High Speed ADCs for Wireline Data Communications

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES
(Dr. sc. ETH Zurich)

Presented by
Abdullah Serdar Yonar
MSc ETH ETIT, ETH Zurich
born on 01.06.1991

accepted on the recommendation of
Prof. Dr. Taekwang Jang
Dr. Pier Andrea Francese
Prof. Dr. Hans-Andrea Loeliger
Dr. Gabriele Manganaro

2023
I would like to thank Prof. Dr. Taekwang Jang for giving me the opportunity to pursue a PhD at ETH Zürich Integrated Systems Laboratory and for his invaluable inputs regarding my research. I would like to express my deepest gratitude to Dr. Pier Andrea Francese from whom I have learned most of the things I know regarding analog design, and for his constant support during the time I have spent at IBM Research. I consider myself extremely lucky to benefit from his deep technical expertise, and he was always ready to help when needed. I am also grateful to my co-examiners, Dr. Gabriele Manganaro and Prof. Dr. Hans-Adrea Loeliger, for fruitful discussions toward finalizing my thesis and for having the time for my PhD examination despite their busy schedule.

I would like to thank Matthias Brändli for all kinds of help related to the CAD tools and anything digital in general, Marcel Kossel, Gain Kim, and Thomas Morf for rewarding discussions on various circuit blocks, Andrea Ruffino for his suggestions on publications and thesis sections, Mridula Prathapan for the time we collaborated.

A special thanks goes to all those who make the whole journey more enjoyable; Martino Dazzi for being the sort of a colleague who can contribute to the overall quality of time in and out of IBM, his language skills that are also utilized for the Italian abstract of this thesis and for his weird taste in music and equivalently good taste in food; Can Fırtına for sharing opinions in a very diverse range of topics during our road trips and weekend excursions; Batuhan Bardak for discussions regarding our
career paths, the burdens of being a PhD candidate and for his suggestions about
the thesis in general; Barkın Sarıtaş for his support in not so technical but equally
important aspects of life; Murat Türe for our rare but deep scientific discussions
in extraordinary places; Julia Bazińska for using every opportunity to invite me to
various events during the writing of this thesis; Idil Kanpolat for our spontaneous
trips to interesting destinations; and for all the others who have left a positive impact
during this journey.

Last but not least, I would like to thank my parents, Fatma and Himmet Yonar, for
raising me and for their unconditional love and support throughout my life, and my
brother Fatih Yonar for always being there when I needed him.

A. Serdar Yonar
Zurich, August 2023
Abstract

The increasing need for faster data traffic between electronic components requires high speed wireline links which establish communication in a variety of platforms ranging from data centers to mobile devices. The demanding channels with high loss and an ever-growing need for faster data rates necessitate advanced signal detection, equalization, and forward error correction algorithms. These capabilities are provided by analog-to-digital converter (ADC) and digital signal processor (DSP) based wireline receiver architectures. The design of the high speed ADC in such a receiver has a profound effect on the overall performance. Among the architectures studied, this thesis presents the implementation of a time-based sub-ADC and a time interleaved successive approximation register (SAR) ADC architecture targeting high speed wireline receivers. Using innovative time-domain techniques, the sub-ADC has the smallest area for 8-bit resolution published up to date, reaching above GS/s sampling speeds with state-of-the-art power efficiency. 64x interleaved SAR demonstrates solid linearity performance and reaches 56GS/s sampling rate. It is tested with different modulation formats, which lead to data rates up to 184Gb/s.

Time-based ADCs offer excellent scaling properties thanks to their digital intensive architectures. The shrinking gate delays due to CMOS scaling allow them to reach beyond GS/s sampling rates with high speed techniques. An 8-bit time-based digital intensive ADC implemented in 5nm CMOS is presented in this thesis. It introduces three techniques: 1) Bipolar ramp-based voltage-to-time converter (BVTC) to eliminate the reference voltage and to allow a wide input swing of $0.75V_{pp,diff}$. 2) $2\times$ interpolating sense-amplifier-latches (SAL) to reach a sub gate delay of 2.3ps in a
power and area efficient manner. 3) Redundancy window to accommodate possible wrong sign bit decisions. These allow it to reach 1GS/s sampling rate at 0.7V supply and 1.25GS/s sampling rate at 0.8V supply with 313 µm² area. It dissipates 1.18mW at 1GS/s and 1.9mW at 1.25GS/s, which corresponds to a Walden figure of merit (FoM) of 16.6fJ/conv-step and 20.3fJ/conv-step, respectively.

A 56GS/s 8-bit asynchronous SAR ADC fabricated in 4nm CMOS technology is presented in this thesis. The 16x4 interleaved ADC introduces a novel gate-drain bootstrapping technique in the first rank T&H switch of the interleaver and a switch-capacitor biased class-AB follower. It achieves a broad input common-mode ($V_{in,cm}$) range; from 0.3V to 0.6V, the total harmonic distortion stays below -52dB at 4.1 GHz with -0.2dBFS amplitude at 0.8V peak-to-peak maximum full scale. It has full foreground calibration capability for offset, gain, skew, and bandwidth mismatches among the sub-channels. The low frequency ENOB is 6.5, and the Nyquist frequency ENOB is 5.2, which is limited by the jitter. The ADC reaches above 27GHz bandwidth with a single 0.8V supply. It achieves a Walden FoM of 47fJ/conv-step. It is tested with PAM4, PAM8, and OFDM modulated data, and the error free operation is attained after forward error correction is applied.
La crescente domanda di traffico dati rapido tra componenti elettronici richiede interfacce di comunicazione su cavo ad alta velocità al fine di stabilire una connessione tra diverse piattaforme che possono variare da data center a dispositivi mobili. I canali di comunicazione caratterizzati da alte perdite ed il sempre crescente bisogno di velocità dati maggiori richiedono tecniche avanzate di rilevamento del segnale, equalizzazione ed algoritmi di correzione diretta degli errori (FEC). Queste funzionalità sono fornite da architetture di ricevitori da cavo basati su convertitori analogico-digitale (ADC) e processori di segnale digitale (DSP). La progettazione dell’ADC ad alta velocità in un tale ricevitore ha un impatto profondo sulle prestazioni complessive. Tra le architetture studiate, questa tesi presenta l’implementazione di architetture sub-ADC a dominio temporale ed ADC ad approssimazioni successive (SAR) con interlacciamento temporale, progettati per ricevitori da cavo ad alta velocità. Usando tecniche innovative del dominio temporale il sub-ADC ha l’area minore finora riscontrata per la risoluzione a 8 bit, raggiungendo velocità di campionamento sopra il GB/s ed efficienza di potenza allo stato dell’arte. Il SAR con interlacciamento 64x dimostra solide prestazioni in termini di linearità e raggiunge una velocità di campionamento di 56GS/s. Viene testato con diversi formati di modulazione che portano la sua velocità dati fino a 184Gb/s.

Gli ADC a dominio temporale offrono eccellenti qualità di scaling grazie alle loro architetture marcatamente digitali. La diminuzione dei ritardi di gate dovuti al ridimensionamento CMOS gli permettono di raggiungere velocità di campionamento oltre il GS/s con tecniche di alta velocità. In questa tesi viene presentato un ADC a 8 bit
e a dominio temporale con architettura spiccatamente digitale implementato in tecnologia CMOS a 5nm. Vengono introdotte tre tecniche nella sua implementazione: 1) Un convertitore tensione-tempo bipolare basato su rampa (BVTC) per eliminare la tensione di riferimento e permettere un’escursione dell’input di $0.75V_{pp,diff}$. 2) Un Sense-amplifier a latch (SAL) con interpolazione 2x che permette di raggiungere ritardi sub-gate di 2.3ps in modo compatto ed energeticamente efficiente. 3) Una finestra di ridondanza per la gestione di possibili decisioni di bit di segno sbagliate. Queste tecniche permettono di raggiungere una velocità di campionamento di 1GS/s con una tensione di alimentazione di 0.7V e una velocità di campionamento di 1.25GS/s con una tensione di alimentazione di 0.8V con un’area di 313$\mu m^2$. La dissipazione è di 1.18mW a 1GS/s e 1.9mW a 1.25GS/s, a cui corrisponde una figura di merito Walden (FoM) di rispettivamente 16.6fJ/conv-step e 20.3fJ/conv-step.

In questa tesi viene presentato un ADC SAR asincrono a 8 bit e con velocità di campionamento di 56GS/s, fabbricato in tecnologia CMOS 4nm. L’ADC con interlacciamento 16x4 introduce una nuova tecnica di bootstrapping gate-drain nello switch T&H nel primo rango dell’interlacciatore e un follower switch-capacitor biased di classe AB. Esso raggiunge un’ampia gamma di tensione di modo comune di input ($V_{in,cm}$); da 0.3V a 0.6V, la distorsione armonica totale rimane al di sotto di -52dB a 4.1GHz con ampiezza di -0.2dBFS alla tensione massima di escursione di $0.8V_{pp,diff}$. Ha piena capacità di calibrazione precedente all’operazione per i disallineamenti in offset, guadagno, skew ed ampiezza di banda tra i sottocanali. Il numero di bit effettivi (ENOB) a bassa frequenza è 6.5 e la ENOB alla frequenza di Nyquist è 5.2, essendo limitato dal jitter. L’ADC ottiene una ampiezza di banda oltre 27GHz con una singola tensione di alimentazione di 0.8V. Inoltre, ottiene una FoM Walden di 47 fJ/conv-step. È stato testato con dati modulati con PAM4, PAM8, OFDM e il funzionamento privo di errore è stato raggiunto dopo l’applicazione della correzione diretta degli errori (FEC).
List of Publications

A. Serdar Yonar, Pier Andrea Francese, Matthias Brändli, Marcel Kossel, Mridula Prathapan, Thomas Morf, Andrea Ruffino, and Taekwang Jang. An 8b 1.0-to-1.25GS/s 0.7-to-0.8V Single-Stage Time-Based Gated-Ring-Oscillator ADC with 2× Interpolating Sense-Amplifier-Latches. In 2023 IEEE International Solid-State Circuits Conference (ISSCC), 2023


Marcel A. Kossel, Vishal Khatri, Matthias Braendli, Pier Andrea Francese, Thomas Morf, Serdar A. Yonar, Mridula Prathapan, Eric J. Lukes, Raymond A. Richetta, and Carrie Cox. 8.3 An 8b DAC-Based SST TX Using Metal Gate Resistors with
1.4pJ/b Efficiency at 112Gb/s PAM-4 and 8-Tap FFE in 7nm CMOS. In 2021 IEEE International Solid-State Circuits Conference (ISSCC), volume 64, pages 130–132, 2021

List of Patents


# Contents

1 Introduction 19

1.1 Thesis Goal ......................................................... 25
1.2 Main Contributions of the Thesis ................................. 26
1.3 Organization of the Thesis ......................................... 27

2 Overview 29

2.1 Modulation Formats ................................................ 29

2.1.1 Pulse Amplitude Modulation .................................. 29
2.1.2 Orthogonal Frequency Division Multiplexing ............... 32

2.2 ADC metrics ....................................................... 35

2.2.0.1 Sampling Rate \( f_s \) ........................................ 35
2.2.0.2 3dB Bandwidth ............................................. 35
2.2.0.3 Signal to Noise Ratio (SNR) ............................ 35
2.2.0.4 Signal to Noise and Distortion Ratio (SNDR) ......... 36
2.2.0.5 Effective Number of Bits (ENOB) ...................... 36
2.2.0.6 Spurious-Free Dynamic Range (SFDR) ................. 36
2.2.0.7 Harmonic Distortion \( HD_k \) .......................... 37
2.2.0.8 Total Harmonic Distortion (THD) ...................... 37
2.2.0.9 Dynamic Range ............................................ 37
2.2.0.10 Walden Figure-of-Merit ................................. 37
2.2.0.11 Schreier Figure-of-Merit ............................... 38
2.2.0.12 Effective Resolution Bandwidth (ERBW) ............ 38
2.3.3.2.3 Phase Interpolation ........................................ 68
2.3.3.2.4 Time Amplification ........................................ 69
2.3.3.2.5 Folding ...................................................... 69
2.3.3.3 Voltage-to-Time Conversion .................................. 70
2.3.3.3.1 Input Voltage to Control the Current .................. 71
2.3.3.3.2 Input Voltage as a Start Voltage ....................... 71
2.3.3.4 Prior Art ....................................................... 72
2.3.4 Time interleaving .................................................. 73
2.3.4.1 Offset Mismatch ............................................... 75
2.3.4.2 Gain Mismatch ................................................ 75
2.3.4.3 Timing Skew Mismatch ....................................... 76
2.3.4.4 Bandwidth Mismatch ......................................... 76

3 Time-based ADC .......................................................... 77
3.1 Motivation ............................................................... 77
3.2 Implementation ........................................................ 79
3.2.1 Top Level Block Diagram ....................................... 79
3.2.2 Timing ............................................................... 81
3.2.3 Voltage-to-Time Conversion .................................... 82
3.2.3.1 Sampling Switch .............................................. 82
3.2.3.2 Bipolar Voltage-to-Time Converter (BVTC) ............. 83
3.2.3.2.1 Pre-decision phase ....................................... 86
3.2.3.2.2 Decision phase .......................................... 86
3.2.3.2.3 Reset phase ............................................. 88
3.2.3.3 Charge Injection CDAC ..................................... 88
3.2.3.3.1 Ramp Voltage Step ..................................... 89
3.2.3.3.2 Redundancy Window .................................... 89
3.2.3.4 Sign Bit Comparator .......................................... 92
3.2.3.5 Zero Crossing Detector .................................... 93
3.2.4 Time-to-Digital Conversion .................................... 94
3.2.4.1 Gated Ring Oscillator (ROSC) ........................................ 95
3.2.4.2 Phase interpolation ..................................................... 96
3.2.4.3 Integer Counter ......................................................... 98
3.2.4.4 Asynchronous Sampling ............................................... 99
3.2.4.5 Regenerative Buffer (BUFLAT) ...................................... 100

3.3 Cascaded Block Performance ............................................. 101

3.4 Measurement Results ..................................................... 102
  3.4.1 Static Tests ............................................................ 105
    3.4.1.1 Folding Point ................................................... 105
    3.4.1.2 ADC Static ENOB ............................................... 107
    3.4.1.3 Sign Bit Comparator Characterization ......................... 108
    3.4.1.4 INL and DNL ..................................................... 108
  3.4.2 Dynamic Tests ........................................................ 108
    3.4.2.1 Input Spectra at Nyquist ..................................... 111
    3.4.2.2 Input Frequency Sweeps ..................................... 111
  3.4.3 Power Dissipation .................................................... 111

3.5 Conclusions and Comparison with Prior Art .......................... 115

4 64x Interleaved SAR ADC .................................................. 117
  4.1 Motivation ............................................................... 117
  4.2 Implementation .......................................................... 119
    4.2.1 Overall architecture and clocking ............................... 119
    4.2.2 First Rank T&H .................................................... 122
    4.2.3 First Rank Buffer ................................................ 127
    4.2.4 Asynchronous SAR (ASAR) Sub-ADC ............................. 131
  4.3 Calibration ............................................................... 138
  4.4 Measurement Results .................................................. 141
    4.4.1 Sinusoidal input tests .......................................... 143
    4.4.2 Modulated Data Experiments .................................... 147
  4.5 Conclusions and Comparison with the Prior Art .................... 150
Chapter 1

Introduction

In the age of data, billions of devices are connected globally. By 2023 it is estimated that two-thirds of the global population will have access to the Internet, and there will be approximately 30 billion networked devices. The number of connected devices is increasing with a compound annual growth rate (CAGR) of 10%, with IoT M2M devices and smartphones accounting for the majority. This leads to the transfer of exabytes of data per month \[6\]. The bandwidth requirement is also increasing in a similar trend. The global fixed broadband and average mobile bandwidth are projected to increase twofold and threefold, respectively, by 2023 in a five-year period \[6\].

