then, in the second phase, all the switches will change their position so: switch \(S_3\) will disconnect the connection to ground; switch \(S_2\) will close the capacitor \(C_f\) in a feedback loop with the amplifier and switch \(S_1\) will connect the output of the analog multiplexer to the input of capacitor \(C_s\).

According to the output provided by the two comparators (that has being latched since the switches have changed their state) the analog multiplexer will provide \(-V_{ref}, 0\) and \(+V_{ref}\). In details, when the output code is 00 the multiplexer will output \(-V_{ref}\); when the output code is 01 the multiplexer will output 0 and when the output code is 11 the \(+V_{ref}\) voltage will pass through the analog multiplexer. Depending on the data from the Digital-to-Analog conversion the following operations are performed by the circuit:

\[
v_{\text{out}}(v_{\text{in}}) = \begin{cases} 
  \left(1 + \frac{C_s}{C_f}\right) \cdot v_{\text{in}} - V_{ref} & \text{if } v_{\text{in}} > \frac{V_{ref}}{4} \\
  \left(1 + \frac{C_s}{C_f}\right) \cdot v_{\text{in}} & \text{if } -\frac{V_{ref}}{4} < v_{\text{in}} < \frac{V_{ref}}{4} \\
  \left(1 + \frac{C_s}{C_f}\right) \cdot v_{\text{in}} + V_{ref} & \text{if } v_{\text{in}} < -\frac{V_{ref}}{4}
\end{cases}
\]

In all the cases the amplifier is connected like in Fig. 19.

The input value has been stored into two equally sized capacitors \((C_f\) and \(C_s\)), these capacitors are then connected in the shown configuration and the quantized analog signal \(V_{ref}\) is applied to the input. When the input is connected, because the OTA’s inverting input is at a virtual ground potential, there will be a voltage

---

8-BIT PIPELINE ADC
8-BIT PIPELINE ADC

$V_{\text{ref}} - v_{\text{in}}$ across the capacitor $C_s$, this voltage drop means that in the capacitor there will be a change of charge of $\Delta q_s = C_s (V_{\text{ref}} - v_{\text{in}})$. This voltage step make the charge changing on this capacitor but the additional charge is taken from the capacitor $C_f$ because no charge can flow outside or inside the inverting input node (it is isolated) hence $\Delta q_s = -\Delta q_f$ and because: $\Delta q_f = C_f (v_{\text{out}} - v_{\text{in}})$

$$C_s (V_{\text{ref}} - v_{\text{in}}) = -C_f (v_{\text{out}} - v_{\text{in}}) \Rightarrow v_{\text{out}} = \frac{C_s}{C_f} v_{\text{in}} + v_{\text{in}} - V_{\text{ref}}$$

If the two capacitors are equally sized then: $C_s = C_f$ and the ratio is 1, as a consequence:

$$v_{\text{out}} = 2v_{\text{in}} - V_{\text{ref}}$$

From here can be deduced that a very good capacitor matching is needed in order to obtain the correct output voltage and this is more important in the first stages. The use of an OTA implies as well high area and power consumption; hence a technique to reduce the number of used OTAs is needed.

3.4.1 Shared amplifier technique

As stated before, a classic sample/hold circuit can be used to store the signal in two equally sized capacitors, the time it takes to charge them is then a function of the capacitor size and of the switches resistance and internal resistance of the source (in case the input is the first one) or of the driving capability of the OTA (in case the input is provided by a previous stage), so the most of the time is spent to charge the capacitors. In the meanwhile the capacitors are charging the flash ADC can operate by following the signal that is being stored. When the capacitors are charged it only needs to wait for the comparators’ propagation time to start the same processing again inside the next stage. While this is happening the amplifier of the 2nd stage is not used, and instead it can by another Pipelined ADC.

![Fig. 20 Two adjacent stages of the realized architecture](image-url)
In Fig. 20 a schematic view of the actual implementation of the 1.5-bits stages is presented; on the upper part of the figure there are two adjacent 1.5-bit stages belonging to a first ADC, and in the lower one there are two adjacent 1.5-bit stages belonging to a second ADC. All the four stages are identical and symmetric but the upper ADC and the lower one are working with opposite phases (like two adjacent stages of the same ADC) in a checkerboard fashion, to allow sharing an amplifier between two stages belonging to different ADCs.

![Fig. 21 global view of the analog signal processing](image)

Sharing the amplifier this way implies that data is going to be processed on both edges of the clock making the architecture a parallel configuration of two independent, time interleaved ADCs.

### 3.4.2 Coding scheme, input voltage and supply

Before to proceed with a description of the 1.5 bit sub-cell basic blocks some considerations have to be carried out.

