Design of an Asynchronous Flash Analog-to-Digital Converter

Gonçalo Pinto Monteiro

STATE OF ART

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Abstract

The proliferation of Internet of Things and biomedical sensors has created a high demand for low-power signal processing interfaces. In many of these applications, the signals of interest are sparse, characterized by long periods of inactivity. Classical synchronous Analog-to-Digital Converters are inherently inefficient in such scenarios, as they consume dynamic power continuously due to the global clock, regardless of the input signal activity.

This dissertation proposes the design and implementation of an Asynchronous Flash ADC. By removing the clock, the proposed architecture aligns power consumption with the input signal activity, theoretically achieving a much lower power consuption. However, the removal of the clock introduces design challenges (acrescentar aqui problemas)

To address this, an offline trimming strategy is proposed to calibrate the comparator offsets without compromising the high-speed operation of the flash topology. The work encompasses the theoretical analysis, schematic design, and validation of the system through Analog/Mixed-Signal co-simulation. The expected outcome is a robust, clockless ADC architecture that offers a superior Figure of Merit for sparse signal applications in compared to conventional synchronous architectures.

Keywords: Analog-to-Digital Converter, Asynchronous Design, Flash ADC, Level-Crossing Sampling, Offset Trimming, Low Power, IoT.

Contents

List of Figures

List of Tables

List of Acronyms

Chapter 1

Introduction

This chapter provides a brief introduction to the topic associated with the work carried out during the dissertation period. Firstly, a contextualization is given, followed by a presentation of the research question and the main motivations for carrying out this study. The main objectives of this dissertation will also be outlined, as well as the methodology used to achieve them. Finally, the structure of the dissertation is presented.

1.1 Context

We live in an era where smart devices are everywhere, known as the Internet of Things, IoT. While the processing and storage of information are inherently digital, benefiting from the scaling of Moore's Law, the physical world remains fundamentally analogic. Physical variables such as temperature, sound and biological potentials are continuous in both time and amplitude. Consequently, the Analog-to-Digital Converter, ADC, serves as the critical interface bridging these two domains, enabling digital systems to interact with the real world.

Energy efficiency has shifted from being a secondary performance metric to a primary design constraint, many of the sensors operate on limited battery or rely on energy harvesting, where every Joule of power dissipation has a direct effect on the system's autonomy. In this landscape, the data conversion block almost always takes place as the most spender in this field, especially in sensor interfaces where the digital transmission is duty-cycled.

However, a fundamental characteristic of many environmental and physiological signals is sparsity in between pulses. Signals such as electrocardiograms, voice activity, or environmental monitoring data are spreaded in bursts in nature: they contain short periods of high information content followed by long intervals of inactivity.

Traditional data conversion approaches, based on uniform sampling and dictated by the Nyquist-Shannon theorem, treat these silence periods with the same computational power and energetic waste as the active periods. This creates a paradigm of inefficiency the system dissipates power to generate redundant digital samples that carry no new information. Recognizing and exploiting this sparsity is the key to unlocking ultra-low-power electronic interfaces.
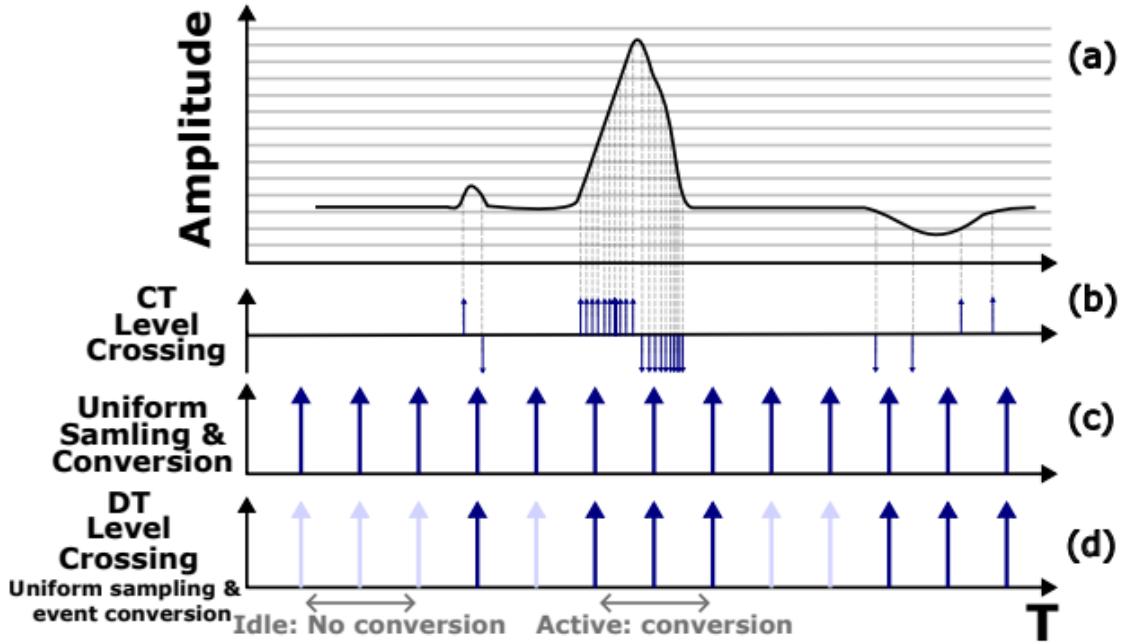
Figure 1.1: Figure adapted from Jixuan Mou (a) Time-sparse signal example. (b) Continuous-time event-driven sampling and conversion, (c) Conventional uniform sampling and (d) Discrete-time sampling and event-driven conversion