The wide adoption of cloud services and speed intensive applications like ultra high definition video (UHD) and virtual reality (VR) streaming are several main drivers. Switching to fifth-generation and sixth-generation (5G/6G) technology standard for cellular networks further paves the way for faster connections. In addition, the computational benefits brought by AI on distributed computing require a tremendous amount of data transfer. On a smaller size scale, major semiconductor companies started adopting chiplet technology. The increasing demand for performance and functionality requires a bigger silicon area, reducing the yield and increasing the cost of production and development. Chiplet technology offers a division of functions by separate chiplets on the same package. This allows each of them cost-optimized and mitigates the yield issues \[7\]. However, it requires the transfer of high speed
data between the chiplets. Hence, the need for rapid data transfer among all platforms requires wireline communication links between electronic components capable of transmitting and receiving high speed data in a robust way.

High speed wireline I/O links are employed in many applications ranging from high performance computing (HPC) clusters to mobile devices to establish the communication between different electrical and optical components. The increase in utilization and transfer of mass data requires the adoption of faster wireline communication standards to keep up with emerging applications. Aggregate data rates can be increased by increasing the number of lanes; however, this results in higher power dissipation, pin count, interconnect costs, and package area. Consequently, the industry has a clear trend of increasing per-lane data rate to reach a power and cost-efficient solution while enhancing the total bandwidth. This is reflected in Fig 1-1 where the per-lane data rate trends of various wireline communication standards are shown. Across a variety of standards, the per-lane data rate doubles every four years.

![Figure 1-1: Per-Lane data rate trend of various wireline communication standards.](image)

Networking equipment in data centers must adapt to support the growing Internet Protocol (IP) traffic. In a data center, the servers are organized in racks, and there
are different wireline connections between the components depending on the length. Optical Interconnecting Forum develops Common Electrical Interface (CEI) to ensure interoperability between the devices of different vendors. OIF developed CEI-112G for the current state-of-the-art 112Gb/s per-lane applications, while CEI-224G is being developed for 224Gb/s per-lane applications at the time of writing. CEI-112G divides the application space depending on the distance:

**CEI-112G-MCM** For ultra-short reach applications with link lengths less than 25mm. The target application can be a Multi-Chip Module that covers 3D stack integration and 2.5D integrated chiplet-to-chiplet applications.

**CEI-112G-XSR** For extra short-reach applications running up to 50mm covering die-to-die and die-to-optical engine connections. The expected loss is 6-10dB at 28GHz.

**CEI-112G-VSR** For very short-reach applications targeting chip-to-module connections. The expected loss is 12-16dB at Nyquist frequency. With VSR applications, a forward error correction code (FEC) can be introduced, which relaxes bit error rate (BER) requirements to $10^{-6}$.

**CEI-112G-MR** For medium range chip-to-chip and midplane connections up to 500mm. The target BER with FEC is $10^{-5}$ and the loss at Nyquist frequency is 20dB.

**CEI-112G-LR** For a long-reach electrical interface up to 1000mm using a backplane or copper cables. At this distance, the loss becomes 28-30dB, and a stronger FEC is applied for a target BER of $10^{-4}$.

Long-reach and medium-reach applications are the most challenging to obtain the target BER. Fig 1-2 shows a typical backplane channel. Each discontinuity that is not ideally impedance matched introduces reflections among the channel. In addition to this, the frequency dependent loss of the backplane trace introduces dispersion. The physical channels show low pass behavior, and I/O links have to deal with considerable loss at higher frequencies with increasing link lengths, as indicated in the
CEI-112G standard. To reduce the data rate limiting effects of the physical channels and increase the spectral efficiency, multi-level signaling can be deployed. The four-level pulse amplitude modulation (PAM4) is the dominant modulation for OIF CEI-112G standard. This reduces the bandwidth requirement by a factor of two compared to two-level pulse amplitude modulation (PAM2, NRZ). However, the reduced voltage margins decrease the SNR. The aforementioned non-idealities give rise to intersymbol interference (ISI), which is the distortion of the symbol due to interference between neighboring symbols. To reduce the effect of ISI and to reach a target BER, an extensive equalization strategy is needed together with FEC, especially for long and medium-reach applications. The increase in equalization complexity necessitates innovation in wireline transceiver architectures to manage demanding specifications.

![Diagram of a typical backplane link](image)

Figure 1-2: The components of a typical backplane link.
I/O link architectures can be classified as mixed signal and ADC-DSP based architectures depending on the domain of equalization. Fig. 1-3 illustrates a mixed signal wireline TX-RX architecture. On the RX side, the received signal is equalized by a continuous time linear equalizer (CTLE), which introduces high pass behaviour and peaking to cancel out the low pass behaviour of the channel. It is followed by a variable gain amplifier (VGA) to amplify the signal up to a desired full-scale level. Depending on the requirements, multiple continuous equalizers or amplifiers can be cascaded. This part of the receiver is referred to as the analog front end (AFE). The signal decision is taken by a comparator array (slicers) with different thresholds. The number of slicers depends on the clocking rate of the RX and the modulation format. It can also vary depending on the decision feedback equalization (DFE) strategy. The number generally increases exponentially with the complexity of the modulation format. The slicers are followed by the DFE that removes the postcursor ISI using the scaled version of the past sample decisions. An advantage of the DFE is that the signal is fed back after the quantizer, which prevents noise amplification \[10\]. However, introducing the feedback requires closing the loop timing in the determined symbol interval. A loop unrolled DFE can relax the timing requirements by increasing the hardware overhead \[11\]. The Feed Forward Equalization (FFE) can also be implemented on the RX side, which uses the scaled and shifted versions of the acquired signal. However, this operation requires accurate delaying, addition, and subtraction capability in the analog domain, which is costly in terms of area and power, limiting the maximum number of FFE taps \[12\]. Instead, FFE functionality can be added on the TX side at the cost of introducing a back-channel. After the equalization, the
received data is deserialized and further processed.

Fig. 1-4 shows a typical block diagram of an ADC-DSP based receiver. In a similar manner, the received symbols are linearly equalized by a CTLE and amplified by the following VGA. Then instead of equalization in the mixed signal domain, it is converted to the digital domain by a high speed analog to digital converter (ADC). The equalization is executed in the digital domain by the DSP. This allows the RX design to leverage the benefits of the CMOS scaling in terms of area and power. In addition, DSP can run more versatile algorithms for equalization and detection, which makes ADC-DSP based architecture a suitable choice for more spectrally efficient modulation formats. To illustrate, FFE can be implemented in a parallel manner and pipelined to relax the timing requirements \[13\]. DFE still has the problem of strict feedback timing requirements. However, this can be relaxed by loop unrolling \[14\] and look ahead multiplexing \[14\]. Hence, an optimized FFE-DFE combination can be applied that works best for both the smooth channels and frequency notches to alleviate the ISI. Furthermore, the calibration of the ADC can further be integrated within the DSP and optimized together with the equalization strategy. As a result, ADC-DSP based architectures are the most promising candidates for reaching higher data rates by using spectrally efficient modulation formats. Most of the architectures reaching 112Gb/s and above utilize ADCs today.

![Figure 1-4: The simplified block diagram of an ADC-DSP based link.](image-url)
1.1 Thesis Goal

The data-intensive applications will impose stricter bandwidth requirements in the future. To not pose a bottleneck in data transfer, the wireline standards must adapt, requiring higher aggregate data rates. For a cost and power efficient solution, the number of pins and physical channels must be sustained, leading to a dramatic increase in per-lane data rate. This trend can be supported through innovations both in link architectures and signaling. Adopting higher order modulation formats to increase spectral efficiency yields higher bit/symbol. This requires an ADC-DSP based wireline receiver capable of advanced equalization schemes, especially in medium to long-reach I/O links. The ADC is the critical block that must satisfy the challenging requirements in several dimensions.

1. Time interleaving is necessary for reaching tens of GS/s sampling speed. The single channel ADC sampling speed must be sufficiently high to reduce the interleaving factor. A higher interleaving factor means higher area and cost, strict interleaver requirements, and more jitter sensitive clocks to distribute. However, a very high single channel sampling rate can also yield a sub-optimal solution due to extra power dissipated stemming from technology limitations and calibration. An ideal sampling rate must be balanced between the interleaving factor and the power/area efficiency of the sub-ADC.

2. The resolution is defined by the target BER, channel profile, equalization scheme, modulation format, and FEC. Overall 6-8-bit nominal resolution is sufficient for current state-of-the-art 112Gb/s and higher links utilizing PAM4.

3. The power efficiency of the ADC is critical for wireline RX to reach the target pJ/bit metric.

4. Considering that multiple ADCs will be time interleaved, the area of a single ADC is of critical importance to keep the total area low and to cut the cost down.
Considering the points listed above, the aim of this thesis is to design a high performance 8-bit ADC architecture tailored for a wireline RX targeting high power efficiency and minimum area. 5nm and 4nm CMOS FinFET technologies were used for the hardware runs. The thesis focuses on two main projects. The first one aims to explore a digitally intensive time-based sub-ADC architecture that offers compactness and high power efficiency. This can be an alternative to a SAR ADC, which is widely used in a time interleaved manner for wireline RX. The goal of the second project was to design a time-interleaved SAR ADC with high linearity, which was tested with PAM and OFDM data to demonstrate the operation with advanced modulation formats.

1.2 Main Contributions of the Thesis

A Single-Stage Time-Based Gated-Ring-Oscillator ADC with 2X-Interpolating Sense-Amplifier-Latches

A digitally intensive time-based ADC is presented. This is a single stage architecture that introduces a bipolar voltage-to-time converter (BVTC). The BVTC generates two opposite polarity ramps, and the differential zero crossing is detected by a zero crossing detector (ZCD) which controls a gated ring oscillator (ROSC). ROSC acts as a fine counter. Each cycle of ROSC increments an asynchronous counter which serves as a coarse counter. A power efficient 2x interpolation strategy is proposed, which utilizes interpolating sense amplifier latches (SAL) to interpolate the outputs of the ROSC only when sampling. A redundancy window is created for possible wrong decisions of the comparator.

The test chip is designed in 5nm FinFET technology. The architecture occupies 313µm² owing to the digital style layout. At 1GS/s, it runs with 0.7V supply and achieves 16.6fJ/conv-step Walden FoM. At 1.25GS/s, it runs with 0.8V supply and achieves 20.3fJ/conv-step. For 1GS/s, the DNL range is [-0.34,0.33] LSBs, and the INL range is [-0.64,0.69] LSBs. For 1.25GS/s, the DNL range is [-0.31,0.52] LSBs,
and the INL range is [-0.85,1.26] LSBs. The ENOBs at Nyquist frequency for both cases are 6.2.

**An 8-bit 56GS/s 64x Time-Interleaved SAR ADC**

A 64x time interleaved asynchronous SAR ADC design is presented. The design accommodates an improved bootstrapped sampler, which is named as constant $V_{gd}$ bootstrapping, for the first rank of the interleaver. It increases the input bandwidth by reducing the capacitive loading at the input node. Furthermore, it has fewer devices on the tracking path compared to conventional implementations, reducing the tracking delay. The first rank buffer has a class-AB topology biased by a switched capacitor network. This not only provides a high current driving capability but also allows the DC signals, eliminating the DC baseline wander problem. An asynchronous SAR ADC with a differential memory cell is utilized as a sub-ADC. This provides extra robustness against metastability.

The design is implemented in 4nm FinFET technology. It occupies 0.078mm$^2$. Running with a single supply at 56GS/s, it achieves a low frequency ENOB of 6.5 and a high frequency ENOB of 5.3, which is limited by jitter at Nyquist frequency. It consumes 240mW power, which corresponds to 47fJ/conv-step.

**1.3 Organization of the Thesis**

The thesis is organized as follows:

**Chapter 2** This chapter gives an overview of the theory. It starts with a short review of the modulation formats PAMN and OFDM. Then ADC performance metrics and errors are explained. The chapter continues with a brief discussion of three ADC architectures that are considered suitable for wireline links, namely: Flash, SAR, and Time-based. Flash ADCs used to be synonymous with the high speed ADC concept for many years. The operation principles and performance improvement techniques for flash ADCs are a starting point.
for discussing ADC architectures. SAR ADC is a proven architecture for interleaved ADCs targeting wireline links. Thus, the operation principle and the techniques for increasing SAR sampling speed are analyzed. Time-based ADCs are promising architectures that favor CMOS scaling and allow compact and power efficient designs. Hence, they are promising candidates for the next generation RX architectures. An overview of the current state-of-the-art time-based ADCs and building blocks are also given in this chapter. Time interleaving technique to increase the sampling rates is introduced.

**Chapter 3** In this chapter, a digitally intensive time-based ADC is introduced. After the architectural overview, the building blocks are discussed in detail. The measurement setup is explained, and the results are given. The chapter is concluded with a comparison of state-of-the-art and critical points.

**Chapter 4** This chapter discusses the design of the 64x time interleaved SAR ADC. The operation of the building blocks of the sub-ADC and the interleaver are explained. The clocking scheme is discussed. The foreground calibration strategy is analyzed. The measurement results are given at the end of the chapter, together with the test runs with different modulation formats. The comparison with the state-of-the-art is also included.

**Chapter 5** The work and contributions are summarized in Chapter 5. The takeaways for both the time-based ADC and the 64x time interleaved SAR ADC are discussed. The performance comparison with the architectures previously published in ISSCC and VLSI is given.
Chapter 2

Overview

This chapter serves to provide a condensed overview of the fundamental theoretical concepts that are vital to comprehending the central focus of the thesis. First, the modulation formats suitable for modern wireline communication applications are briefly presented. Second, the working principles of flash ADC, SAR ADC, and time-based ADC are covered. These constitute the majority of high speed interleaved architectures today. Third, challenges regarding time-interleaving are explained.

2.1 Modulation Formats

Baseband transmission is the dominant choice of digital transmission used for wireline communication systems. PAM is one of the most efficient baseband modulation formats for optimum power and bandwidth [15].

2.1.1 Pulse Amplitude Modulation

PAM is the most widely used modulation technique for wireline links where the data is encoded in different pulse amplitude levels. The number of levels indicates the modulation order and defines the spectral efficiency.

In PAM2 format, the binary data is encoded in two voltage levels of the pulse as
Figure 2-1: The time domain waveforms, eye diagrams, and power spectral density of (a) PAM2 (b) PAM4, and (c) PAM8. The symbol periods, vertical eye opening, and Nyquist frequencies are also illustrated. The eye diagram is constructed after low pass filtering the time domain signal while the power spectral density shows the ideal bit sequence.

shown in Fig. 2-1 (a). Since the range of full scale voltage available is limited by the nominal supply voltage and the dynamic range of TX and RX, PAM2 achieves the highest SNR among multi-level amplitude modulation formats. Thus, it is historically the natural choice for early communication systems. However, the increase in data rate demand and band limited channels require solutions that can provide more data per symbol. For the same baud rate, this can be achieved by increasing the number of levels that the symbol can accommodate in pulse amplitude modulation. In PAM4 there are four distinct levels that allow two binary bits to be encoded in one symbol, whereas PAM8 accommodates eight distinct levels and allows three binary bits to be encoded as illustrated in Fig. 2-1. For a symbol period of $T_b$, this corresponds to a data rate of $1/T_b$ for PAM2, $2/T_b$ for PAM4, and $3/T_b$ for PAM8. The theoretical bandwidth requirement reduction can be generalized for a PAMN signal as $\log_2(N)$.