First of all, the project deals with 0.18µm double well technology with low voltage supply (1.8Volts). As a consequence, whatever signal is applied to the ADC input (that is coming from an amplifier) is in between 0V and 1.8V. Hence, for convenience, it has been chosen to set the ground voltage in the middle of this range: $V_{GND} = 900m\text{V}$ (all the quantities shown in the next chapters are referred to this ground voltage).

As a convention it is assumed that a constant signal of $V_{GND}$ voltage will be converted with the code “1000000”; all the codes from “00000000” to “01111111” (128 possible codes) will represent negative inputs (below $V_{GND}$) and all the codes from “10000001” to “11111111” (127 possible codes) will represent positive inputs (more than $V_{GND}$).
3.4.3 Proposed Operational Transconductance Amplifier

3.4.3.1 Design

The OTA, as discussed above, is the most critical component, together with the capacitor matching for the first stage; it is the most area consuming component and should be designed in order to drive two large capacitors in the required time (huge slew rate), maintaining its area and power consumption as smallest as possible. Moreover it has to provide an output signal that is linear with the input voltage and has the same swing (the maximum possible in order to reduce the conversion error), this means that $V_{ds, sat}$ of the last stage transistors have to be very small (this means to lower the gate voltage as much as possible maintaining an high value of the current hence bigger size).

The implemented topology is the folded cascode OTA with self-biased high-swing cascode current mirrors, shown in Fig. 22. With this configuration a high gain can be achieved because of the high output resistance provided by the cascode current mirror, and moreover, for the self compensation that makes the amplifier less sensible with respect to the power supply variations (a high power supply rejection ratio PSRR is essential for the future use of this amplifier).

At this point the design of the OTA starts from a trade-off solution among different issues involving the resolution of the ADC (LSB), the power consumption, and the area occupied by the capacitors. In details:

1. Capacitors have to be sized in order to have a $kT/C$ noise under a certain percentage of the quantization noise power;
2. To know the LSB it is needed to know the full scale voltage range that have to be chosen as high as possible;
3. The full-scale voltage range depends on the maximum linear output swing that can be achieved by the output stage of the OTA;
4. The output swing is limited by the saturation voltage of transistors M5, M7, M9, M11 (Fig. 22); it can be reduced but has to be more than $4U_T$ (to operate in strong inversion);
5. Then, the sizes of these transistors have to be chosen according to the current needed to drive the capacitors (slew rate).

A good compromise has been obtained with these values:

$$V_{DS,sat} = 196mV \cong 7U_T; \quad V_{FullScale} = 1.020V;$$
$$v_{out,max} = 508mV; \quad LSB = 4mV; \quad C_s = C_f = \frac{K_BT}{V_{noise,RMS}^2} = 107fF.$$ 
$$v_{out,min} = -512mV; \quad v_{noise,RMS} = 5\%LSB = 200\mu V;$$

Values for $v_{out,max}$ and $v_{out,min}$ are chosen according also to the convention explained in section 3.4.2, thus they are not symmetric because of the missing code. With the values obtained before it is possible to find the current of the OTA’s output stage by computing the Slew Rate.

As stated at the very end of section 2.3 the minimum clock speed for 10 ADCs per side is 10MHz (to meet exactly the frame-rate speed specification of 20kfps) but for safety purpose the design has been carried out with a clock speed of 15MHz, with period of 66.67ns. Within this period the amplifier is used twice because of the amplifier sharing and the capacitors have to be charged before a semi period is passed (accounting for the switching activity).

If the output has to swing from the lowest possible voltage to the highest one (or vice versa) in the required time (less than 40% of the period), the value of the Slew rate can be carried out:

$$SR_{MIN} = \frac{V_{FullScale}}{26.4ns} = \frac{1.020V}{26.4ns} = 38.636\mu A$$

From which:

$$I_3 > 10.2\mu A$$

From this value on, all the transistors of the operational amplifier have been extracted; in addition, the two resistors have been substituted with two switch connected n-MOS in an always closed state and the biasing stage has been shared with more than one amplifier in order to save area. In Tab. 2 is presented the sizing of the designed amplifier.

<table>
<thead>
<tr>
<th>Transistors</th>
<th>Width (µm)</th>
<th>Length(µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1, M2</td>
<td>20</td>
<td>0.36</td>
</tr>
<tr>
<td>M4, M5, M6, M7, M13, M14</td>
<td>10.75</td>
<td>0.36</td>
</tr>
<tr>
<td>M8, M9, M10, M11</td>
<td>1.04</td>
<td>0.5</td>
</tr>
<tr>
<td>R_A ⇒ M_A</td>
<td>0.24</td>
<td>1</td>
</tr>
<tr>
<td>R_B ⇒ M_B</td>
<td>1.38</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Tab. 2 OTA sizing
Transistor M3 is not present in the table because it is used to provide the biasing current to the differential pair and since the reference voltage that has to be used to bias all the differential pairs is still unknown, it is still not possible to size it.