1.2 Motivation and Problem Statement

The energy waste in massive-scale IoT deployments and the inefficiency of conventional synchronous architectures when processing sparse data shows the need for a fundamental shift toward architectures that activate only when meaningful data events occur.

1.2.1 Inefficiency of Synchronous Sampling

Traditional Analog-to-Digital Converters operate under a fixed-rate sampling rate, dictated by a global clock signal (f_s). As established by the Nyquist theorem, this rate is determined by the maximum possible bandwidth of the signal, $f_s \geq 2B$. However, in many real-world applications, the signal is often active for only a fraction of the time, meaning the true information content is much lower than the theoretical maximum bandwidth suggests.

This leads to a massive waste of energy due to the fundamental relationship governing dynamic power consumption in CMOS circuits:

$$P_{dynamic} = \alpha \cdot f \cdot C_L \cdot V_{DD}^2 \tag{1.1}$$

Where:

- $P_{dynamic}$ is the dynamic power dissipated.

- α is the activity factor (or switching factor).

- f is the clock frequency.

- C_L is the load capacitance being switched (e.g., in the clock tree).

1.2.2 Architectural Problem: Flash ADCs

While the Flash ADC architecture is highly attractive due to its single-cycle, high-speed conversion capability, it suffers from severe limitations regarding power and area, particularly when reaching medium-to-high resolution. The number of required comparators scales exponentially with resolution (2^N). This exponential scaling leads directly to an increase in input capacitance and static power dissipation.

Even in low-power, synchronous designds, the dynamic power of the comparator switching is the dominant factor in the power budget.

Therefore, combining the high speed of the Flash architecture with the high power efficiency required for IoT demands a radical shift away from the synchronous clocking scheme.

1.2.3 Device Limitation: The Mismatch Problem

To mitigate the power and area issues mentioned above, designers are compelled to minimize the physical dimensions of the core devices, particularly the transistors within the comparators. However, this introduces a fundamental physical constraint:
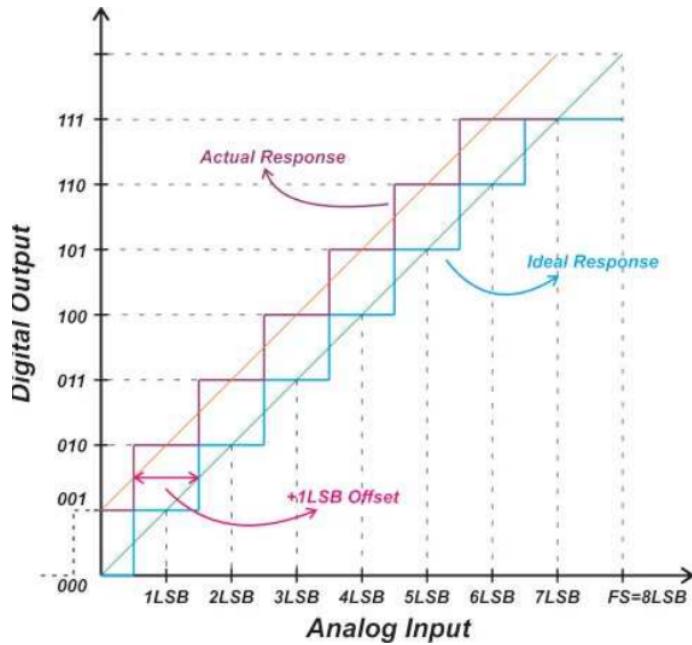