For a state-of-the-art 224Gb/s TX-RX, this would mean a Nyquist bandwidth of
112GHz, 56GHz, and 37.3GHz for PAM2, PAM4, and PAM8, respectively.

The increase in modulation order provides a way to increase the data rate for the same baud rate, which helps relax the bandwidth requirements of the TX-RX. However, this comes at the cost of decreased signal integrity and more complex design requirements. The SNR decreases due to smaller decision level spacing when the modulation order is increased. This can be observed from the reduction of vertical eye opening (EH) in [2-1] Assuming a similar RMS noise at the receiver and the same full swing input amplitude, the decreased decision level spacing results in $|20 \log (1/3)| = 9.5dB$ less SNR for PAM4, $|20 \log (1/7)| = 16.9dB$ less SNR for PAM8 and $|20 \log (1/(N-1))|$ less SNR for PAMN compared with NRZ. A higher N number leads to more sensitivity to amplitude and time domain noise. Since ISI has a detrimental effect on both vertical and horizontal eye openings, increasing N has a significant impact on BER. To illustrate, the sensitivity of PAMN is $(N-1)$ times bigger than NRZ, as shown in [10]. This increases the equalization efforts leading to a more complex DSP and an increased number of FFE and DFE taps. Furthermore, a higher N number gives rise to a more significant difference between the amplitude of minimum and maximum transitions. This makes small transitions more susceptible to coupling from bigger transitions under the effect of crosstalk. Another challenge posed by PAMN is the asymmetry of the eyes in the eye diagram. Since the eyes do not have the same shape in a PAMN modulation format due to multiple possible transitions, the worst eye limits the BER. The eye width and height should be optimized to reach the target pre-FEC BER.

In general, it can be stated that opting for a higher N results in relaxing the timing requirements of the wireline TX-RX. Yet, it brings challenges in terms of signal integrity and power dissipation. However, with the increasing per-lane data rate demand, moving toward higher spectral efficiency is inevitable, which increases TX-RX complexity by demanding higher resolution with similar power efficiency.
2.1.2 Orthogonal Frequency Division Multiplexing

OFDM is a modulation technique that utilizes different carrier frequencies for transmitting parallel data. The concept was first developed by Bell Labs in 1966 [17]. It was further developed by introducing FFT to precisely place the channels close to each other in frequency domain [18]. The concept was used extensively in wireless systems as a remedy against multi-path propagation. In 1991 it was introduced to wireline communications and used for a high bit rate digital subscriber line application (HDSL) [19].

![OFDM diagram](image)

Figure 2-2: OFDM in frequency domain, time domain, and data sequence including cyclic prefix (CP).

OFDM uses multiple narrow frequency bands called sub-channels (sub-carriers) for transferring data, as illustrated in Fig. 2-2. Generally, each channel is quadrature amplitude modulated (QAM). If the sub-channel bandwidth is narrow enough, they can be assumed to have a flat frequency response. The parallel transfer of data allows the period of the symbol to be extended without affecting the total data rate, significantly reducing the effect of ISI. The concept is similar to frequency division multiplexing (FDM), except that the sub-carriers are orthogonal to each other. Orthogonal sub-carriers can be recovered in the RX without the effect of neighboring carriers which allows them to be placed close or overlapping in the frequency domain.
for increasing the spectral efficiency as shown in Fig. 2-2. To further increase the resilience of the link against ISI, a cyclic prefix (CP) code can be used, as illustrated in Fig. 2-2 in the time domain. CP is a guard interval used between the OFDM signals where some portion of the last samples are placed in front of the corresponding symbol to reduce the ISI between two symbols.

Figure 2-3: An OFDM RX-TX block diagram.

The block diagram of a typical RX-TX pair for OFDM modulation is given in Fig. 2-3. On the TX side, the QAM modulated data is generated. The generated vector is forwarded to Inverse Discrete Fourier Transform block, which converts the frequency domain components to the time domain signal. Then the CP code is added, and the data is serialized. A DAC transmits the serialized data. On the RX side, the transmitted data is first conditioned by the AFE and then converted to the digital domain by an ADC. The CP is removed, and the received data is converted back to the frequency domain by the Discrete Fourier Transform block. The data is equalized in the digital domain and then demodulated.

After propagation in the physical channel, the received QAM symbols are attenuated and shifted in phase. One significant advantage of the OFDM is that the received symbols can be amplified, and the phase can be corrected by a single multiplication per subcarrier [20]. This allows an efficient equalization solution by shifting the computation overhead from the analog domain to the digital domain. Another significant benefit of the OFDM is that it allows the flexibility of choosing the modulation order
and the power of each sub-channel. This is called bit and power loading, respectively. Bit loading allows the sub-channels with less SNR due to the frequency response of the physical channel to employ a relaxed modulation strategy by reducing the order, as illustrated in Fig. 2-4. The same can be applied to the power of the sub-channels: a higher loss can be compensated by allocating a higher power on the TX side to the specific sub-channel, increasing the signal-to-noise ratio (SNR). An OFDM modulation utilizing this kind of technique is called as discrete multitone (DMT).

![Diagram of Bit Loading and Power Loading](image)

Figure 2-4: Bit loading relaxes the modulation format depending on the frequency response of the channel.

The spectral efficiency combined with simple equalization of the individual sub-channels supports OFDM as a promising candidate as a modulation technique for the next generation communication standards targeting above 224Gb/s. The effectiveness of OFDM in I/O links has already been demonstrated with data rates between 56-112Gb/s [21][22][23]. The feasibility of DMT for high speed communications is further discussed in [24].
2.2 ADC metrics

ADC metrics are used to characterize the ADC and quantitatively describe its performance. The metrics used in this thesis are listed below.

2.2.0.1 Sampling Rate \( (f_s) \)

The sampling rate is the frequency at which the input signal is sampled. Samples per second \((\text{S/s})\) is the unit to describe the sampling rate. The sampling instants of the ADC are not necessarily uniformly spaced. There are types of ADCs exploiting the spectral properties of non-uniform sampling to relax the anti-aliasing filter requirements \cite{25, 26}.

2.2.0.2 3dB Bandwidth

3dB bandwidth is the frequency at which the fundamental spectral component of the output decreases by 3dB.

2.2.0.3 Signal to Noise Ratio (SNR)

SNR is the ratio of the rms signal and rms noise powers at the ADC output. In the ideal case, noise consists only of the quantization noise component. The signal-to-quantization noise ratio (SQNR) can be written as:

\[
SQNR = 6.02N + 1.76
\]  \hspace{1cm} (2.1)

where \(N\) is the nominal number of bits of the ADC which means there is an additional 6.02dB SNR increase with each additional bit due to shrinking quantization error. The derivation comes from the definition of the quantization noise, which will be explained in the following chapters. Thus, SQNR is the theoretical maximum SNR that an ADC with \(N\) bits can achieve.
2.2.0.4 Signal to Noise and Distortion Ratio (SNDR)

SNDR is the ratio of the signal power to all error powers stemming from the distortion and the noise of the ADC at the output. It is generally measured with full-scale amplitude or an amplitude slightly backed off from the full-scale to correctly characterize the nonlinearity. There are several ways to measure the SNDR, and two of them are widely used:

**Sine-fit Method** In this method, a sine wave with a specified frequency and amplitude is applied to the ADC. The reconstructed sine at the ADC output is fitted to an ideal sine wave. The difference between the sampled points and the ideal sine wave represents the error at each time instant. The error powers at each point can be summed up to find the total error power that can be used to calculate the SNDR 27.

**FFT Method** The output of the ADC can be analyzed in the frequency domain by taking the Fast Fourier Transform (FFT) at the output. The noise power spanning the Nyquist zone can be extracted together with the power from harmonics, and SNDR can be calculated from these components.

2.2.0.5 Effective Number of Bits (ENOB)

ENOB indicates the practical resolution of the ADC by taking noise, distortion, and other artefacts arising in the conversion process into account. The ENOB is measured with full scale or close to full scale input. It can be calculated from the SNDR value as:

\[
ENOB = \frac{SNDR - 1.76}{6.02}
\]  

(2.2)

2.2.0.6 Spurious-Free Dynamic Range (SFDR)

SFDR is the ratio of the signal power to the power of the largest frequency spur within a given frequency interval. It is generally defined in dBc as the ratio of the spur to the carrier.
2.2.0.7 Harmonic Distortion ($HD_k$)

Harmonic distortion is the ratio of the power of the kth harmonic to the signal power.

2.2.0.8 Total Harmonic Distortion (THD)

THD is the ratio of the total power of the harmonics to the signal power. Unless stated otherwise, the harmonics from the second to the tenth are taken into consideration.

2.2.0.9 Dynamic Range

Dynamic range is the ratio of the full scale input to the smallest detectable signal. In general, the smallest detectable signal corresponds to the point where the SNDR decreases to zero.

2.2.0.10 Walden Figure-of-Merit

Walden FoM is a metric to assess the ADC performance by taking into account the power dissipation, ENOB, and the sampling frequency $[28]$. It is written as:

$$FoM_W = \frac{P}{2^{ENOB} \times f_s}$$ (2.3)

It is a widely used metric today in the literature and has units of joules per conversion step (J/conv-step). Using Walden FoM as a comparison metric yields more accurate results when comparing the ADCs with similar resolutions. This is because the power scales more than two when adding an additional ENOB in noise limited designs. To illustrate, transconductance or capacitance has to be increased by four to halve the noise power. This indicates a power increase by a factor of four to obtain an additional 6dB SNDR compared to a two-fold power increase assumed by Walden FoM.
2.2.0.11 Schreier Figure-of-Merit

Another performance metric used for ADCs is the Schreier FoM. It can be written as \[29\]:

\[
FoM_S = SNDR_{dB} + 10 \log_{10} \frac{f_s}{2P}
\] (2.4)

Schreier FoM indicates a four times power increase per additional bit, which is more appropriate for noise limited designs.

2.2.0.12 Effective Resolution Bandwidth (ERBW)

ERBW is the input frequency at which the SNDR of the ADC drops by 3dB, which corresponds to half ENOB.

2.2.0.13 Third Order Intermodulation Distortion (IMD3)

The increasing signal complexity can render single tone test insufficient when the frequency band of interest does not span the whole Nyquist frequency range. The input signal can contain components from different frequencies simultaneously. The non-linearity of the converter gives rise to intermodulation products of these frequencies which can cause distortion within the frequency interval of interest. IMD3 is used to quantify this behaviour. To illustrate, two-tone input signal with frequencies f1 and f2, which are close to each other, can be applied to the ADC. The output will contain the fundamental components f1 and f2, intermodulation products 2f1-f2 and 2f2-f1, and the harmonics of f1 and f2. IMD3 is the ratio of the power of the intermodulation product to the input signal. This is a standard test to characterize the third order distortion of the RF ADCs used for wireless communications where the frequency interval of interest is within the sub-channel and does not span the whole Nyquist range.
2.2.1 ADC Errors

2.2.1.1 Quantization Noise

Quantization noise is the inherent noise of an ADC stemming from the operation of mapping an infinite domain of input values to a finite domain of output values. Assuming there are enough input values sampled, this noise can be treated as a uniformly distributed random noise the rms value of which decreases with increasing number of bits due to finer quantization intervals. Assuming it has a uniform distribution, which is valid for most systems, it takes values between 0.5LSB and -0.5LSB. Then the rms value of the quantization noise can be written as:

\[ V_{Q,rms} = \frac{V_{LSB}}{\sqrt{12}} = \frac{V_{fs}}{2^N \sqrt{12}} \]  \hspace{1cm} (2.5)

where \( V_{fs} \) is the full scale input and \( V_{LSB} \) is the LSB value of the ADC. Assuming a sinusoidal input, the rms value of the input signal can be written as:

\[ V_{in,rms} = \frac{V_{fs}}{2\sqrt{2}} \]  \hspace{1cm} (2.6)

and the SQNR value can be written as:

\[ SQNR = 20 \log_{10} \left( \frac{2^N \sqrt{12}}{2\sqrt{2}} \right) = 6.02NdB + 1.76dB \]  \hspace{1cm} (2.7)

2.2.1.2 Differential and Integral Nonlinearity (DNL,INL)

DNL and INL define the deviation of the ADC transfer function from an ideal quantizer. They are generally given in LSBs. DNL describes the deviation of the difference between two consecutive transition points on the analog input axis from an ideal LSB step as illustrated in Fig. 2-5. INL characterizes the deviation of the transfer function from the ideal one, as shown in Fig. 2-6. The ideal transfer function is either a straight line that connects the endpoints or the best fit line of the transition points. DNL is the derivative of the INL. DNL bounded between ±1 guarantees no missing
codes at the ADC output. The sum of the DNL over the whole range of output code is zero:

\[ \sum_{k=1}^{2^N} DNL[k] = 0 \] (2.8)

While measuring INL and DNL, the noise of the ADC must be taken into account, which suggests using histogram based statistical methods. Two of the methods widely used are: Ramp Histogram and Sine Histogram [27].
2.2.1.2.1 Ramp Histogram:

For this method, a slow, highly linear ramp with a bigger range than the ADC input range is applied to the ADC. A slow ramp allows the same code to appear multiple times at the output of the ADC. Then the bins indicating the occurrence number of the codes at the output can be constructed for each code. The transition levels can be written as [27]:

\[ T[k] = C + AH_c[k - 1] \]  

(2.9)

where A is the gain, C is the offset factor that can be extracted from the output data, and \( H_c \) is the cumulative histogram that can be written as [27]:

\[ H_c[j] = \sum_{i=1}^{j} H[i] \]  

(2.10)

where \( H[i] \) is the number of occurrences in bin i.

2.2.1.2.2 Sine Histogram:

Although a ramp histogram is a straightforward method to find the code transitions, the challenge is the difficulty of creating a linear ramp waveform with the desired accuracy. It is more convenient to generate sinusoidal signals by frequency selective filtering. The code transitions for a sine histogram test can be written as [27]:

\[ T[k] = C - A \cos \left( \frac{\pi H_c[k - 1]}{S} \right) \]  

(2.11)

where A and C are the amplitude and offset of the sine, respectively, S is the total number of samples, and \( H_c \) is the cumulative histogram.

After obtaining the transition levels from either of the methods, the DNL can be calculated as:

\[ DNL[k] = \frac{G(T[k + 1] - T[k]) - Q}{Q} \]  

(2.12)

where G is the static gain and Q is the ideal bin width. Then the INL can be
calculated as:
\[ INL[k] = \sum_{i=1}^{k-1} DNL[k] \]  
(2.13)

Generally, the shape of the INL curve defines the distortion at the ADC output. The effect of the INL and DNL on SNDR for an ADC without any other errors can be written as [30]:
\[ SNDR_{INL,DNL} = \frac{(2^N)^2}{1/2 + \frac{1}{4}\sigma^2_{DNL} + \sigma^2_{INL}} \]  
(2.14)

where \( \sigma_{DNL} \) and \( \sigma_{INL} \) are the standard deviation of the INL and DNL. The effect of INL is significantly higher than that of DNL.