It is necessary to point out the fact that the differential pair is constituted of two n-MOS transistors that are working in weak inversion. This means that an high gain is obtained but the bandwidth of the amplifier is sacrificed. Since the slew rate aspect dominates and this specification is satisfied, also the bandwidth requirements are satisfied.

3.4.3.2 Experimental results

The proposed architecture for the operational amplifier has been designed and tested with Cadence Virtuoso® Front to Back Design Environment. The realized layout view of this amplifier is reported in Fig. 23. It consists of three main blocks: The first one on the left side is the differential pair, the one in the middle is the upper part of the output stage (p-MOS mirrors and common gate stage) while the right part is the n-MOS active load of the output stage. Each block is protected by a guard ring from interference of the neighbour circuits and dummy devices have been used to obtain a high grade of symmetry both from the physical and electrical point of view.

This device is the most area consuming component, its area is 33.080μm x 16.120μm. Simulations that have been performed on this structure deal mainly with its speed and linearity.

In Fig. 24 is presented the simulation result for the slew rate of the amplifier before parasitic extraction. Marker M0 is the differential measurement fot the slew rate; Markers M2 and M7 point the peak overshoot (respectively for a low to high and high to low transition); Markers M3 and M8 point to the settling time (respectively for a low to high and high to low transition). It has been chosen to insert in input a square wave signal with 1.020V amplitude, which is the maximum input swing for the incoming signal and a frequency of 15MHz (blue line). By connecting the
amplifier in a buffered configuration with a load of 200fF, the output is observed (Red line). From the graph obtained the slew rate has been extracted.

The actual value for the Slew Rate is $65.459 \frac{\mu V}{\mu s}$, for a low to high transition, that is higher than the required value. For a high to low transition the Slew Rate is $-53.09 \frac{\mu V}{\mu s}$ that is still higher than the required value. Always before parasitic extraction the settling time is $19.83\,ns$ and $24.89\,ns$ (respectively rising and falling transition), that is enough to ensure that the capacitors are charged in half a period. The main reason for the difference of the slew rate is because the path of the current trough the output stage of the amplifier is different while on the rising or on the falling edge of the device. While in the first case (rising edge) the current is sourced from the p-MOS transistors that are bigger (so more current will pass trough tem in less time) in the other case the current is sink to the Vss node through the n-MOS devices that are smaller and cannot drive the current like the upper p-MOS.

In Fig. 25 is presented the result of a post layout simulation of the Slew Rate. The post layout values for the positive and negative Slew Rate are respectively
55.730/\mu and 49.327/\mu. As expected these values are lower than the pre-extraction values but they are still not critical for the purpose the amplifier has been designed.

The Open loop AC simulation before parasitic extraction is reported in Fig. 26.

From this simulation has been extracted the gain value of the amplifier, the frequency of the dominant pole and the phase margin:

- Open loop gain: 48.05 dB
- Dominant pole frequency: 476kHz
- Phase margin: 66.42°

The phase margin is more than 60° this guarantees a good stability of the amplifier.
After parasitic extraction (Fig. 27) these values are worsen but still good enough:

- Open loop gain: 47.52 dB
- Dominant pole frequency: 491kHz
- Phase margin: 64.8°

For what concern the phase margin for the stability of the circuit it has to be checked at 6dB of the open loop gain bode plot, but if at 0dB the phase margin is already good enough the stability check at 6dB can be avoided.

From the closed loop AC simulation (Fig. 28) after parasitic extraction instead, the graph shows which is the error that the amplifier makes at the working frequency. By intersecting the curve at 15 MHz the resulting gain is -433.9mV which means that the introduced relative error is -4.872%.

Other extracted figures of merit for the presented amplifier are shown in Tab. 3.

<table>
<thead>
<tr>
<th>Figure of merit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMRR</td>
<td>107.299dB</td>
</tr>
<tr>
<td>PSRR+</td>
<td>87.76dB</td>
</tr>
<tr>
<td>PSRR-</td>
<td>96.4dB</td>
</tr>
<tr>
<td>Input Offset</td>
<td>-653.9μV</td>
</tr>
<tr>
<td>Power consumption</td>
<td>80.478μW</td>
</tr>
</tbody>
</table>

Tab. 3 OTA figures of merit
3.4.4 Proposed Comparator

3.4.4.1 Design

Comparators are the second most area consuming component, they don’t need to be very precise because of the error correction processing but they have to be fast to produce the output in small fractions of half a period.