Figure 1.2: Figure adapted from [Mohammed Abdulmahdi Mohammedali / New and Accurate Self-Test Methods for Offset Error of 10-Bit SAR-ADC] show how mismatch cause offset error in ADCs.

The reduction of transistor size, while increasing speed and reducing capacitance, drastically increases random variations in parameters like the threshold voltage (V_{th}), leading to a significant Input Offset Voltage (V_{os}) in the comparators.

1.3 Objectives and Contribution

Based on the limitations identified in traditional synchronous data conversion when dealing with sparse sensor signals, the primary objective of this work is to design and implement an energy-efficient Analog-to-Digital Converter (ADC) architecture. This ADC must exploit signal inactivity to achieve a significant reduction in power consumption.

1.3.1 Specific Objectives

To achieve the overall goal of developing a highly efficient ADC, the following specific objectives are defined for this dissertation:

1. Asynchronous Architecture Design: To develop the circuit-level design of an asynchronous Flash ADC, focusing on minimizing dynamic power consumption by ensuring that power is only dissipated when an input signal event occurs.

2. Offline Trimming Implementation: To implement a robust Offline Trimming mechanism capable of detecting and compensating for the input offset voltage of the comparators.

3. Mixed-Signal Validation: To validate the complete system through comprehensive Analog and Digital co-simulation. This validation must confirm the functionality and accuracy of both the asynchronous core and the trimming logic.

4. Performance Benchmarking: To quantify the energy efficiency of the proposed design using established Figures of Merit (FoM). The results must be rigorously compared against the current state-of-the-art synchronous ADCs and relevant asynchronous solutions.

Chapter 2

Literature Review

This chapter presents the state-of-the-art in Analog-to-Digital Converters, reviewing theoretical foundations and analyzing survey data to justify the proposed architecture.

2.1 Fundamentals of Analog-to-Digital Conversion

This section provides the theoretical background required to understand the operation and performance characterization of data converters. It covers the fundamental steps of the conversion process such as sampling, quantization, and coding.

2.1.1 The Data Conversion Process

2.1.1.1 Ideal Data Conversion

The analog-to-digital conversion process acts as the bridge between the continuous physical world and the discrete digital domain. Conceptually, it involves two distinct operations: discretization in time, sampling, and discretization in amplitude, quantization. Ideally, this process should be instantaneous and lossless within the signal bandwidth of interest.

2.1.1.2 The Sampling Operation

Sampling converts a continuous-time signal $x(t)$ into a discrete-time sequence $x[n]$. Mathematically, this is modeled as the multiplication of the input signal by a periodic train of Dirac delta functions with period T_s(VER ESTA EXPLICAÇÃO MELHOR).

Sampling Theorem: According to the Nyquist-Shannon sampling theorem, a band-limited signal with maximum frequency B can be perfectly reconstructed if the sampling frequency f_s satisfies:

$$f_s \geq 2B \tag{2.1}$$

Violating this condition results in aliasing, where high-frequency spectral components mixes into the baseband, becoming indistinguishable from the original signal.

2.1.1.3 The Reconstruction Operation

Reconstruction is the inverse operation, converting the digital sequence back into a continuous signal. In an ideal scenario, this corresponds to convolving the discrete samples with a sinc function, which removes the spectral images generated during sampling, leaving only the original baseband content.