### 2.2.2 Sampling

A S&H block is necessary for ADCs in high speed applications since a fast varying signal at the converter input can introduce considerable error during conversion. This can be simply illustrated by fixing the maximum tolerable error during the sampling. To reach an error less than 1LSB for an 8-bit 1GS/s ADC with 800ps conversion time and 200ps tracking time, the maximum input signal frequency can be calculated as follows [31]:
\[ v(t) = V_{LSB}(2^n/2)\sin(2\pi ft) \]  
(2.15)

Then the maximum rate of change is:
\[ \frac{dv(t)}{dt}\bigg|_{\text{max}} = V_{LSB}2\pi f(2^n/2) \]  
(2.16)

and the maximum input frequency is found as:
\[ f_{\text{max}} = \frac{dv(t)}{dt}\bigg|_{\text{max}} = 1.55MHz \]  
(2.17)

where \( \frac{dv(t)}{dt}\bigg|_{\text{max}} \) is \( V_{LSB}/800ps \). This is orders of magnitude lower than the Nyquist frequency of the 1GS/s ADC. A S&H is often unavoidable in a modern high speed ADC.
2.2.2.1 Sampling Noise

The sampled voltage has a noise component stemming from the finite sampling switch resistance $R_{on}$, $4kTR_{on}$, where $k$ is the Boltzmann Constant ($1.38 \times 10^{-23} JK^{-1}$), and $T$ is temperature. The switch resistance $R_{on}$ together with sampling capacitor $C_s$ forms a low-pass network and the power spectral density of the sampled noise is shaped by the transfer function:

$$V_{samp,n}^2 = \frac{4kTR_{on}}{1 + (2\pi f R_{on} C_s)^2} \tag{2.18}$$

Then the total noise power can be written as:

$$\sigma_{samp}^2 = \int_0^\infty V_{samp,n}^2 df = \frac{kT}{C_s} \tag{2.19}$$

The sampled noise component magnitude is independent of the switch resistance. This imposes a limit on the SNDR of the sampling operation. To reduce the effect of the sampling noise, the capacitor size can be increased; however, this limits the input bandwidth by reducing the input pole frequency. Other techniques are also developed for reducing the effect of sampling noise through $kT/C$ noise cancellation, which makes use of amplifiers in the sampling network.

A differential ADC architecture is generally chosen due to resilience against supply noise and maximum allowed voltage swing. The differential sampling scheme is shown in Fig. 2-7 where $R_{term}$ is the termination resistor. In this case, the capacitors con-
nected in series result in a \( C_s/2 \) capacitance leading to \( V_{samp,diff,n}^2 = \frac{2kT}{C_s} \). Assuming a sinusoidal input, SNR can be calculated given a differential peak-to-peak voltage at the input, \( V_{ipp,diff} \):

\[
SNR_{samp} = \frac{\left( \frac{1}{2\sqrt{2}} V_{ipp,diff} \right)^2}{\frac{2kT}{C_s}} = \frac{1}{16} \frac{V_{ipp,diff}^2 C_s}{kT} \tag{2.20}
\]

Equation 2.20 defines a lower limit for the sampling capacitor given a sampling resolution. The upper limit can be found for a given 3dB bandwidth:

\[
C_s = \frac{1}{2\pi R_{term} || R_s f_{3dB}} \tag{2.21}
\]

where \( R_{term} \) is the termination resistor, \( f_{3dB} \) is the 3dB bandwidth and \( R_s \) is the source resistance. In practice, other design limitations, such as area or the integration noise if the sampling capacitor is used in a voltage-to-time converter, also define the capacitor sizing.

### 2.2.2.2 Sampling Time Uncertainty

In practice, the input signal is not sampled at perfect uniformly spaced instants. The deviation can be caused by sampling clock jitter, input dependent sampling time, and the non-idealities from the switch. This can cause noise and distortion on the sampled voltage. Jitter becomes one of the main challenges in modern high speed ADCs for wireline communications. Since the effect of jitter becomes more punishing at high input frequencies, the first rank samplers of the time-interleaved ADCs are severely affected. The timing jitter to voltage error conversion can be done assuming a sinusoidal input and using the slew rate of the signal [33]:

\[
\sigma_{V,jitter} = \sigma_j \left. \frac{dV}{dt} \right|_{rms} = \sqrt{2\pi f_{in} A \sigma_j} \tag{2.22}
\]
where $A$ is the amplitude and $f_{in}$ is the input frequency and $\sigma_j$ is the rms value of the jitter. Then the SNR due to timing jitter can be written as:

$$SNR_{jitter} = 10 \log_{10} \frac{(A/\sqrt{2})^2}{(\sqrt{2} \pi f_{in} A \sigma_j)^2} = 20 \log_{10} \frac{1}{2\pi f_{in} \sigma_j}$$

(2.23)

As can be observed from equation (2.23), the SNR decreases 20dB per decade of the signal frequency given an rms jitter value. Considering the trend of increasing wireline data rates, this imposes strict requirements on the sampling network and the clocking strategy. To demonstrate this, the effect of jitter on sampling ENOB is plotted in Fig. 2-8. Reaching 8 sampling ENOB at 14GHz Nyquist frequency requires less than 36fs rms jitter. The maximum allowable rms jitter must be halved for every frequency octave resulting in 9fs for 56GHz Nyquist frequency. The effect of the quantization can be added to demonstrate the effect of jitter at the output of the ADC. In the SNR equation, noise powers of jitter and quantization noise can be summed:

$$SNR_{jitter,quant} = 10 \log_{10} \frac{(A/\sqrt{2})^2}{\sigma_{V,jitter}^2 + \sigma_{V,quant}^2}$$

(2.24)

The results are shown in Fig. 2-9 for an 8-bit ADC. At low $\sigma_j$ the ENOB is dominated by the quantization noise, while for higher values, the jitter creates the bottleneck.

Figure 2-8: The sampling ENOB vs. the standard deviation of the sampling jitter for three different Nyquist frequencies: fin=14GHz, fin=28GHz, and fin=56GHz.
Figure 2-9: The 8-bit quantized sampling ENOB vs. the standard deviation of the sampling jitter for three different Nyquist frequencies: fin=14GHz, fin=28GHz, and fin=56GHz.

This increases the power dissipation of the clocking circuitry. Especially VCO power dissipation for a given maximum SNR degradation increases abruptly with increasing ADC resolution.

The SNR decrease is calculated using the Nyquist frequency, which assumes the worst case scenario. However, in PAM systems, the majority of the signal power is located in the low frequency band of the spectrum, as can be seen from Fig. 2-1. Similarly, for an OFDM system, the bit loading can compensate for the degradation due to jitter by relaxing the modulation format close to the Nyquist frequency. A specification of the jitter requirement for the RX can be found with a detailed system analysis also considering the modulation format.

2.3 High Speed ADC Architectures for Wireline Communications

Wireline communications demand high sampling rates and medium resolutions from RX ADCs. Among various ADC architectures, flash and SAR ADCs are amenable to time interleaving and provide high sampling rates per channel. CMOS scaling favors
architectures that are digitally intensive. SAR is widely adopted as an efficient topology for wireline receivers due to its digital properties. Emerging ADC architectures also utilize time-based techniques for their distinct scaling advantage by minimizing analog blocks and employing simple digital building blocks. The popularity of time-based techniques has increased in recent years due to the resolution improvement in the time domain by the reduction of single gate delay. The resolution improvement in the time domain becomes even more appealing when compared with the voltage domain under the effect of reduced supply voltages. Digital edge transition of a signal provides more resolution than a voltage step in deep-submicron CMOS processes [35]. This section describes the foundational concepts of the following high speed ADC architectures and their design techniques.

- Flash ADCs
- Successive-Approximation-Register ADCs
- Time-Based ADCs

The experimental results for the latter two architectures are also demonstrated in the following chapters.

2.3.1 Flash ADC

Flash ADCs are among the fastest ADC architectures due to their single step conversion nature employing multiple parallel comparators. The simplicity and fast conversion rate can be attained at the expense of higher power dissipation and area overhead. In Fig. 2-10 a simple n-bit Flash ADC block diagram is illustrated. The input is compared with multiple decision levels created by a resistive ladder. There are a total of \(2^N - 1\) comparators corresponding to each decision level, where \(N\) is the resolution of the ADC. The resistive ladder divides the total voltage range between \(2^N\) levels. The output is thermometer coded, and the comparators with threshold levels smaller than the input return 1s while the others return 0s.

The exponential relationship between the resolution and hardware resources requires
a large number of comparators, leading to a high capacitive input load. This complicates the design of the front end by limiting the bandwidth. In addition, the non-linear input capacitance contributed by the comparators can introduce distortion. Furthermore, the comparators operating simultaneously poses challenges for power delivery, especially for the recent CMOS nodes, where a considerable amount of power must be delivered to a small area. It may require reducing the biasing currents of comparators and trading speed against reliability. Another problem is the kickback noise from each comparator. The kickback noise scales with the square of the number of comparators [36]. Thus, fully flash ADCs are not practical for medium to high resolutions, mainly if they are used in an interleaved manner.

Different approaches to flash conversion have been developed over the years to break the exponential dependence of the resolution to hardware resources. The most prominent ones are:

- Two-step architectures
- Folding
- Interpolating
2.3.1.1 Two-Step Architectures

Two-step architectures break the exponential relationship between the resolution and hardware overhead in a flash ADC by dividing the conversion process into different stages. The cost of this is reduced sampling speeds since two different stages process the data in a sequential manner. The block diagram of a two-step flash ADC is shown in Fig. 2-11. First, the input voltage is sampled by the S&H block. At the end of the track mode, the coarse flash ADC starts converting the input voltage into a digital value (MSBs), which is then converted back to an analog value estimated by the following DAC. This estimate is subtracted from the input, and the residue is fed to a fine flash ADC which outputs the least significant bits (LSBs). Then the total number of comparators can be written as $2^N_c + 2^N_f - 2$ where $N_c$ is the coarse resolution, and $N_f$ is the fine resolution. For an 8-bit flash ADC, this can lead to an approximately 90% decrease in the total number of comparators if both stages resolve 4 bits. However, additional circuitry for the DAC and subtraction operation is needed.

Two-step flash architectures generally suffer from speed-accuracy trade-offs. The DAC settling and the subtraction have to be within a specific time interval. Furthermore, the DAC and the subtractor contribute to the nonlinearity of the converter. If there is no amplification before the second stage, fine flash ADC has to resolve voltages smaller than one LSB. In the case of amplification, the interstage gain must be accurately controlled to fit the fine range to the coarse range. This may necessitate redundancy to relax the interstage gain matching, which eventually increases the number of comparators. To remove the need for the subtractor, a technique called
sub-ranging can be applied \[37, 38\]. This technique first finds in which coarse step the input voltage lies and subdivides that range into fine intervals. Even if the requirement for the DAC and the subtractor can be eliminated, the two-step architecture does not preserve the speed advantage of the single-step purely flash ADC. To reduce the hardware requirement with the single-step conversion, folding and interpolating techniques were developed.

### 2.3.1.2 Interpolation

Interpolation is used for interpolating two reference levels to create an intermediate level that can be used for comparing the signal. This reduces the number of pre-amplifiers depending on the interpolation factor and relies on the fact that pre-amplifiers have a finite gain. The interpolation technique is explained in Fig. 2-12 with \(V_{p1,n1}\) being the differential outputs of the pre-amplifier \(P_1\) and \(V_{p2,n2}\) being the differential outputs of the pre-amplifier \(P_2\). The pre-amplifier outputs change polarities when the input is at decision levels \(V_{d1}\) and \(V_{d2}\). As can be observed from Fig. 2-12, \(V_{n1}\) and \(V_{p2}\) crosses between \(V_{d1}\) and \(V_{d2}\) which provides an additional reference \(V_{int1}\). Hence, if \(V_{n1}\) and \(V_{p2}\) are input to a differential latch, it provides an additional decision. Alternatively \(V_{p1}\) and \(V_{n2}\) can also be used. Thus, the number of comparator pre-amplifiers can be halved with an interpolation factor of two.

The interpolation factor can further be increased by blending the pre-amplifier out-
puts to create more crossings. This can be done resistively [39], capacitively, or actively [40]. Interpolation also helps reduce the maximum DNL error by redistributing the nonlinearity due to offset [36].

Generally, the interpolation factor is limited by the increase of INL [40]. Furthermore, the increased load capacitance and resistance in the interpolation network leads to longer settling times, limiting the speed. Additionally, the gain of the pre-amplifiers must lie within a specific range for a successful interpolation.

2.3.1.3 Folding

Folding is an analog pre-processing technique that allows fine and coarse quantizers to operate simultaneously, preserving the one-step conversion nature of the full flash ADC. The folding block is used before the fine quantizer to generate a folded version of the input voltage range and compute the residue before forwarding it to the fine quantizer, as shown in Fig. 2-13. This eliminates the need for the DAC and the subtractor. The ideal folding characteristic is also shown in Fig. 2-13. In practice, it is challenging to realize piecewise-linear sawtooth functions and folding characteristic is generally produced by differential amplifiers with limited input range as demonstrated in Fig. 2-14, which approximates the triangular waveform. Since the obtained folding waveform has a nonlinear behaviour, it deviates from the triangular waveform. However, zero crossings occur precisely at the location of those of the ideal behaviour; hence do not introduce error. This is exploited in folding interpolating flash ADCs [41].

Folding and interpolating techniques are generally combined as a folding-interpolating ADC [42], where a folding block is used before interpolation. Multiple zero crossings at each sub-divided interval can be obtained by interpolating two folded signals. If the interpolation factor is sufficient to obtain the total resolution, then only the polarity information in these sub-intervals is of interest and can be resolved with comparators. The folding-interpolating ADCs can reduce the power consumption, input capacitance, and hardware overhead without sacrificing the one-step conversion.
Figure 2-13: Folding flash ADC block diagram.

Figure 2-14: Folding circuit with practical folding waveforms implemented by differential pairs.
nature of the flash ADC, hence the speed.

Folding-Interpolating converters suffer from precise placement of zero crossings under PVT variations, especially when the interpolating factor increases. In addition, the increased area gives rise to larger time constants slowing down the operation.

2.3.1.4 Prior Art

Flash ADCs used to be the first choice for GS/s ADCs. Although the advantages of CMOS scaling have increased the popularity of the pipelined and SAR ADCs, the flash ADCs are still fast and reliable choices for wireline communications.