Among all the possible architectures there are some that allows latching the output signal like the one adopted for the purpose of this project showed in Fig. 29. It consists of four stages: the differential pair, used as a preamplification stage; the latching stage; the power stage and the output stage [15].

![Four stages latched comparator schematic](image)

The first stage is used to simply amplify the differential signal and contribute also to the overall gain of the comparator (by increasing the sizes of M1 and M2 an higher gain can be obtained).

The second stage is called decision stage, this stage must be capable of discriminating mV level signals; in order to have a faster decision a feedback system has been adopted. Assuming that a differential signal is applied at the input of the comparator, an unbalanced current will flow in the two differential branches \( I_1 \neq I_2 \Rightarrow I_3 \neq I_4 \); in case the Latch signal is low the M12 transistor is switched on and diodes M11 and M10 are working and a voltage will appear between node A and B. Supposing that current \( I_4 \) is higher than current \( I_3 \) then \( v_{BA} > v_{AB} \) and because the gate of M8 is connected to node B its drain current is higher than the drain current of M9, this means that M8 will take current from the diode M11, further lowering the voltage \( v_{AB} \) that drives the gate of M9 that will drain less current than before increasing further the voltage \( v_{BA} \). This way the voltage \( v_{BA} \) keep increasing more and more in a very fast way. From here it is possible to deduce that this circuit is in a dynamic equilibrium and that a slightly different current can produce a really high voltage difference between node A and B immediately.
This circuit has to be extremely symmetrical because it could be affected by hysteresis problems; hence a particular layout design has been conceived to achieve this specification and will be presented in section 3.4.4.2.

The across node A and B is sent to a power stage that is a fully Complementary Self-Biased CMOS Differential Amplifier (CSDA).

In the CSDA, devices M14 and M19 operate in the linear region. Consequently, the voltages on their drain may be set very close to the supply voltages. Since these two voltages determine the output swing of the amplifier, the output swing can be very close to the difference between the two supply rails. This large output swing makes interfacing the CSDA to ordinary CMOS logic gates straightforward, since it provides a large margin for variations in the logic threshold of the gates [16].

Another consequence of the linear-region operation of devices M3 and M4 is that the CSDA can provide output switching currents that are significantly greater than its quiescent current. This capability of supplying momentarily large current pulses makes the CSDA especially suitable for high-speed comparator applications, where it is necessary to rapidly charge and discharge output capacitive loads without at the same time consuming inordinate amounts of power [16].

Finally, the output stage consists of simple inverters that have the purpose to increase further the gain and make the comparator more precise.

<table>
<thead>
<tr>
<th>Transistors</th>
<th>Width (µm)</th>
<th>Length(µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1, M2</td>
<td>5</td>
<td>0.3</td>
</tr>
<tr>
<td>M4, M5, M6, M7</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>M8, M9, M10, M11</td>
<td>0.24</td>
<td>0.7</td>
</tr>
<tr>
<td>M12, M13</td>
<td>1</td>
<td>0.36</td>
</tr>
<tr>
<td>From M14 to M23</td>
<td>0.24</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Tab. 4 Comparators sizing

3.4.4.2 Experimental results

The proposed architecture for the comparators has been designed and tested with Cadence Virtuoso® Front to Back Design Environment. Comparators layout required a careful design of the latching stage because of the hysteresis problem. Therefore because of the high matching required between the left and right branches of the decision stage, transistors M8, M9, M10 and M11 have been split each one in two series connected devices and dummies have been added. A cross shaped structure has been chosen to have central symmetry and to avoid mismatches due to the different doping caused by the different crystal orientation of the silicon substrate.

The resulting layout is shown in Fig. 30. The above discussed decision stage is placed in the upper right part of the figure and can be recognized from the particular geometry of the design, the n-MOS differential pair is placed in the bottom left part and consists of a classical interdigitated structure. The p-MOS mirrors are placed on the upper left part. All this blocks are enclosed with protective guard ring to avoid external devices’ interference. The Output stage and the output buffer are placed on the bottom right corner of the figure, this part consists of small sized transistors and it does not need a guard ring because the signals inside this block is quite robust and the
effect of noise is limited. The area consumption for this component is 33,080μm x 16,120μm.

Several simulations have been carried out, all of them show good results. The most important characteristic is generally the input-output voltage curve (Fig. 31). As stated in section 3.2 because of the error correction the input offset is not a huge problem, anyway the presented architecture has a really reduced offset of -42,72μV.