2.1.1.4 The Quantization Operation

While sampling discretizes time, quantization maps the continuous amplitude of each sample to one of a finite number of levels. An *N*-bit ADC divides the input range (V_{ref}) into 2^N discrete levels. The step size between adjacent levels is the Least Significant Bit (LSB):

$$V_{LSB} = \frac{V_{ref}}{2^N} \tag{2.2}$$

Unlike sampling, quantization is non-reversible and introduces a deterministic error. This error is typically modeled as additive white noise (*quantization noise*) with a uniform probability distribution. For an ideal quantizer, the Signal-to-Quantization-Noise Ratio (SQNR) is given by the formula:

$$SQNR_{dB} \approx 6.02N + 1.76 \tag{2.3}$$

2.1.1.5 Coding

The final stage is encoding the quantized level into a binary format. The choice of coding scheme (e.g., straight binary, two's complement, Gray code) depends on the system requirements for data processing and transmission, but does not affect the fundamental analog performance.

2.1.1.6 Undersampling and Oversampling

- Undersampling: Intentionally violates the Nyquist criterion for the carrier frequency to down-convert Radio-frequency signals directly.

- Oversampling: Involves sampling at a rate much higher than the Nyquist rate ($f_s \gg 2B$). This technique spreads the fixed quantization noise power over a wider frequency range. A subsequent digital filter can then remove the out-of-band noise, effectively increasing the SNR and resolution beyond the intrinsic bit-depth of the hardware.

2.1.1.7 Decimation and Interpolation

- Decimation: Reduces the sampling rate of an oversampled signal by filtering and downsampling, trading speed for resolution.

- Interpolation: Increases the effective sampling rate in the digital domain (zero-stuffing and filtering), often used in DACs to relax the requirements of the analog reconstruction filter.

Figure 2.1: Undersampling and Oversampling

2.1.2 Performance Metrics

To evaluate and compare different data converters objectively, a standard set of metrics is used. These are categorized into dynamic and static parameters.

2.1.2.1 Resolution and Sampling Rate

Resolution defines the theoretical dynamic range, while the sampling rate determines the maximum signal bandwidth. However, these are nominal values; real-world performance is limited by noise and non-linearities.

2.1.2.2 Signal-to-Noise-and-Distortion Ratio (SNDR)

Signal to noise ratio or SNDR is the primary dynamic metric. It is the ratio of the signal power to the total power of all noise and harmonic distortion components. From SNDR, the Effective Number of Bits, ENOB, is derived:

$$ENOB = \frac{SNDR_{dB} - 1.76}{6.02} \tag{2.4}$$

This value represents the true resolution of the converter at a specific input frequency.

2.1.2.3 Spurious-Free Dynamic Range (SFDR)

SFDR is defined as the ratio between the fundamental signal power and the power of the largest spurious component in the spectrum (typically a harmonic). A high SFDR is crucial in communication systems to detect small signals in the presence of strong interferers.

Figure 2.2: Decimation and Interpolation

2.1.2.4 Differential and Integral Non-Linearity (DNL and INL)

Static linearity is characterized by measuring the deviation of code transition levels from their ideal positions.

- DNL: Measures the deviation of a single step width from the ideal 1*LSB*. A DNL less than -1 LSB implies a missing code in the transfer function.

- INL: Is the cumulative sum of DNL errors, representing the deviation of the transfer curve from a straight line. Specific INL patterns (like "S-curves" or "saw-tooth" shapes) can reveal systematic errors in the architecture, such as gain mismatch or non-linear biasing.

Figure 2.3: DNL and INL

2.1.2.5 Offset and Gain Error

These are linear errors. Offset is a constant shift of the transfer characteristic, while gain error is a deviation in the slope. Unlike non-linearity, these errors preserve the signal shape and can often be calibrated out simply.

Figure 2.4: Offset and Gain Error

2.1.2.6 Bit Error Rate (BER)

BER quantifies the probability of the converter producing an incorrect digital code. This is often caused by metastability, a phenomenon where internal decision circuits fail to resolve a valid logic level within the allocated time when the input is extremely close to a decision threshold.

2.2 Synchronous Architectures

Synchronous ADCs have a global clock signal that dictates sampling instances and synchronizes internal operations. While these architectures are highly mature and widely used, they face fundamental power-efficiency trade-offs, particularly when operating at high frequencies or processing sparse data.