The mixed-signal wireline links inherently employ an array of comparators which are called slicers as discussed in Chapter [1]. Then flash ADC provides a natural way to move to an ADC-DSP based solution using a similar architecture. The interleaved flash ADCs constitute the main block of such a receiver. In [43], a 12GS/s 8-way interleaved fully flash ADC is demonstrated in 65nm CMOS. The digitally tuned minimum sized comparators are used to reduce the input capacitance. [44] shows a 16GS/s 6-bit architecture with 63 background calibrated comparators without pre-amplifiers. The lack of a pre-driver increases the effect of offset but a statistics based offset cancellation technique is applied. In [45], a 10GS/s time-interleaved Flash ADC is discussed, which utilizes sub-ranging to reduce power consumption. A 2-bit coarse flash activates one of the four 4-bit fine flash sections. [46] discusses the design of a 10.3GS/s, 6-bit Flash ADC for 10G Ethernet Applications. Each comparator of the array can be powered down depending on the resolution requirement, and the comparators can be reordered to reduce the offset calibration range. The comparators have kickback cancellation in addition to the offset calibration. A 1-bit voltage domain interpolation is used for a 6-bit flash ADC in [47], which runs at 20GS/s. Another distinct advantage of the flash is that it allows different thresholds to be chosen for different comparators. This is exploited in [48] demonstrating a 64Gb/s PAM4 receiver utilizing a 1-bit folding Flash ADC with adaptive thresholds. This helps increase the power efficiency for low loss channels and reduces the BER.
Besides time interleaved architectures, Flash ADCs can also assist other architectures in realizing an ADC suitable for a wireline RX. With this method, the fast conversion and low latency properties of a flash conversion can be exploited while hardware overhead is reduced. In [49], a 2-bit flash ADC is combined with a 5-bit SAR converter to obtain a 1.75GS/s ADC. The fast flash conversion not only increases the conversion speed but also produces low latency data to clock data recovery (CDR) circuit for better jitter tracking, while the SAR conversion provides power and area efficiency for the remaining LSB bits. [50] demonstrates a similar technique in a time-interleaving scheme where 3-bit Flash ADC is used in common for 4-way interleaved SAR sub-channels. [51, 52] also demonstrate high-speed single channel flash architectures that can be used for wireline communications.

As can be observed from the prior art, the maximum practical resolution attained for the target sampling speeds is around 6-bit. The flash ADCs rely on heavy offset calibration for minimum sized comparators without pre-drivers for extra power efficiency. High reconfigurability is one of the strengths of flash ADCs. The folding and interpolating degrees are kept low to maintain the linearity degradation at a minimum.

2.3.2 Successive Approximation Register ADC

SAR ADC uses a binary search algorithm to approximate the input voltage to the corresponding output value. The block diagram is shown in Fig. 2-15. After the input is sampled, it is compared with the output of the DAC using a comparator. The output of the DAC is updated after each decision step depending on the output value of the comparator, and the operation is organized by a SAR state machine. The DAC is generally implemented as a capacitive array which allows the input to be sampled on the DAC instead of a separate sampling capacitor. In the synchronous case, the clocks for the S&H and the comparator are provided externally. Since the search for the output code is done in a binary fashion (DAC output is updated such that it changes half of the value of the previous DAC output), an ADC of N-bit
resolution requires N comparisons. This provides a slower operation compared to the single step nature of the full flash ADC.

The linearity of the ADC is generally defined by the S&H and the CDAC. The sampling accuracy can be affected by the aperture jitter, charge injection, clock feedthrough, and the varying on resistance of the sampling switch. CDAC static linearity is affected by the mismatch of the unit capacitors. However, in advanced FinFET processes, the mismatch shows a relatively weak effect, especially when considering 6 to 8-bit resolutions targeted for wireline communications. Instead, the linearity degradation stems from the settling errors of the CDAC.

There are several factors affecting the speed. A high comparator bandwidth can reduce the decision time, which speeds up the whole conversion process. For a high bandwidth comparator, the resulting noise is also higher, which is generally the dominant noise source in a SAR ADC. As mentioned previously, each step requires the DAC to settle within a certain accuracy, which also puts an upper bound in the sampling speed. In asynchronous SAR architectures, the internal clock generation can also affect the conversion speed of the ADC. However, in modern SAR ADCs, a big challenge in increasing the sampling speed is comparator metastability. Comparators have a clock to output delay which increases exponentially with shrinking input voltages. Hence, in a case where the input voltage is small enough, the comparator cannot take a decision within the allocated time interval. The probability of metasta-
bility increases from the MSB to LSB since the differential input range at the input of the comparator decreases. However, one compensating feature is that the accuracy degradation also decreases towards LSB decisions, and the last decision is mainly buried under the noise. Hence, the metastability in the LSB range does not affect the accuracy of the ADC considerably. Although it is more probable towards LSB decisions, metastability can occur at each decision and occurs only once. As a result, for synchronous architectures, sufficient time must be allocated for each decision step which slows down the whole conversion process. An asynchronous approach relaxes this limitation. A detailed analysis of metastability in synchronous SAR ADCs can be found in [53].

There are several design techniques targeted for the high-speed medium, resolution SAR ADCs which are explained next. These are:

- Asynchronous operation
- Loop unrolling
- Pipelining
- CDAC redundancy
2.3.2.1 Asynchronous operation

Asynchronous SAR is the dominant choice for high speed SAR applications thanks to high sampling rate and power efficiency. The clocks for the comparator and internal organization of the conversion are generated within the ADC instead of distributing high speed clocks to the ADC externally as in the case of a synchronous SAR ADC. A block diagram is given in Fig. Most of the time, only the sampling clock is distributed externally. The output of the comparator is observed by a clocking block. When the decision is taken, the clocking block initiates the comparator reset, and when the reset is completed, it starts the next decision step. Optionally the SAR state machine can also communicate with the clocking block to organize the memory capture.

The asynchronous SAR architecture allocates each decision step the exact time required. In the case of lower input voltages to the comparator, this time is longer, and in the case of higher voltages, this time is shorter. Compared to synchronous SAR, where each decision step must be adjusted for the worst case scenario, asynchronous operation increases the sampling speed considerably. Furthermore, the time left after the conversion is finished until the next sampling instant can be utilized for resolving any possible metastable decision. It can be demonstrated that if sufficient time is allocated, then the bit errors mainly stem from the thermal noise in asynchronous SAR ADCs. Thus, asynchronous SAR ADCs provide a time efficient operation by minimizing the unutilized time intervals. Another advantage is that they do not require the distribution of high speed clocks. Since an N-bit ADC requires N conversion steps, the internal clock frequency must be N times the sampling clock frequency assuming the sampling clock is provided externally. In a synchronous approach, this means high speed clocks must be routed over distance which increases the power consumption and reduces the routing resources. These advantages lead to the broad adoption of asynchronous SAR ADCs, especially for time interleaved architectures.
2.3.2.2 Loop Unrolling

Loop unrolling is a speed enhancement technique that uses N comparators for an N-bit SAR instead of a single one [56]. Each comparator takes one decision and operates in a domino sequence from MSB to LSB. This allows the next decision cycle to start without waiting for the reset phase from the previous comparison. Instead of N memory cells, now N comparators can hold the decisions to control the CDAC switching activity. All the comparators are connected to a common CDAC. This can lead to an increase in the sampling rate if the conversion process is limited by the reset time of the comparator. In other words, if the CDAC settles within an accuracy of 1/2LSB before the end of the reset phase of the comparator. However, in modern ultra-scaled FinFET technologies, highly resistive lower metal layers increase the time constants of CDAC settling [57]. Loop unrolling increases the area, leading to longer routing paths, and fortifies this effect. The design choice must be made whether the sampling rate of the design is settling limited or comparator reset phase limited. The target resolution also plays a role in architecture definition since a lower resolution ADC has more relaxed settling requirements. A recent adoption of this technique is applied to reach 1.75GS/s per sub-ADC sampling speed for a 6-bit ADC in a 112GS/s wireline receiver [58].

2.3.2.3 Pipelining

Although pipelining is not specific to a SAR ADC architecture, it is commonly used for increasing the sampling rates for SAR ADCs [59, 60]. The main idea of pipelining
is to divide the conversion process into multiple steps that can run in parallel, as shown in Fig. 2-17. Since the conversion must only be partially finished before the next value is sampled, the sampling speed is increased. In addition, the noise and the linearity specifications of the coarse and fine stages can be optimized separately to reduce the power and area.

One major challenge associated with the pipelined SAR ADC is the need for residue amplification between the stages. The residue amplifier gain must be fine tuned to prevent any distortion. This causes additional power dissipation. Considering that the gain varies in different PVT conditions, a calibration circuit may be required. The accuracy of the interstage gain can also be relaxed by introducing redundancy in the second stage. Several residue amplifier topologies have emerged lately. Ring amplifiers are inverter based closed loop structures with stabilization. They offer high power efficiency with wide output swings and scale well with the technology [61, 62, 63]. This increased the resolution of pipelined ADCs with GS/s sampling rates. However, another problem associated with pipelining is the extra latency introduced. Since the conversion of one sample is completed in M cycles where M is the number of stages, the latency increases by the degree of the pipelining. This leaves a shorter time for critical DSP functions like FEC. Additional latency, together with the extra calibration required for interstage gain, limits the usage of pipelining in fast interleaved ADCs with high interleaving factors.

2.3.2.4 Redundancy

Redundancy is a technique to introduce error tolerance in ADCs either through extra quantization steps or additional hardware. The most common method for SAR ADCs is CDAC redundancy. The purpose of introducing redundancy in a CDAC is to compensate for the settling errors. Previously this is stated as one of the speed limiting factors in modern SAR ADCs. The time it takes for DAC to settle within 1/2LSB can be written as [64]:

\[ T_{settling} = (n + 1)\tau ln(2) \]  

(2.25)
where $\tau$ is the settling time constant. To shrink this time interval a CDAC with a radix less than 2 can be implemented. The main idea is to overlap the consecutive decision ranges to increase the error tolerance such that full settling of the CDAC is unnecessary. In a binary weighted CDAC, once a search range is eliminated by a decision, it is not considered again. However, overlapping the search ranges provides the recovery of errors up to a certain amount defined by the radix. The concept is illustrated in Fig. 2-18 for a 4-bit ADC with a radix of 1.78. An extra decision step is required for the full conversion. The design decision is whether the allowed limited settling time compensates for the additional time consumed by extra decision steps. This is critical, especially regarding asynchronous SAR ADCs, where the internal clock generation also consumes a considerable portion of the conversion period.

### 2.3.2.5 Prior Art

SAR ADC is a mature architecture that offers good power efficiency and scalability. Asynchronous SAR is the dominant choice as a sub-ADC architecture for the time interleaved ADCs today. [65] proposes a power efficient loop unrolled SAR architecture achieving 16.6fJ/conv-step while running at 1GS/s. [66] demonstrates an 8-bit partially interleaved SAR with noise reduction technique reaching 12fJ/conv-step.
shows a 2.4GS/s SAR operation with background calibration, which starts the
conversion with 1-bit per step and then switches to the 2-bits per step for the re-
mainder of the conversion for extra speed. uses a charge injection based DAC
instead of a conventional CDAC which reduces the area through charge injection cell
sharing and increases the speed by shrinking the settling times. shows one of
the early examples of highly interleaved ADCs reaching up to 90GS/s with a 64 times
interleaved SAR. In , a 24 to 72GS/s 8-bit interleaved SAR is demonstrated with
64 sub-ADCs. It shows an SNDR of 30dB at 36GHz input frequency. shows
another interleaved architecture with 40 to 97GS/s sampling rate showing excellent
power figures. It has 96fJ/conv-step at 36GHz input frequency.

2.3.3 Time-Based ADC

A time-based ADC utilizes time for processing the input information. The input
data in most of the communication systems today is encoded in voltage levels. To
use time-domain techniques, this input quantity is converted to a time or frequency
domain quantity first. This can be done through a voltage-to-time converter (VTC)
or a voltage or current controlled oscillator (VCO or CCO). In the latter case, the
frequency-voltage or current characteristic is highly nonlinear. Even in the ideal case
where the mismatch and noise are neglected, this gives rise to harmonic distortion
and frequency spurs. For specific applications, the distortion can reside out of the
frequency band of interest; however, in most cases, it degrades the SNDR consid-
erably. Hence, for medium resolution ADCs, either circuit techniques to cancel the
nonlinear behaviour partially or a linearization scheme in the digital domain address-
ing this nonlinear characteristic must be introduced. For highly interleaved ADCs,
this increases the design complexity or latency, leaving a shorter time for FEC and
DSP operations. As a result, VTC-TDC based inherently linear time-based ADCs
are deemed more suitable candidates for highly interleaved RX ADC, and they are
covered in this section. For this thesis, they are referred to as time-based ADCs.

A generic time-based ADC block diagram is shown in Fig. 2-19. After the input
voltage is sampled by an S&H, it is converted to a time domain quantity which is generally a pulse width or the delay between two pulses. This time-domain quantity is then mapped to the digital domain by a time-to-digital converter (TDC). Depending on the output code of the TDC, an encoder may be required to convert the output to binary bits.

Time-domain techniques are in use for a long time in areas like particle detectors [72, 73], light detection and ranging (LIDAR) [74], all digital phase-locked-loops (PLLs) [75, 76]. The popularity of time-domain techniques in data conversion systems has seen a considerable increase in the last decade. The driving factor for this is the CMOS scaling since time-based ADCs allow digital intensive designs. The reduction of supply voltages gives rise to an SNR decrease in voltage-domain circuits. On the contrary, decreasing gate delays increases the timing resolution and allows faster conversion in the time domain. Another point is related to the high parasitic resistance, and capacitance of the interconnects in more advanced technology nodes. The area reduction, thanks to the digital style layout that can be realized by time-based ADCs in this case, helps reduce the wiring parasitics and decreases the performance penalty. Since they are digital intensive, the time domain ADCs are also amenable to synthesis. They may be integrated into a design flow similar to a digital design flow which can reduce the physical design time considerably. The building blocks of time-based ADCs are discussed next.
2.3.3.1 Time-to-Digital Converters

2.3.3.1.1 Flash TDC

Flash TDCs are the most straightforward TDC architectures that can be built with a delay line and time comparators. An example is demonstrated in Fig. 2-20. In analogy with flash ADC architectures, the reference delays are generated with a delay line instead of reference voltages generated by a resistive ladder. Time comparators are used instead of voltage comparators to decide which input edge comes earlier. A reference pulse can be input to the delay line, and an input pulse can be used for clocking the time comparators. The time from the rising edge of the reference pulse to the rising edge of the input pulse indicates the time to be resolved. The output sequence depends on the length of this time interval. In other words, the first $N$ comparators will output 0 and the remaining ones 1 if the time interval is between $N\tau$ and $(N + 1)\tau$.

The exact delay throughout the delay line is subject to PVT variations. This can be controlled precisely with a phase detector and a filter to construct a delay-locked loop (DLL). However, the single delay step $\tau$ is still subject to variations and introduces DNL errors.

As in the case of voltage-domain flash ADCs, the number of comparators and delay elements scales exponentially with the resolution. Especially long delay lines are not feasible not only in terms of area and power dissipation but also linearity. The
effect of mismatch and noise accumulates through the delay line resulting in high values of INL. The solution is similar to the ones used in voltage-domain flash ADCs which require the division of the delay line into fine and coarse stages or necessitate hierarchical TDC architectures.

### 2.3.3.1.2 Two-step TDC

In order to break the exponential relationship between hardware resources and resolution, the TDC can be divided into two blocks. The coarse time information can be obtained from the first stage, and then the remaining time residue can be forwarded to the second stage to be resolved as fine bits. To resolve the residue, the second stage must have finer time steps. This can be accomplished by using delay elements with smaller delays in the fine TDC or amplifying the time between the two stages.

Two-step time-to-digital conversion brings challenges related to residue generation and amplification. Since time is not a quantity that can be stored, the generated residue can be converted into the voltage domain to be stored in between two stages. An additional time-to-voltage conversion step is needed for this, which naturally reduces the effectiveness of the time-domain technique. An alternative way is to let the pulse propagate through dummy delay lines to store the residue; however, this increases the power dissipation considerably. Another challenge is to obtain an accurate gain for the time amplifier across the PVT variations. The harmonics contributed by the time amplifier further disturb the spectral purity. On the other hand, a two-step architecture reduces the number of delay elements and time comparators. Moreover, it allows a pipelined conversion that can reach high sampling rates.