Surprisingly, a post layout simulation shows not only a really reduced hysteresis but also a further reduced input offset voltage

![Fig. 30 Comparator Layout](image.png)

![Fig. 31 Input Output characteristics before (Red) and after (Blue) parasitic extraction](image.png)

The most important figure of merit for the realized comparator is the propagation delay. In Fig. 32 a transient analysis is reported: the red signal is the signal applied to the non inverting input of the comparator, while the inverting input is connected to ground. To have a propagation delay measurement a square wave with frequency of 15MHz has been applied. A propagation delay of 1.86ns and 1.39ns (rising and falling edge respectively) have been measured without parasitic extraction and a
propagation delay of 2.45ns and 2.39ns (rising and falling edge respectively) have been measured with parasitic extraction.

![Fig. 32 Transient response before (Green) and after (Purple) parasitic extraction](image)

### 3.4.5 Proposed Switches and Analog Multiplexers

#### 3.4.5.1 Design

All the switches presented in this work are basically CMOS transmission gates, designed in order to reduce the dependence of their resistance with respect to the input signal.

A CMOS transmission gate consist of a parallel connection of an n-MOS and a p-MOS transistors where the n-MOS gate is connected directly to the driving signal and the p-MOS gate is connected to the inverted driving signal as shown in Fig. 33.

![Fig. 33 a) CMOS transmission gate b) bidirectional switch c) analog multiplexer](image)

With the transmission gate (Fig. 33 a), the bidirectional switch (or 2 to 1 analog multiplexer) have been designed by connecting the outputs of two switches and cross connecting the gates (Fig. 33 b) and the analog multiplexer has been designed just by connecting the output of three switch (Fig. 33) . The big issues with the use of the CMOS pass gate are the nonlinear on resistance and the charge feed-trough and charge injection effects.

The nonlinear on resistance happens because the mobility of carriers is different in the n channel device and in the p channel device: when the input voltage is low the n-
channel dominates while when the input is high the p-channel dominates. And if the mobility of carriers are different also the resistances will be different. This problem can be avoided by a proper sizing of the two transistors.

The clock feedthrouch instead takes place when the analog switch turns on and off, it happens that a small amount of charge can be capacitively coupled (injected) from the digital control line to the analog signal path. This effect is more important if the switch is used for a sample and hold circuit because it can introduce or drain charge from the capacitor, increasing or lowering the voltage across it.

The charge injection is due to the fact that a small charge is located inside the switch channel so that when a clock transition occurs part of this charge will flow in one side and part on the other side according to the node impedance that is present on each side.

If the switch size is too big then this effect become more important and can introduce a non negligible error.

3.4.5.2 Experimental results

The proposed architecture for all the analog multiplexers has been designed and tested with Cadence Virtuoso® Front to Back Design Environment.

In Fig. 34 to Fig. 36 three different implementations for the analog multiplexers are presented. Each of them integrates both the switches and the inverters needed to drive the p-MOS device with the inverted signal hence it is needed only one wire to rive each switch.
Resistivity simulation have been carried out to check the linearity of the ON switch resistance with respect to the input voltage, and to check if the OFF resistance is high enough to be considered an open circuit for the given timescale and for the capacitor sizes.

![DC Response](image)

Fig. 37 I/V characteristics (Left) and resistances (Right) before (Red and Green curves) and after (Purple and Blue curves) parasitic extraction.

A sweep analysis of the output voltage has been performed and the results for the pre-parasitic extraction and post parasitic extraction are shown in Fig. 37. It can be noticed that the ON resistance is almost constant at 5KΩ resistance with the varying input voltage (Green curve). The OFF resistance is even much more stable and constant (its value is 500GΩ for the entire input voltage range).

However post layout simulations show that these switches are largely affected by parasitics. A first effect is that the OFF resistance even if symmetric is drastically reduced of an order of magnitude but it can still be considered like on open circuit. The very bad result concern the ON resistance that has a huge variation in the p-MOS dominant side (positive input voltages).
3.4.6 Proposed Delay lines

3.4.6.1 Design

A delay element is a circuit that produces an output waveform similar to its input waveform, only delayed by a certain amount of time. In the present master’s project it is used to generate several delayed copies of the clock signal in order to correctly drive the switches of the 1.5-bit sub converters. Since The amplifier that has been designed is very fast and can drive the load capacitors from the lowest value to the highest one in less than 20ns it is not required for the delay line to produce a very precise delay for the clock copies. Anyway it has been designed to generate tree clock copies delayed of 1ns one to each other.