2.2.1 Synchronous Flash ADC

The Flash ADC is conceptually the simplest and fastest architecture, performing a complete conversion in a single clock cycle. It utilizes a resistive ladder to generate $2^N - 1$ reference voltages, which are compared simultaneously to the input signal by a large bank of comparators. This parallel comparison produces a "thermometer code" that a digital logic block then converts into a standard binary output. Its primary limitation is the exponential increase in hardware complexity, area, and power consumption as resolution increases.

Figure 2.5: Flash ADC Architecture: A parallel bank of comparators driven by a global clock.

2.2.2 Data Criticism: The Power Bottleneck

The primary drawback of the synchronous Flash architecture is its exponential scaling. Survey data clearly illustrates that "High Speed = High Power" in these designs. For example, a 6-bit 500 MS/s CMOS Flash ADC reported in 1999 already consumed 400 mW. More extreme cases, such as a 22 GS/s 5-bit design, can consume up to 3 W.

The continuous clocking of a massive comparator bank creates a significant "power floor," where energy is dissipated even when the input signal is static. This lack of adaptivity makes synchronous Flash architectures increasingly unsuitable for battery-powered or energy-harvesting applications.

2.2.3 SAR (Successive Approximation Register)

The SAR ADC operates using a binary search algorithm. For each sample, the internal logic tests one bit at a time, starting from the Most Significant Bit (MSB). A single comparator compares the input to the output of an internal Digital-to-Analog Converter (DAC). If the input is higher, the bit is set to '1'; otherwise, it is set to '0'. This process repeats for N cycles. Because it uses very few active components, it is the most energy-efficient choice for low-to-medium speeds.

Figure 2.6: SAR ADC Architecture: Sequential bit-estimation using a single comparator.

2.2.4 Pipeline

The Pipeline ADC breaks the conversion into several sequential stages. Each stage resolves a few bits of information, quantizes them, and passes the remaining error (the residue) to the next stage after amplification. This "assembly line" approach allows the ADC to work on multiple samples concurrently, obtaining very high throughput and high resolution, albeit with an inherent processing latency.

Figure 2.7: Pipeline ADC Architecture: Concurrent processing of multiple samples across stages.

2.2.5 Sigma-Delta (ΔΣ)

The $\Delta\Sigma$ architecture relies on oversampling (sampling much faster than the Nyquist rate) and noise-modulation. A modulator integrates the difference between the input signal and a feedback version of the quantized output. This process pushes the quantization noise into higher frequencies, outside the signal band of interest. A subsequent digital filter then removes the high-frequency noise, resulting in extremely high resolution and precision.

Figure 2.8: Sigma-Delta ADC Architecture: High-resolution through oversampling and noise-modulation.

2.2.6   Dual-slope

This integrating architecture measures the time required for a capacitor to charge and discharge. In the first phase, the capacitor is charged by the input voltage for a fixed period. In the second phase, it is discharged by a known reference voltage. The time it takes to return to zero is proportional to the average value of the input signal. It is highly accurate and immune to high-frequency noise but is much slower than other architectures.

Figure 2.9: Dual-slope ADC Architecture: High precision via time-based integration.

2.3   Asynchronous Architectures

Asynchronous, or event-driven, architectures represent a fundamental paradigm shift in data conversion. Unlike synchronous ADCs, which are bound by a global cloock and the Nyquist-Shannon sampling theorem, asynchronous ADCs operate based on the signal's activity. This approach offers a more efficient alternative for processing sparse signals.

2.3.1   Level-Crossing Sampling (LCS)

The core principle behind many asynchronous ADCs is Level-Crossing Sampling (LCS). In traditional uniform sampling, the signal is captured at fixed time intervals (T_s), and the amplitude is quantized. In LCS, the process is inverted: the amplitude levels are fixed (quantization thresholds), and the ADC records the exact time at which the input signal crosses these thresholds.

This method is particularly powerful for signals that remain constant or change slowly over long periods. Instead of generating redundant samples that capture no new information, the LCS ADC remains idle, only producing a digital event when the signal effectively changes by more than the defined threshold.