### 2.3.3.1.3 Successive-Approximation-Register TDC

SAR TDC is another architecture where the number of delay elements and time comparators does not scale exponentially with the resolution. The scaling in this case is linear thanks to the binary search algorithm similar to a SAR ADC. An example architecture is given in Fig. 2-21. Each stage consists of time comparators,
Figure 2-21: SAR TDC architecture with binary encoded delays.

delay paths, and multiplexers. Depending on the comparator decision at each stage, either the positive pulse \( V_{pp} \) or the negative pulse \( V_{pn} \) is delayed. The delays \( T_{dn} \) are binary weighted between the stages. Multiplexers choose which pulse to delay. Hence, an N-bit SAR TDC requires only N stages.

SAR TDCs are efficient architectures that reduce hardware resources. They also allow sub-gate delay resolution since the delay difference between the two paths can be used as a unit delay. This is true, especially in the LSB stages. Additional delays must be placed to accommodate the comparator decision time and metastability in addition to multiplexer combinational delay. This increases the power dissipation and complexity. Furthermore, the delay paths must be calibrated to restrict DNL errors. The sampling speed of a SAR TDC is limited by the longest path of delays in addition to the total multiplexer and comparator delays contributed at each stage.

2.3.3.1.4 Ring Oscillator (ROSC) based TDC

A ROSC-based TDC has a similar topology to a two-stage TDC. The difference is that the first stage defines the fine bits while the second stage decides on the coarse bits. It can be thought of as accumulating fine time steps to achieve a course time step in contrast to a two stage TDC where the coarse time step is divided into finer time steps. This eliminates the necessity of generating and processing the residue between the stages. An example architecture is shown in Fig. 2-22. Leading pulse \( V_{pp} \) can be injected into the ROSC while the lagging one \( V_{pn} \) is connected to the clock inputs of the samplers. The leading pulse triggers the oscillation in the ROSC.
Figure 2-22: A ring oscillator based TDC architecture.

and is delayed by the delay elements. Each time it travels around the ROSC, the counter is incremented by one. When the lagging pulse arrives, it samples the state of the counter and the ROSC. The state of the ROSC represents the residue from the incomplete cycle and constitutes the fine bits, while the output of the counter represents the coarse bits.

The ROSC based TDC resolves issues with residue generation and amplification by utilizing a hierarchical topology. This not only reduces the complexity but also decreases the power consumption. It further eliminates the need for interstage gain calibration since each ROSC cycle exactly equals one integer count. The ROSC topology limits the INL errors emerging due to the accumulation of mismatches of the delay cells since the INL shows a cyclic behaviour. Another advantage is that it occupies a small area because a long delay line is avoided, and a ROSC consists of compact digital-style delay cells. The decoupling of the ROSC must be sufficient enough to reduce the effect of supply noise on the oscillation frequency which can reduce the area advantage since a high number of decoupling cells may be required. The resolution of this architecture is limited by the delay of the single delay cell. While designing the ROSC, undesired mode of oscillations must be prevented. Furthermore, the ROSC can not be arbitrarily short since the counter is clocked with the oscillation frequency, and shorter rings provide oscillation frequencies that can be higher than the counter can support.
2.3.3.2 Time-to-Digital Converter Resolution Techniques

The biggest challenge of the TDC architectures is reaching to sub-gate delay resolution in a power and area efficient manner that is robust against PVT variations. This relaxes the resolution-sampling speed trade-off and allows TDCs to reach medium-high resolutions at GS/s sampling rates which is critical for modern high speed wireline communications. Several techniques that are widely used are explained below.

2.3.3.2.1 Vernier Delay Line

Vernier Delay Line employs two different delay lines with different delays as shown in Fig. 2-23. The lagging pulse $V_{pn}$ is routed to the faster delay line while the leading one $V_{pp}$ is routed to the slower delay line. Arbiters are used at each stage to decide which edge comes first. The delay difference between two delay cells $T_{d1} - T_{d2}$ represents the LSB time step which offers a sub-gate delay resolution.

The cost of achieving a sub-gate delay with a Vernier Delay Line is doubling the power consumption and area due to the requirement of two different delay lines. Also, calibration may be required to have a well defined time step across PVT conditions.

2.3.3.2.2 Pulse Shrinking Delay Line

Pulse shrinking delay line is an alternative way to reach sub-gate delay timing resolution. The time quantity is encoded in the pulse width of a pulse $V_p$, which
Figure 2-24: Pulse shrinking TDC architecture.

propagates throughout the delay line. The delay line is designed such that the pulse shrinks at each stage by controlling the rising or falling time of the edges. The pulse vanishes at the stage, which is proportional to the input time.

2.3.3.2.3 Phase Interpolation

Phase interpolation is similar to interpolation in flash ADCs to obtain additional reference levels without increasing the number of preamplifiers. The primary motivation of interpolation in time-based ADCs is to obtain intermediate phases, which divides the unit delay into finer time intervals to obtain sub-gate delay resolution. An inverter based interpolation technique is shown in Fig. 2-25. As can be observed, an intermediate phase can be obtained by blending the outputs of $I_1$ and $I_2$. This phase gives an additional zero crossing. By adding more levels, the interpolation factor can be further increased. The price paid for the additional inverters is the increase in the power dissipation and the area. Furthermore, excessive capacitive loading on the delay line can extend the time steps. Since the delay line can not be too long due to linearity and jitter requirements, $T_d$ is likely to be higher than a single gate delay which makes multiple levels of interpolation necessary to reach sub-gate time delay. One significant advantage provided by the interpolation is that if it is used in a two stage TDC, then the interstage gain is automatically defined by the interpolation factor. This leads to a PVT robust solution without any need for calibration of the time step [80].
2.3.3.2.4 Time Amplification

Time amplification can eliminate the need for reaching sub-gate delay resolution. Since the input time is amplified, a coarse time step can be used to resolve the input given a certain TDC resolution. However, it also means that the conversion takes a longer time. This can be used in a time-interleaved architecture where the sub-ADC conversion rate is defined by the interleaving factor, or it can be used in a pipelined time-based ADC which can afford a slow conversion rate without impairing the sampling rate at the cost of increased latency.

There are several approaches to time amplification. The time amplifier can employ open loop slew rate control [81], a ring oscillator [82]. [83] demonstrates a time amplifier with a small linear input range utilizing an SR latch. [84] shows a pulse train based technique. Similar to the voltage domain amplifiers, the linearity of the time amplifier is critical for the conversion process, which is hard to preserve at high gain. Furthermore, some of the amplifier architectures inherently convert the time back to voltage for amplification which leads to extra power dissipation.

2.3.3.2.5 Folding

Folding is another technique inherited from voltage-domain flash ADC applications. A ring oscillator naturally provides folding capability due to its cyclic behaviour. A
A folding TDC architecture with ideal folding waveforms.

Figure 2-26: A folding TDC architecture with ideal folding waveforms.

A ring oscillator with N stages can output 2N different states in one cycle. Input can be mapped to any of these states without the cycle information. The cycle information can be resolved by the coarse TDC. The coarse TDC can be realized by one of the discussed TDC architectures. [85] demonstrates a folding architecture where the coarse TDC is realized by differential delay lines. An example block diagram illustrating the folding principle is shown in Fig. 2-26 where the fine TDC defines the location within one fold region and the coarse TDC chooses the fold region.

Folding provides an area efficient way to increase TDC resolution. Similar to a two stage TDC, it breaks the exponential dependence of resolution on hardware resources. Compared to voltage domain folding, which is normally realized by differential amplifiers with limited input voltage range, time domain folding done by a ROSC is inherently linear, assuming no errors from the ring oscillator output. Moreover, the folding factor can be increased at the cost of sampling time, provided that the coarse TDC has enough dynamic range.

2.3.3.3 Voltage-to-Time Conversion

Voltage-to-time converters (VTCs) convert the input voltage to time domain quantities. There are several approaches to voltage-to-time conversion. Input voltage can be used to control the current in a current-starved inverter topology which adjusts the delay between two pulses [85]. The sampled input voltage can also provide a starting point for a linear ramp with a constant slope where the ramp duration will
be proportional to the input voltage for a given slope [86]. The input voltage can be compared with a continuously generated ramp [87]. The first two are of interest considering the resolution and the sampling speeds required by the wireline RX.

### 2.3.3.3.1 Input Voltage to Control the Current

A typical implementation of this type of VTC is given in Fig. 2-27. Input voltage controls the strength of the discharge path. A smaller sampled voltage leads to a slower falling edge at the output, which is converted to a delay by a threshold detector, while the rising edges have the same strength.

This type of VTC architectures offer high speeds. However, the voltage-current characteristic of a transistor is highly nonlinear. This permits the resolution achievable by this technique around 6-bits. The linearity can be improved in the analog domain by introducing additional non-linearity with the opposing characteristic to cancel out the existing effect using circuit techniques or processing in the digital domain.

### 2.3.3.3.2 Input Voltage as a Start Voltage

In this technique, the input voltage defines the starting point of a voltage ramp that is created by the current from a current source. The working principle is shown in Fig. 2-28. After sampling the input on a sampling capacitor, the current source starts discharging the sampling capacitor, and when the voltage $V_c$ decreases below a threshold level, a pulse is triggered.
Figure 2-28: The VTC architecture where the input voltage serves as the voltage ramp starting point.

The advantage of this technique is that it is inherently linear since the current from the current source is constant if it stays in a linear operating range. Furthermore, it also allows GHz operating speeds while the gain is easily adjustable by controlling the current that defines the slope using a current DAC.

2.3.3.4 Prior Art

Time-based ADCs have architectures that mainly differentiate with VTC and TDC implementations.

In [88], a time-based ADC is implemented with a two step coarse-fine TDC and a high linearity VTC achieving 18.7fJ/conv-step. In [89], a hybrid ADC running at 4GS/s is proposed. The first stage operates as a SAR, and after the residue is obtained from the first stage conversion, the second stage resolves it in the time domain using a ring-oscillator based TDC. [90] shows a 10GS/s two times interleaved SAR TDC based ADC. SAR TDC utilizes selective delay tuning cells to create the binary weighted delays, and the sampling speed is increased through pipelining the operation of comparators at each delay stage. [91] shows a background skew calibration technique for a 20GS/s interleaved time-based pipelined ADC. [92] achieves a 6GS/s sampling rate with a flash ADC with latch interpolation technique exploiting the fact that propagation delays of the dynamic comparators are input voltage dependent. In [93], voltage is converted to time by voltage-to-time amplifiers, and the outputs of the amplifiers are folded further to reach a finer resolution. [94] discusses the implementation of 4x time interleaved ADC where the interleaving is done in the time
domain by multiplexing the pulses obtained from the VTC. Each channel employs a coarse-fine TDC approach, and the VTC is common for all of them. Fine TDC is an interpolated gated ring-oscillator that also realizes the time domain folding. [94] introduces a 10GS/s 4x interleaved ADC utilizing inverter based interpolation technique. In [80], a 4x interleaved coarse-fine TDC based ADC is implemented. Fine TDC is obtained by 16x time interpolation of the coarse TDC, eliminating the need for interstage gain and time step calibration. [95] presents the remainder number system quantization technique for a flash time-based ADC. [96] uses voltage and time domain stages in a pipelined manner to reach 14-bit nominal resolution. [97] presents a 5-bit 5GS/s ADC with flash TDC and a dual edge triggered VTC.

2.3.4 Time interleaving

The wireline link receivers demand sampling rates that can not be achievable by a single ADC. One standard method to increase the sampling rate is time interleaving, where multiple ADCs sample the input in an interleaved manner and then operate in parallel to resolve the input. Any ADC architecture discussed previously can be used as a sub-ADC in a time-interleaved architecture.
A generic architecture of an interleaved ADC is shown in Fig. 2-29 where N sub-ADCs with a sampling rate of $f_s/N$ operate in a parallel manner that results in an aggregate sampling rate of $f_s$. This kind of architecture is called direct interleaving, where all the ADCs are connected directly to the input. It does not require an interleaver which reduces the area and power consumed, yet its performance is limited. First, all of the sub-ADCs load the input network in addition to wiring capacitance and inductance. If the interleaving factor is high, this leads to a severe limitation in the input bandwidth. Second, all the ADCs see the full input bandwidth leading to the stringent T&H requirements for all the sub-ADCs causing a higher clocking power. These limit the use of direct interleaving only for architectures with small interleaving factors.

Another approach is to use hierarchical interleaving. This kind of approach involves multiple stages. A generic example with two stages is shown in Fig. 2-30. The input is first sampled into the M first rank sampling capacitors. Then the sampled voltage is buffered to drive the N ADCs in the second rank. This leads to a total interleaving rate...
of $M \times N$. As can be observed from Fig. 2-30, the input is shielded from the capacitive loading of a high number of ADCs. Only the first ranks load the input, which leads to an increased bandwidth. The number of jitter sensitive high speed clocks is also reduced to the M. Second rank sub-ADCs only see a sampled input which greatly relaxes the requirements for the second rank S&H. These benefits are obtained at the cost of increased hardware complexity. Additional buffers and first rank T&Hs are required, which gives rise to extra power dissipation and area. Furthermore, the interleaver also contributes to the noise, which reduces the overall SNR. It is essential to ensure adequate linearity at the output of the first interleaver stage to maintain INL performance.

The real challenge of interleaving stems from the mismatches in the interleaver and sub-ADCs. The errors are caused mainly by offset, gain, timing skew, and bandwidth mismatches between the sub-channels. These errors introduce frequency spurs in the output spectrum leading to distortion and decreased linearity. The mismatches must be corrected using a background or foreground calibration strategy.

2.3.4.1 Offset Mismatch

Offset mismatch creates frequency spurs located at (99):

$$f_{\text{spur,ofs}} = k \times \frac{f_s}{N}, k = 1, 2, .. N$$

(2.26)

where $f_s$ is the sampling frequency, and $N$ is the interleaving factor. If one sub-ADC channel has an offset, it is observed at the output with a periodicity of $N \times T_s$. Additionally, the aliases from the other Nyquist regions fold into the first zone. The effect of offset is constant and independent of the input amplitude and frequency.

2.3.4.2 Gain Mismatch

Gain mismatch creates frequency spurs located at (99):

$$f_{\text{spur,gain}} = k \times \frac{f_s}{N} \pm f_{\text{in}}, k = 1, 2, .. M$$

(2.27)
This can be thought of as the input signal amplitude modulated by the gain error that repeats with a frequency \( f_s/N \) due to the cyclic nature of the interleaving \([100]\). As expected, the total power of spurs depends on the signal amplitude and gain error standard deviation.

2.3.4.3 Timing Skew Mismatch

Another error source is the timing skew mismatch in the ADCs. The frequency spurs due to timing skew mismatch occur at \([99]\):

\[
f_{\text{spar,tm}} = k \times f_s/N \pm f_{\text{in}}, k = 1, 2, \ldots N
\]

(2.28)

Timing skew mismatch is observed at the output with a periodicity of \( N \times T_s \) similar to the gain mismatch. It can be seen as the phase modulation of the input signal \([100]\). Since a higher timing skew with respect to signal period introduces a more significant error on the sampled voltage, the power of frequency spurs is proportional to the input frequency \( f_{\text{in}} \) and the standard deviation of the timing skew.