Several architectures have been proposed in literature, they belong mainly to three different categories:

1. Transmission Gate based;
2. Cascaded inverter based;
3. Voltage controlled based;

The transmission gate delay element delay is provided essentially by the delay that the signal needs to charge the load capacitance through the switch resistance; the working principle is that one of an RC circuit, with the advantage of a really reduced area and power consumption. However the main drawback of this kind of delay element is the signal integrity that is defined as the transition time for the output to go from 10% to 90% of the drain voltage.

Because of this an improved version has been studied with a cascaded Schmitt trigger that is a circuit that generates a fast, clean output signal from a slowly or noisy varying input signal (Fig. 38 a)

![Fig. 38 Element with Shmitt trigger (a) Cascaded inverter based (b) Voltage controlled based]  

Due to the use of positive feedback, the signal integrity remains virtually unchanged as the delay value increases.

The cascaded inverter delay element is simply a cascade of inverter stages, where the total delay is obtained by summing all the gate’s propagation delay. The main
One of the most interesting implementation is the m-Transistor cascaded inverter, in this architecture the resistance of the n-MOS and the p-MOS are increased in order to slow down the current that flows from $V_{DD}$ and $V_{SS}$ node to the output of each stage through all the devices. Actually this is the main Idea that is behind the third family of delay elements that is the Voltage controlled one. In Fig. 38 c) an np-voltage controlled cascaded delay element is presented, it employs two controlling voltages: $V_p$ and $V_n$, in order to increase respectively the resistivity of the upper p-channel MOS devices, and the resistivity of the lower n-channel MOS devices in order to reduce both the currents that are sourced by the $V_{DD}$ node and drained to the $V_{DD}$ node. This circuit is the one that provides the best signal integrity but it has the largest power consumption.

The Implemented architecture for the delay element uses a working principle that includes all the three techniques explained before. It is basically a transmission gate delay element with 2 cascaded inverters, moreover the gates of the switch is connected to the reference voltages ($127mV$ and $-128mV$ ) that will be generated from the reference generator, hence the delay will be quite independent from the variation of the supply voltages and from variations of the temperature. The major advantage of this kind of connection is the fact that is possible to have the same resistance of a switch connected directly to the voltage supply with a considerably smaller area. The schematic view of the implemented topology is presented in Fig. 39.
3.4.6.2 Experimental results

The proposed architecture for the delay lines has been designed and tested with Cadence Virtuoso® Front to Back Design Environment.

Results from transient analysis after a post layout extraction are reported in Fig. 41. An input square wave of 15MHz with an excursion range of 1.8V is applied to the input of the delay line (light blue curve) and the output, which is connected to an inverter as load, is read at each different output points of the same delay line (red,
green and purple curves). It is shown that the delay is almost constant between all the delayed copies and has a value of 1.33\(\text{ns}\) for the rising edge and of 1.78\(\text{ns}\) for the falling edge, this means that the layout design doesn’t influence significantly on the delay line performances.

### 3.5 Digital processing

Digital processing consists of controller units (to drive all the switches of the analog part), shift registers (to align the output produced by each stage) and a ripple carry adder for the error correction. For long pipelines within the registers is often included a code converter in order to reduce the number of registers, but for an 8 bit ADC, with 0.18\(\mu\text{m}\) technology, the area required to make an intermediate conversion is higher than the space required for a D flip-flop unit, hence it has been decided to ease the design and make it compact by simply using a final ripple carry adder.

Unlike the analog part that has shared components between two parallel ADC, the digital part for each parallel ADC works independently from the other one and with opposite clock phase. The two ADCs hence will output at different clock edges, in details, the upper ADC will give an output at the rising clock edge and the lower ADC will produce the output at the falling clock edge, moreover each ADC is composed by opposite phase adjacent stage, this means that the Flip flop register chains have to work with opposite phases and the final adder has to sum them all together in the same time. In the next paragraphs a detailed description will be presented on how the digital part has been analyzed and synthesized and on how the issues explained above have been solved.
3.5.1 Control signals

To describe how to generate control signals for all the switches in the circuit, a temporal sequence of switch configurations belonging to four adjacent stages will be shown. The analysis start from the case in which one of the first two stages is in sampling configuration and the other one is in amplifying configuration, each case will be named according to the state of the two first stages belonging to the upper ADC.

**TIME (1):** in Fig. 42 the first configuration has been shown: the upper left sub-converter is sampling the incoming signal while the lower left is in the amplifying configuration, hence is using the amplifier and it is sending the output to the lower right sub-converter that is in sampling configuration.

While the circuit is in this state the amplifier load consist of the sum of the sampling capacitor and of the input capacitance of four comparators, this means that if an high speed application was needed, It was needed to take into account also this last contribution because the input stage of the comparators has a non negligible input capacitance. This capacitance has to be taken into account even if the value of the sampling capacitor would be small enough to be compared to it.