2.3.2   Asynchronous Flash ADC

The Asynchronous Flash ADC adapts the parallel structure of a standard Flash ADC but removes the sampling clock entirely.

- Continuous-Time Comparators: In this architecture, the comparators are not latched or gated by a clock. They operate in continuous time, constantly monitoring the input voltage V_{in} against the reference ladder.

- Event Generation: When the input signal crosses a threshold, the corresponding comparator changes its output state immediately. This transition triggers an asynchronous logic to generate a digital output pulse.

2.3.3   Advantage

The primary motivation for adopting asynchronous architectures is the direct relationship between signal activity and power consumption. In a synchronous system, the clock tree and the comparators switch at every clock cycle, regardless of whether the input is changing. The power consumption is dominated by the fixed clock frequency:

$$P_{sync} \approx f_{clk} \cdot C_{total} \cdot V_{DD}^2 \tag{2.5}$$

In an asynchronous ADC, the switching frequency is replaced by the *event rate*. If the signal is "silent" (constant or slowly varying), the comparators and digital logic remain static, leading to the following proportionality:

$$P_{async} \propto \text{Activity} \times V_{DD}^2 \tag{2.6}$$

This property ensures that the energy consumed is always proportional to the information content of the signal. This makes these ADCs significantly more efficient for sparse signal environments, where they can achieve near-zero power during periods of inactivity.

2.3.4   Comparitive Evaluation

An analysis of state-of-the-art converters shows that asynchronous ADCs excel in energy efficiency across a wide range of sampling rates, particularly for low-to-medium bandwidths.

While Synchronous Flash ADCs are designed for peak performance at a specific frequency and suffer from a much higher power consuption caused by the continuous clocking, asynchronous designs scale their power consumption linearly with the input activity.

2.4   Figures of Merit for ADCs

To evaluate and compare the performance of various Analog-to-Digital Converter (ADC) architectures, the scientific community relies on standardized Figures of Merit (FoM). Since ADCs vary widely in terms of resolution, speed, and power consumption, these metrics serve as a normalized benchmark to determine which design offers the best energy efficiency for a given performance target.

2.4.1   Walden Figure of Merit (FoM_W)

The Walden FoM is the most prevalent metric for evaluating low-to-medium resolution ADCs, such as the Flash and SAR architectures. It measures the energy consumed per conversion step and is defined as:

$$FoM_W = \frac{P}{2^{ENOB} \cdot f_s} \tag{2.7}$$

where P is the power dissipation and f_s is the sampling frequency. Expressed in femtojoules per conversion-step ($fJ/\text{conv} - \text{step}$), a lower Walden value indicates a more energy-efficient design. This metric is particularly useful for this dissertation to demonstrate how an asynchronous architecture can reduce the numerator (P) by eliminating clock-related overhead.

2.4.2 Schreier Figure of Merit (FoM_S)

While Walden focuses on energy per step, the Schreier FoM is typically used for high-resolution ADCs where performance is limited by thermal noise (kT/C). It accounts for the dynamic range and bandwidth (BW):

$$FoM_S = SNDR_{dB} + 10\log_{10}\left(\frac{BW}{P}\right) \tag{2.8}$$

A higher Schreier value indicates better performance. Although this work focuses on a Flash architecture (where Walden is the standard), the Schreier FoM provides a broader context regarding the physical limits of analog design.

2.5 Calibration and Trimming Techniques

In high-speed Flash ADCs, the accuracy of the system is fundamentally limited by the precision of the comparators. While the architecture fundamentals were established in the previous sections, practical implementations must address the non-idealities of the fabrication process, specifically the Input Offset Voltage (V_{os}).

2.5.1 The Component Mismatch Problem

In deep sub-micron CMOS technologies, transistors that are drawn with identical dimensions on the layout will exhibit slight differences in their electrical parameters after fabrication. This phenomenon, known as mismatch, affects the threshold voltage (V_{th}) and the current gain factor (β) of the differential pair in a comparator.