2.3.4.4 Bandwidth Mismatch

The paths from the input of the interleaved ADC to the output of each sub-ADC show a different frequency response. This leads to magnitude and phase differences at the output of each sub-ADC. Since the nature of these errors is similar to the timing skew and the bandwidth mismatch, the frequency spurs occur at the same locations:

\[
f_{\text{spar,bw}} = k \times f_s/N \pm f_{\text{in}}, k = 1, 2, \ldots N
\]

(2.29)

Bandwidth mismatch is more observable at high frequencies close to 3dB frequency since the low frequency input signals are affected less by the bandwidth limitations.
Chapter 3

Time-based ADC

This chapter presents an 8-bit time-based ADC running at 1GS/s with 0.7V supply and at 1.25GS/s with 0.8V supply. The ADC is implemented in 5nm CMOS technology and achieves 16.6fJ/conv-step and 20.3fJ/conv-step Walden FoM respectively. The total active area is $313 \mu m^2$.

3.1 Motivation

As new process technologies enable further CMOS scaling, circuit performance benefits from improvements in speed, area, and power dissipation. However, besides the benefits, scaling also brings new challenges to be addressed with innovative techniques. Threshold voltages remain relatively the same after the transition to FinFET technology; however, the supply voltage continues to decrease [57], which is mainly driven by mobile applications. This limits the achievable voltage swing and dynamic range. The interconnects in lower level metal layers become highly resistive, and the RC time constants can increase the settling times. It also makes power delivery more challenging. The electromigration rules limit the allowed current and achievable speed.

These challenges can be alleviated by compact and digital intensive designs. Time-
Figure 3-1: Resolution and sampling speed trade-off in time-based ADCs.

Based ADCs naturally allow this by utilizing logic gates and minimizing the requirement of precise analog blocks. Furthermore, gate delays are shrinking, which allows finer quantization and relaxes the sampling speed and resolution trade-off. This is one of the major challenges of time-based ADCs, since all time steps have to fit in a conversion period, indicating exponential scaling of the conversion period with the resolution. This is illustrated in Fig. 3-1. Time domain techniques like phase interpolation further help refine the time steps and allow exceeding the sampling rates of widely used voltage domain ADCs.

From a system level perspective, the next generation wireline receivers demand 6-8-bit resolution for modulation formats PAM4 and higher [101]. Sampling rates are dictated by the data rate and the modulation format. They must exceed 100GS/s for the next generation links. This means a high interleaving factor is required, which makes area an important concern. The state-of-the-art sampling speeds for single channel sub-ADCs can reach above 5GS/s. It can be increased by design techniques such as pipelining to reduce the interleaving factor; however, the interface with DSP poses a latency bottleneck in that case since the DSP can not be clocked at these clock rates. Moreover, power efficiency starts to suffer when operating close to technology limitations. In the 5nm CMOS technology used, the power performance area (PPA) sweet spot for the DSP resides within 0.7-0.8V supply and 1-2GHz clock frequency range. If the sub-ADC clock rate and supply match the DSP, latency is reduced, and level shifting is avoided yielding also a power efficient solution. Thus, this work targets 1GS/s and 1.25GS/s sampling speeds with 8-bit nominal resolution, and the priority is to reach a compact, digital intensive and power efficient solution. For this, the proposed ADC introduces three techniques:
1. Bipolar ramp-based voltage-to-time converter (BVTC) to eliminate the reference voltage and to allow a wide input swing of $0.75V_{pp,diff}$.

2. $2 \times$ interpolating sense-amplifier-latches (SAL) to reach a sub gate delay of 2.3ps in a power and area efficient manner.

3. Redundancy to accommodate possible wrong sign bit decisions of the sign bit comparator, allowing a minimum size design without any calibration.

### 3.2 Implementation

#### 3.2.1 Top Level Block Diagram

The working principle of the implemented ADC is similar to the one given in Fig. 2-19. After sampling, the input voltage is converted to time using a voltage ramp VTC using the sampled voltages as a starting point. Then, this time quantity is mapped to the digital domain by a ring oscillator based TDC.

The overall ADC architecture is shown in Fig. 3-2. Transmission gate switches with feedthrough cancellation sample the differential input with VDD/2 common mode into 25fF sampling capacitors. A sign bit comparator detects the polarity of the input signal, and the output serves as the MSB bit used for folding and configuring the BVTC. Thanks to folding, the resolution is doubled. However, the folding point must be determined to compute the sign-corrected converter output. The BVTC operates with two ramps with opposite polarities, where one ramps down from the higher sampled voltage while the other ramps up from the lower sampled voltage. At the start of the ramp, a feed-forward ring oscillator (ROSC) implemented with eight delay cells starts oscillating from a reset state. The differential crossing of the ramps is detected by a zero-crossing detector (ZCD), which stops the ROSC. Each ROSC cycle increments a 4-bit asynchronous integer counter [102]. The integer counter is driven by one of the ROSC outputs (Vrop<0>) that is conditioned by a regenerative buffer (BUFLAT). The difference between the two consecutive counts represents four
Figure 3-2: The time-based ADC top level block diagram.
binary-coded coarse bits. The ROSC state after stopping the oscillation represents the fine bits. The fine bits are the residual fractional part of the last incomplete oscillation. No interstage gain adjustment is required because an integer step exactly equals a ROSC cycle. The ROSC state is $2 \times$ interpolated directly by the sense amplifier latches (SAL) to obtain eight cyclic thermometer-coded fine bits. The fine bits correspond to four binary bits, and together with the four coarse bits and the sign bit, the converter reaches 9-bit nominal resolution. The extra ADC bit is needed to accommodate the delay between the start and stop signals applied to the ROSC when the input signal is zero. The SALs sampling the fine and the coarse bits have a full conversion period to make their decision, which diminishes metastability events compared to a SAR ADC. The sign bit comparator is the most critical block in case of metastability. Thus, a time-out mechanism is implemented that forces a decision within the allowed time interval.

### 3.2.2 Timing

The timing of the operation is illustrated in Fig. 3-3. The conversion interval can be divided into three phases.

**Phase I:** Phase I lasts $T_{samp}/8$ seconds where $T_{samp}$ is defined as the sampling period of the ADC. In this region, the input is tracked, and the sign bit comparator is reset.
The BVTC currents are steered to VDD and GND.

**Phase II:** Phase II lasts $2T_{samp}/8$ seconds. At the beginning of this phase, the input is held, and the comparator clock is asserted, which initiates the sign bit decision process. In the first half of this phase, ZCD is reset, and in the second half, ZCD is enabled prior to the start of the voltage ramps. The ROSC is also enabled in the second half such that it reaches a steady state before the ramp starts.

**Phase III:** Phase III lasts $5T_{samp}/8$ seconds. This is the phase where the sampled voltages are ramped up or down. Whenever the zero crossing is detected by ZCD, the ROSC is stopped. Clock for sampling the integer counter and ROSC is generated after ROSC is stopped and the fine and coarse bits are sampled. The sampling capacitors are shorted at the end of this phase to reduce the memory effect.

The output bit mapping is also shown in Fig. 3-3. Interpolated 8-bit cyclic-thermometer code from the ROSC gives four binary bits (LSB[0:3]) since it allows counting sixteen different states. Another four binary bits are obtained (MSB[4:7]) from the integer counter. Finally, the sign bit is shown as the most significant (Sign Bit[8]) bit.

### 3.2.3 Voltage-to-Time Conversion

#### 3.2.3.1 Sampling Switch

The differential sampling switch is shown in Fig. 3-4. Transmission gate switches are used for compact area and for 0.4V mid-rail input common-mode voltage ($V_{cm}$). Cross coupled switches are added to increase the isolation when the switch is off. The sampling capacitors are sized targeting around 8.5 sampling ENOB with the maximum swing according to the equation 2.20. This gives 25fF as the capacitor size. Then the switches are sized for optimizing the sampling ENOB at the maximum input frequency.
3.2.3.2 Bipolar Voltage-to-Time Converter (BVTC)

In this work, a bipolar VTC approach is utilized where the voltage ramps of the opposite polarity are created on the sampling capacitors. The polarity of the voltage ramps is defined by the sign bit comparator which takes the polarity decision for folding. Fig. 3-5 illustrates the proposed and the conventional VTC approach through voltage ramp waveforms. After the differential voltage is sampled, the BVTC is configured such that the higher voltage ramps down and the lower voltage ramps up by using the polarity information. The time of interest is when the two voltage ramps cross which is detected by the ZCD that controls the ENB pulse driving the ROSC. The conventional approach using a pseudo-differential VTC is also shown in Fig. 3-5. For the conventional approach, the voltage ramps are moving in the same direction, and they are compared against a threshold. Whenever each ramp crosses the threshold, a pulse is triggered.

Pseudo-differential VTC relies on a detection threshold level which is depicted as $V_{det,th}$ that must be placed outside the input sampling window. This puts a limit on the possible voltage swing that can be used. As shown in Fig. 3-5, there should be enough margin between the lowest possible sampled voltage to the $V_{det,th}$, which is labeled as unutilized input range. This eventually creates a dead time interval.
Figure 3-5: Voltage ramp waveforms of the BVTC and the conventional pseudo-differential VTC.

\[ T_{\text{dead}} \text{ which does not contribute to the conversion process. Furthermore, achieving an arbitrarily small } V_{\text{det,th}} \text{ is not straightforward which may necessitate a common mode voltage jump at the input. Assuming perfect matching for the conventional architecture, the output time } T_{\text{out, } pd} \text{ and the dead time interval } T_{\text{dead}} \text{ can be written as:} \\

\begin{align*}
T_{\text{out, } pd} &= \frac{V_{\text{in, } \text{diff}} C_s}{I_{\text{vtc}}} \tag{3.1} \\
T_{\text{dead}} &= \frac{(V_{\text{in, } \text{min}} - V_{\text{det,th}}) C_s}{I_{\text{vtc}}} \tag{3.2}
\end{align*}

\]

where \( I_{\text{vtc}} \) is the VTC current defining the voltage ramp slope, \( C_s \) is a single sampling capacitor, and \( V_{\text{in, } \text{min}} \) is the minimum sampled voltage and \( V_{\text{in, } \text{diff}} \) is the sampled differential input voltage. The effect of unutilized time interval can be illustrated with some numbers. For a voltage ramp duration of 400ps on 25fF capacitors, and an input voltage swing of 0.6\( V_{\text{pp, diff}} \), an input common mode of 0.6V and a detection threshold \( V_{\text{th, det}} \) of 0.4V, the required discharge current can be found as 21.9\( \mu \)A. This creates a 57ps dead time interval and a maximum output \( T_{\text{out, max}} \) 343ps. For an 8-bit ADC with 1-bit folding, this value corresponds to an LSB step of 2.68ps.

The output time mismatch has three sources for this approach; VTC currents, sampling capacitors, and threshold levels.

On the other hand, the bidirectional architecture relies on zero crossing of the differ-
ential ramp. This allows full utilization of the input voltage range and eliminates the necessity of creating a $V_{\text{det,th}}$. Output time $T_{\text{out,bd}}$ can be written as:

$$T_{\text{out,bd}} = \frac{V_{\text{in,diff}}C_s}{2I_{\text{vtc}}}$$  \hspace{1cm} (3.3)

where the factor of 2 comes from the ramps of the opposite polarity. The required LSB step for this approach can be calculated using the same example of an 8-bit folding ADC with a ramp time of 400 ps. The same input swing $0.6V_{\text{pp,diff}}$ can be applied now with a common mode voltage of 0.4V. The required current can be found as $9.4\,\mu A$. This time all the ramp interval can be allocated to the $T_{\text{out,max}}$ since there is no dead zone, leading to LSB steps of 3.13ps. As a result, for a fixed voltage ramp time, the proposed approach offers a better SNR with lower power under the assumption of the same jitter conditions.

Hence, BVTC provides a wider input range with better utilization of the current, further enhancing power efficiency. Moreover, it eliminates the necessity of $V_{\text{th,det}}$ generation. Thus, there is one less source of mismatch. Furthermore, it inherently realizes the XOR function that is required to create an enable pulse for the ring oscillator. One challenge with this architecture is that the polarity decision by the sign bit comparator for folding must be taken in time to configure the BVTC.

The implementation of the BVTC is shown in Fig. 3-6. Two current mirrors (M6-M9) provide the charging and discharging currents to BVTC. These currents are either steered either into the sampling capacitors $C_{sp}$ and $C_{sn}$ depending on the input polarity or to the supplies. The sourcing and sinking currents are provided separately by a global current DAC with 6-bit resolution. M0-M5 are switches that are used to steer the currents. The outputs are taken from nodes $V_{\text{ramp},p}$, which indicates the positive ramp, and $V_{\text{ramp},n}$, which indicates the negative ramp. The operation involves three phases:

1. Pre-decision Phase
2. Decision Phase
3. Reset Phase

3.2.3.2.1 Pre-decision phase

This phase starts with the sampling of input voltages and ends when the sign comparator takes a decision. The operation in this phase is illustrated in Fig. 3-7. Since the polarity information is not yet known, the currents are steered to VDD by M5 and to GND by M4. This allows the drain nodes of M7 and M9 to remain at a constant voltage close to VDD and GND, respectively. M7 and M9 do not enter the cut-off region, which reduces the switching transients at the beginning of the ramp. Furthermore, the voltage levels at $V_{ramp}=p$ and $V_{ramp}=n$ reset the ZCD and prevent any memory effect.

3.2.3.2.2 Decision phase

The decision from the sign bit comparator is known during this phase, and BVTC is configured such that the currents are steered accordingly, as demonstrated in Fig. 3-8. If the sign bit $D_{sign}$ is 1, then the charge is extracted from $C_{sp}$ and injected to $C_{sn}$ by enabling M0 and M2 while M1 and M3 are in cut-off. This creates a voltage ramp.

Figure 3-6: The BVTC schematic.
Figure 3-7: The configuration of BVTC in the pre-decision phase.

(a) (b)
M0 M3
M1 M2 M4
M5
M6 M7
M8 M9

Figure 3-8: The configuration of BVTC in decision phase: (a) the sign bit equals 1 and (b) the sign bit equals 0.
with positive slope on $C_{sn}$ and a voltage ramp with negative slope on $C_{sp}$. $V_{rampn}$ follows $V_{sampp}$ with a voltage difference which equals to voltage drop on M0. Similarly, $V_{rampp}$ follows $V_{sampn}$.

If $D_{sign}$ is 0, then M1 and M3 turn on while M0 and M2 are in cut-off, which means $V_{sampp}$ has a positive slope and $V_{sampn}$ has a negative slope. This time $V_{rampp}$ follows $V_{sampp}$ and $V_{rampn}$ follows $V_{sampn}$.

$V_{ramp}$ always has a positive slope irrespective of the sign bit decision, while $V_{rampn}$ has a negative slope. This allows an analog XOR function to be realized at the output of the BVTC. Since ZCD inputs always have the same polarity, ZCD has a polarity dependent precharged design.

### 3.2.3.2.3 Reset phase

The reset phase starts after the zero crossing. The sampling capacitors are shorted by enabling M10 and M11 in Fig. 3-6 before acquiring a new sample. This reduces the distortion due to the memory effect.