When this configuration occurs it requires the most of the time to be completed because as stated before in section 3.4.3, when the amplifier is used to load the huge output capacitors it needs enough time to charge them up to the correct voltage value.

**TIME (2):** Once the Sampling-Amplifying configuration is finished, hence when the capacitors are charged and the propagation time of the comparators have been waited, the latching signal can be sent and the sampling capacitors $C_S$ and $C_F$ can be disconnected completely and isolated from the circuit in order to hold the voltage value across them. This situation is described in Fig. 43. It can be noticed that the amplifier is left in amplifying configuration but the load consist only of the input
capacitance of the four comparators but it has no more influence on the sample capacitors.

Fig. 43 Time 2: Disconnected-Amplifying configuration

**TIME (3):** When the capacitors have been disconnected the configuration of the previously amplifying stages can start changing releasing the input of the amplifier and connecting one node to ground like shown in Fig. 44.

Fig. 44 Time 3: Disconnected-Load Connected configuration

It happens that the $C_s$ capacitor will be charged with the reference voltage provided by the Analog multiplexer and the $C_f$ capacitor will load the amplifier of its stage and will charge with an undefined voltage. This situation should be held as short as possible in order to avoid huge changes on the voltages across the capacitors, mainly
because with high probability they are going to be charged with a new voltage value that is close to the previous one, so that if the remaining value is not changed too much they can reach the final value in a shorter time.

**TIME (4):** After the Input of the amplifier has been released a new connection can be established with the capacitors of the upper left sub converter that were storing the new value that has to be processed. In the meanwhile, the capacitors of the lower left sub comparator are going to be connected together with the effect of a redistribution of charge between the two capacitors; this will reduce the error generated in the previous configuration.

Anyway, the connection between one node of the capacitors and the input of the amplifier is a dangerous connection because of two reasons. The first reason is the fact that some charge can be lost because of the switching activity and the original voltage value stored can change; the second one is that the amplifier is changing his input to a different voltage level, this means that the output could be different from the previous configuration and continue affecting the capacitors of the next stage probably increasing the time required to charge again the sample capacitors.

![Fig. 45 Time 4: Floating capacitors-Sampling configuration](image)

**TIME (5):** At this point simply by connecting the switches $S_1$ and $S_2$ to the output of the amplifier and to the output of the amplifier the configuration is the opposite of the first one (Amplifying-Sampling configuration). From now on the circuit will pass from the following configurations:

1. **TIME(6):** Amplifying-disconnected (opposite of TIME(2))
2. **TIME(7):** Load connected-disconnected (opposite of TIME(3))
3. **TIME(8):** Sampling-floating capacitors (opposite of TIME(4))
4. **TIME(9) = TIME(1):** Sampling-Amplifying

This is what happens in one clock period. These requirements have been successfully obtained by using four copies of the clock signal delayed of 1ns one each other and synthesizing directly from these copies the switches’ controlling signals.
3.5.2 Delay chains and final adder

The outcoming signal has to be processed in order to obtain a final digital corrected output code in 8 bits per ADC. The major task that the digital correction circuit has to solve is to align the code produced from the Analog part. As the analog signal is passing through the chain of sub converter it is being converted. While a sample is being evaluated in a stage of the chain, the next sample is being evaluated in the previous one and the previous sample is being evaluated in the next stage, therefore the output produced from each stage has to be stored for the amount of time is necessary to have a complete conversion and this storing time is different and depends on the position of the stage within the chain.

Another issue has to be solved. The output produced by the sub converter stages of each ADC consist of 12 bits per channel plus 2 last bits produced from the full flash final unit. While each stage is producing 2 bits in a clock period on dedicate channels, the last two bits are sent in a shared channel so that the data has to be read twice in the same time (both on rising and falling edge). Moreover each adjacent stage of each ADC is producing the output with different phases.

A general view of the global timing is presented in the following time versus space diagram:

![Tab. 5 Time versus space diagram](image)
The diagram is divided vertically in two sections: the left one represent the upper ADC and the right one represent the lower ADC; on the top side the clock is represented with associated the numbers of the clock period. Each row represents a stage in the circuit and its associated output bits. The red shaded area means that the ADC is in an unknown initial state. Because the Pipeline ADC gives the most significant bits at the first stage the voltage values to converted are coming from the down side of the table (Black arrows identify the path of the signals) and are sampled in the specified time. To describe this situation the cell of the corresponding stage that is sampling has been marked with an “S”, highlighted with a light blue colour and a number is assigned to identify the sample that is being processed (i.e. “1” means that the first voltage value in temporal order applied at the input of the corresponding ADC is going to be processed). After the sampling, the output can be read and appears on the corresponding channels on the rising edge of the clock (Green divider) or on the falling edge (Blue divider).