According to Pelgrom's Law[**johns_martin_analog**], the standard deviation of the threshold voltage mismatch ($\sigma_{V_{th}}$) is inversely proportional to the square root of the transistor area ($W \cdot L$):

$$\sigma_{V_{th}} = \frac{A_{V_{th}}}{\sqrt{W \cdot L}} \tag{2.9}$$

Where $A_{V_{th}}$ is a technology-dependent constant. This creates a critical trade-off: to minimize offset without calibration, transistors must be made very large, which increases parasitic capacitance and drastically degrades the ADC speed and power efficiency. Therefore, small transistors are used for speed, and calibration is employed to correct the resulting offset.

2.5.2 Calibration Classifications

Calibration techniques can be broadly categorized by their timing and domain:

- Timing: *Foreground* (Offline) calibration interrupts normal operation to measure and correct errors, while *Background* (Online) calibration operates continuously but adds significant complexity.

- Domain: *Digital* calibration corrects the output code mathematically, whereas *Analog* calibration adjusts the circuit biasing or load conditions to nullify the offset at the source.

For an asynchronous architecture, avoiding continuous clock activity is crucial to maintain low power during idle periods. Thus, Analog Foreground Calibration (Trimming) is the preferred approach.

2.5.3 Resistive Trimming Techniques

Resistive trimming aims to compensate for the imbalance in the input differential pair (M_1, M_2) by intentionally creating an opposing imbalance in the load resistance or the reference path.

2.5.3.1 Internal Resistive Loading

This method involves placing a variable resistive network in parallel (or series) with the output loads of the comparator's pre-amplifier stage.

- Mechanism: By digitally switching small resistors (or MOS switches operating in the triode region) in parallel with the load branch, the effective resistance R_L is modulated.

- Effect: Since the gain of the pre-amplifier is $A_v = g_m R_L$, changing R_L on one side adjusts the output DC level. If the differential pair has an offset $+\Delta V$, the trimming network is adjusted to introduce $-\Delta V$, effectively zeroing the error.

- Implementation: A binary-weighted bank of PMOS transistors is typically used as the variable resistance. The digital code to control these switches is determined at startup and stored in a register.

2.5.3.2 Reference Ladder Trimming

Alternatively, instead of modifying the comparator internally, the reference voltages (V_{ref}) supplied to the comparators can be adjusted.

- Mechanism: The main resistive reference ladder is tapped using a local switching network that allows fine-tuning of the tap voltage connected to each comparator input.

- Context: This technique was successfully demonstrated in asynchronous Flash ADCs, such as in the work of Chen et al. [**chen_async_2006**], where the reference ladder itself acts as the calibration DAC.

2.5.4   Advantages of Resistive Trimming for Asynchronous ADCs

Compared to dynamic techniques like Auto-Zeroing (which requires accurate clock phases ϕ_1, ϕ_2 and storage capacitors), resistive trimming offers distinct advantages for the proposed work:

1. Static Operation: Once the calibration bits are set (during the offline phase), the trimming network becomes static. It does not switch and does not require a clock, preserving the "event-driven" nature of the ADC.

2. No Switching Noise: Since the calibration is constant during conversion, it introduces no injection noise or clock feedthrough.

3. Speed Preservation: It does not add significant capacitive load to the high-speed nodes of the comparator, allowing for maximum bandwidth.

2.6   Related Works

2.7   Summary

Chapter 3

Future Work Planning, Methodologies and Tools

This chapter outlines the development strategy for the dissertation, detailing the methodologies, tools, and the schedule for the remaining phases.

3.1 Methodologies and Tools

3.2 Work Plan

As the work done so far was related to the first and second chapters of the dissertation, future work planning will focus on the three remaining sections. Chapter 3 will be about the mathematical formulation of the problem while Chapter 4 will involve simulation and discussion of results. Finally, in Chapter 5 are mentioned the conclusions and possible future works. To successfully complete the dissertation with respect to both time and quality, a Gantt Chart presented in Figure 3.1 was built considering the main tasks involved in the process and the expected number of days needed to conclude the tasks.

Chapter 4

Conclusions

This document presented the preparatory work for the design of an Asynchronous Flash ADC with calibration. The literature review confirmed that asynchronous architectures offer superior energy efficiency for sparse signal processing, validating the research direction. The identified challenge of comparator offset will be addressed through the proposed trimming methodology, as outlined in the work plan.