Next step is the amplification to the next stage, which has been marked with an “A” letter, highlighted with a yellow colour and the number that identifies the processed value. The table shows how the pipeline works on both the ADC in a specified time in order to allow the amplifier sharing. Actually the two sides of the table can be overlapped to clearly see that there are no conflicts among signals.

The pipeline ends with a Full Flash ADC that is used by both ADCs (The box of the Full Flash is shaded with a purple colour on both the sides to describe this situation). Since its input has been changed, it produces the output just after its propagation delay. To illustrate this situation a yellow divider has been used and it is placed a bit later, after the amplifier of the previous stage has sent the signal. As we stated before this ADC is shared so it can be used from the upper ADC or the lower one (this case is highlighted green and the letters mean “U” upper ADC and “L” lower ADC and the number of the sample).

From this graph has been obtained the literal representation of the aligned bits. For only one ADC this representation is:

\[ b_{13,1,R}, b_{12,1,R}, b_{11,1,F}, b_{10,1,F}, b_{9,2,R}, b_{8,2,R}, b_{7,2,F}, \]
\[ b_{6,2,F}, b_{5,3,R}, b_{4,3,R}, b_{3,3,F}, b_{2,3,F}, b_{1,3,3,F+\Delta t}, b_{0,3,3,F+\Delta t} \]

From this representation it is easy to find the number of D flip-flop cells needed on each register and which of them have to be driven with an opposite phase clock. The last bits that are produced are those of the full flash ADC just few nanoseconds after those of the last sub-comparator cell, hence the data has to be ready just a bit after the third falling edge of the clock. The resulting register structure contains 18 D flip-flop six of them are driven by an opposite phase clock.

Assuming that all the bits corresponding to the code of a processed input voltage sample are aligned, the error correction operation that has to be performed is:

\( F + \Delta t \) subscript means that it is needed to wait the propagation delay of the Full Flash ADC
This operation can be done by a simple ripple carry adder connected to the output of the shift registers. The outline scheme of the realized structure is shown in Fig. 46.

The realized digital output processing has been tested. In Fig. 47a test bench simulation is shown: an input signal “1011010111011” is applied and the expected result (9) is reached after three clock cycles as expected.
Chapter 3

Conclusion

From the results presented in the previous chapter it can be deduced that each block has good characteristics (except for switches) to be integrated in the final structure. A floorplan of the layout and a complete schematic view have been obtained, so that it has been possible to run a simulation of the complete ADC in order to check if the main requirements are met.

In Fig. 48 the result of a mixed mode simulation is shown. This plot shows the outputs of the first three stages when a ramp is applied to the ADC channels. It confirms that the entire circuit (both analog and digital parts) is operating in a proper way.

![Mixed mode simulation](image)  
**Fig. 48 Mixed mode simulation: Outputs of the first three stages when a ramp signal is applied on both channels**
But some issue due to timing are present and it can be seen from the output of the second and the third stage: just the semi-period after the comparator of the first stage changes its output the second stage will miss the correct value. This error is most probably due to timing issues and has to be analyzed and corrected like the results that have been obtained for the switch resistance. Another mixed signal simulation has been carried out to check if also the addition operation is performed in order to obtain the final code. These simulation shows that the output code is not stable. Hence because of these issues, it is still not possible to evaluate some important figure of merit of the ADC (SNDR, SFDR, INL/DNL).

Another component that has not been treated in details is the full flash final ADC. For the moment it has been designed with the same comparators that have been used for the stages comparators and uses a ladder of high resistance diffusion resistors and the terminals of the ladder are connected to the voltage references. This kind of ADC has to be revisited and treated in more details because maybe it could be possible to not use a ladder resistance that consumes a lot in terms of area and power, and because the actual prototype has been designed just to test the functionality of the whole circuit.

Reference generators have not been implemented yet and generally some more detailed simulations still have to be carried out, like Montecarlo simulations and the noise influence on the circuit. Anyway a detailed global floorplan has been developed, all the blocks have been studied and simulations show that further optimization can be done and that a much more performing converter can be obtained by revisiting some critical components. For instance if the design of the comparator could be eased thanks to the error correction it is not needed a really precise comparator so it can be substituted with a smaller one, saving area for other components. Or slightly changing the switching timing it is possible to use less clock phases saving area.

Moreover, a preliminary place and route showed that the digital part is more compact than expected and, because it is arranged in a striped way on both sides of the ADC, the place and route leaves some free spaces, so that other digital functionalities can be further thought to occupy those free spaces.
REFERENCES