### 3.2.3.3 Charge Injection CDAC

A CDAC is placed in parallel with the sampling capacitors $C_{sn}$ and $C_{sp}$ as can be observed from Fig. 3-2. The CDAC serves two purposes. First, it introduces input polarity dependent charge on the sampling capacitors prior to the ramp start. This charge injection creates voltage steps at the beginning of the ramp which ensures a minimum ramp time in the case of zero differential input. Second, it provides a redundancy window against wrong sign bit decisions due to comparator artefacts.

Fig. 3-9 shows the charge injection CDAC operation on a single sampling capacitor. During the sampling instant, the upper half of the capacitive bank is connected to VDD, and the lower half is connected to VSS. After the sign bit decision is taken, some of the capacitors in the lower half or the upper half are switched while the opposite operation is done on $V_{sampn}$. This creates a polarity dependent differential voltage jump. $D_{cp} < 0 : 3 >$ and $D_{cn} < 0 : 3 >$ are generated from the outputs of
the sign bit comparator and the control signals determining the number of capacitors
to be switched. The amount of charge injected is controlled by a 4-bit thermometer
code and each unit capacitor of the CDAC is 0.25fF. Custom fringe capacitors are
designed for the CDAC to reduce the total area.

3.2.3.3.1 Ramp Voltage Step

Zero differential input is a challenging case for a VTC since ZCD does not detect any
ramp crossings. It is equivalent to trying to map the zero-differential input to time
which is an absolute quantity. To resolve this ambiguity, an input polarity dependent
differential voltage step is introduced at the beginning of the ramps by the CDAC as
demonstrated in Fig. 3-10. This voltage step introduces an offset time \( T_{offs} \), which
is the minimum ramp duration. In the ZCD, this translates to increasing the enable
pulse width by \( T_{offs} \), and it leads to extra counts in the TDC denoted as \( N_{offs} \) on
top of the full scale count 128 as shown in Fig. 3-10.

3.2.3.3.2 Redundancy Window

The sole purpose of this voltage step is not to resolve the zero differential input. The
provided offset time \( T_{offs} \) can also be utilized to ensure a ramp crossing in the case
of wrong sign bit comparator decisions. This can be seen as a way of introducing
redundancy. This case is demonstrated in Fig. 3-11. As can be observed, a zero
crossing is guaranteed even if the sign bit is wrong, provided that the introduced
voltage jumps are sufficient to cover the input error range of the comparator. The
Figure 3-10: The voltage step introduced to resolve the zero differential input voltage, corresponding enable pulse, and the total count scale.

Figure 3-11: The voltage step introduced to resolve zero differential input voltage also provides a redundancy window. In the case of a wrong decision, the total ramp time is shorter than $T_{offs}$ and the output count $N_{out}$ is smaller than $N_{offs}$. 
Figure 3-12: StrongArm architecture with time-out modification which is used as a sign bit comparator.

The input error range in this concept refers to the maximum differential input voltage that can lead to a wrong comparator decision. In case of a wrong decision and sufficient redundancy, the time it takes to reach a zero-crossing is less than $T_{offs}$. ENB pulse shrinks by an amount $T_{red}$ which is proportional to the input voltage when the wrong decision is taken. This corresponds to a TDC output that is less than $N_{offs}$ by $N_{red}$. The output being less than $N_{offs}$ indicates an error in the sign bit comparator, and since $N_{red}$ is proportional to the input voltage, the output value can be recovered without any error. This allows wrong decisions to be tolerated and eliminates the need for calibration of the sign bit comparator. The lack of additional calibration circuitry reduces the area since the comparator calibration is generally done by capacitor banks or resistive ladders, which occupy a considerable area. Furthermore, the comparator can be sized smaller accepting high mismatch and input referred noise while reducing power dissipation and area.
3.2.3.4 Sign Bit Comparator

The redundancy window proposed in the previous section provides a solution against the wrong decisions of the comparator stemming from the offset and noise. However, unsettled polarity decisions can also cause a major problem for the ADC operation since, in this case, the BVTC can not be configured. The unsettled decisions are the results of the metastability of the comparator for low differential input voltages. To counteract this problem, a modified StrongArm latch architecture is used as a comparator, as shown in Fig. 3-12.

StrongArm offers a single stage architecture with high bandwidth and requires a single clock phase. Among the topologies evaluated, it provides the steepest input voltage vs. clock-to-q delay characteristic, which is of primary concern for this application since the decision has to be taken within a specific time interval. During normal operation, the time-out modification is not active, and M9, M10, and M11 are in cut-off state. The StrongArm operation can be divided into three operation phases. The first phase is the reset phase. During this phase, M7, M8, M12, and M13 charge the output nodes $V_{on}$ and $V_{op}$ and the drains of the input differential pair $V_x$ and $V_y$ up to VDD when CKP is low. In the second phase, the integration phase, CKP is high, and M1 and M2 draw different currents from the nodes $V_x$ and $V_y$ depending on the input voltages, which creates a differential voltage build-up. If the voltage on $V_x$ and $V_y$ drops by one NMOS threshold voltage, then the cross coupled devices M3 and M4 turn on, which starts the regeneration phase. With $V_{on}$ and $V_{op}$ reaching $V_{DD} - |V_{thp}|$, M5 and M6 also turn on, and the output node voltages regenerate to the rails with a regeneration time constant inversely proportional to the size of M3-M6, and proportional to the load capacitor.

The $CK_{to}$ enables the time-out modification branch after a fixed time which is $\frac{T_{samp}}{16}$ in this design. When the time-out branch is activated, the regeneration gain increases abruptly and regenerates whatever noise accumulated on $V_{on}$ and $V_{op}$, forcing a decision. This helps resolve metastable states. Since the redundancy window counteracts the wrong decision, a forced wrong decision does not affect the operation.
of the ADC.

3.2.3.5 Zero Crossing Detector

ZCD has an inverter based differential amplifier topology in the first stage and a precharged logic based second stage. The schematic is given in Fig. 3-13. As discussed in the BVTC section, the inputs to ZCD always have the same polarity as in Fig. 3-13. When CKN is low and CKP is high, the outputs ENBN and ENBP are precharged to VDD and GND, respectively, and the first stage is disabled. During this time, the input is being sampled, and the sign bit comparator has not taken a decision. $T_{\text{samp}}/8$ before the voltage ramps start, the sign bit is ready, and the BVTC is enabled. CKN becomes high, and CKP becomes low. Since $V_{\text{rampp}}$ always starts from a voltage close to GND and $V_{\text{rampn}}$ starts from a voltage close to VDD, M1 charges $V_x$ up to VDD and M4 sets $V_y$ to GND. Under these conditions, M5 and M10 are in cut-off region, and ENBN and ENBP stay intact. This defines the reset state before the voltage ramps start. As $V_{\text{rampp}}$ rises and $V_{\text{rampn}}$ falls, M2 starts discharging $V_x$ and M3 starts charging $V_y$. M5 and M10 slowly turn on, injecting and subtracting charges at the input of the inverters $I_4$ and $I_3$. After the voltages at $V_x$ and $V_y$ reaches the trip point of the cross-coupled inverters, the regeneration starts, and the ZCD outputs flip completely, ENBN decreases to GND while ENBP rises to VDD, indicating the zero crossing is detected, and the ROSC is disabled.
It is important to note the ZCD is functional without the cross coupled inverters \( I_1 \) and \( I_2 \). These devices are added to introduce positive feedback. After the zero crossing happens, the signal coupling or noise can lead to ramps falling back to the zero crossing point again. This can lead to glitches in ENBN and ENBP pulses, leading to a wrong ROSC state. Positive feedback prevents multiple zero crossings.

A similar discussion to StrongArm noise and offset performance can also be followed here. The second stage has a negligible effect when referred to the input since it turns on only after sufficient gain is developed in the first stage. The same argument also applies to \( I_1 \) and \( I_2 \). Furthermore, the noise contributions from the devices in the later stages are divided by the first stage gain when referred to the input, which reduces their contribution considerably. The input transistors M1-M4 dominate the offset and noise for this architecture.

### 3.2.4 Time-to-Digital Conversion

The block diagram of the TDC is given in Fig. 3-14 which is an architecture similar to the one in [113]. As mentioned previously, a gated ROSC is controlled by the enable signal (ENB) from the ZCD. Phase \( V_{rop} < 0 > \) from the gated ROSC is buffered by a regenerative buffer block (BUFLAT) and drives the clock inputs of the asynchronous integer counter which consists of two cascaded frequency dividers. The outputs of both gated ROSC \( (V_{rop} < 0 : 3 >, V_{ron} < 0 : 3 >) \) and the integer counter \( (D_{int,x} < 0 : 3 >, D_{int,y} < 0 : 3 >) \) are sampled by sense amplifier latches with asynchronous sampling logic. The outputs of the gated ROSC are interpolated during
sampling to obtain a shorter time step. Four of the eight outputs of the integer counter are chosen by the asynchronous sampling logic, which prevents sampling errors.

3.2.4.1 Gated Ring Oscillator (ROSC)

A ROSC is one of the most compact oscillators since it can be built with only inverters. Furthermore, it provides a high oscillation frequency which continues to increase with the emerging CMOS nodes. Due to the compactness, simplicity, and speed it offers, ROSC is utilized as a fine quantizer in this ADC.

A gated ROSC is designed to output eight cyclic thermometer coded fine bits with 2x phase interpolation. This requires four differential phases (four positive and four negative) and a ROSC with eight stages. Typically a ROSC with an even number of stages can not oscillate. Hence, feedback branches must be added to fulfill the Barkhausen Criteria, which suggests the phase shift around the loop must be $2m\pi$ where $m$ indicates possible oscillation modes and the loop gain must be more than or equal to 1. Considering this, only three feedback topologies give rise to oscillations with a non-zero oscillation frequency. The detailed analysis of these topologies can be found in [104]. Since the speed is the main specification for this application, these topologies are evaluated by simulations and compared with respect to the oscillation frequency. The one illustrated in Fig. 3-15 is used, which offers the highest oscillation frequency. The extracted simulation results show an oscillation frequency of 25.4GHz for the slowest corner. The topology has a direct path (D) and a feedback path (F), which connects two nodes that are two stages apart. The gated inverters on the direct path have double the size of the feedback path. Extra switches are added for reset functionality which resets $V_{rop} < x >$ nodes to VDD and $V_{ron} < x >$ nodes to GND, except the nodes triggering the oscillation. The gated ROSC is enabled and disabled by header and footer devices which require differential inputs that are provided by ZCD. The ROSC is always reset before starting a new oscillation.

The clocks interpolating and sampling the ROSC outputs are generated with a fixed delay after the falling edge of the ENB signal. The outputs are sampled before waiting
for the end of the conversion period since leakage can cause ROSC state errors after all the stages are disabled.

### 3.2.4.2 Phase interpolation

A ramp time of $5T_{samp}/8$ requires at least 3.9ps and 4.9ps time steps for achieving 1GS/s and 1.25GS/s sampling rates with 8-bit resolution and 1-bit folding. In reality, this allocated ramp time also includes the combinational delays of the operation. Also, the PVT variations can further shrink this step. This means that a time step smaller than the single inverter delay in 5nm CMOS is required. As discussed in section 2.3.3.2.3, phase interpolation is required to refine the time step further. However, the discussed techniques, such as inverter based interpolation or resistive interpolation, can not be applied here due to two reasons. First, the gated ROSC operates in a cyclic manner which means the interpolation is carried out at each cycle, giving rise to dynamic power dissipation continuously. Second, the gated ROSC has a reset state where the positive $V_{rop} < x >$ phases are reset to VDD, and the negative $V_{ron} < x >$
phases are reset to GND. This causes crossover currents if the interpolation is done by inverters since two similar inverters in the interpolation network drive the output node to different directions. Moreover, conventional approaches increase the area and power dissipation. Hence, a new interpolation strategy is required, ideally without increasing the power dissipation and area.

The principle of phase interpolation applied in this ADC is demonstrated in Fig. 3-16. The differential phases from the gated ROSC $V_{rop < 0 >}$, $V_{ron < 0 >}$ have a zero crossing at $T_0$. Similarly, the consecutive differential phases $V_{rop < 1 >}$ and $V_{ron < 1 >}$ have a zero crossing at $T_1$, which is one time step away from $T_0$. As shown in Fig. 3-16 phases $V_{rop < 1 >} - V_{ron < 0 >}$ or $V_{rop < 0 >} - V_{ron < 1 >}$ yield an additional zero crossing $T_i$ which resides in the middle of $T_1$ and $T_0$. This can be exploited in a similar manner to flash interpolating ADCs by utilizing additional latches to detect this crossing.

The sense amplifier latch configuration for interpolation is illustrated in Fig. 3-17 $V_{rop < 1 >} - V_{ron < 0 >}$ and $V_{rop < 0 >} - V_{ron < 1 >}$ are connected to double differential SAL inputs. As can be observed in Fig. 3-16 the intersections can result in two common mode voltage levels $V_{cm,h}$ and $V_{cm,l}$ depending on the stop time of the gated ROSC. As NMOS input pairs are used, a lower input common mode can drive the input differential pair into the subthreshold region, which leads to considerably less integration gain. This reduces the sensitivity of the SAL and can lead to LSB

![Figure 3-16: Phase interpolation waveforms.](image-url)
Figure 3-17: SAL schematic and the input configuration for interpolating and non-interpolating latches.

errors. Double differential input pairs are used such that whenever one pair has a low input common mode, the other one has a higher common mode and defines the output accuracy. For the non-interpolating outputs, the same SAL architecture is used but with the differential phases from the ring connected to both differential inputs, as shown in Fig. 3-17.

This interpolation approach allows finer time steps with negligible area and power overhead since the phases from the gated ROSC are interpolated only when the sampling clock arrives. Because the same SAL architecture is also used for non-interpolating inputs, it does not require additional design effort. The measured DNL performance indicates that sufficient linearity can be obtained for 8-bit resolution.

3.2.4.3 Integer Counter

The integer counter has to be clocked with gated ROSC frequency, which limits the choices of counter topologies. A long propagation delay does not permit a high clocking speed for the counters; thus, a counter architecture that minimizes the combinational delay between the latches is the most promising. This is provided by frequency
Two cascaded divide-by-four frequency dividers are used as an integer counter [102]. Each divider has four cascaded latches, as illustrated in Fig. 3-18. There are a total of eight outputs taken from the integer counter, yet four are needed for the coarse bits. Extra outputs are utilized in an asynchronous sampling scheme to capture only the non-transitioning outputs when the sampling clock arrives. The latches have complementary pass transistor logic (CPL) architecture.

### 3.2.4.4 Asynchronous Sampling

Asynchronous sampling is necessary for the TDC since the gated ROSC and the integer counter operate asynchronously with the sampling clock. This means that the output of both the ROSC and the integer counter can be sampled wrong if they are in transition when the sampling clock arrives. The technique presented in [102] is summarized in Fig. 3-19. The operation can be explained as follows: since the first divide-by-four block is clocked by the BUFLAT, only two phases in the first stage, $D_{int,y} < 0 >$ and $D_{int,y} < 1 >$, are stable when $V_{rop_buf} < 0 >$ is high while the other
two phases $D_{int,x} < 0 >$ and $D_{int,x} < 1 >$, are in transition. Thus, the multiplexers at the outputs pass $D_{int,y}$ as the valid phases if the sampled value of $V_{rop_{buf}} < 0 >$ is 1 and $D_{int,x}$ if the sampled value of $V_{rop_{buf}} < 0 >$ is 0. Similarly, the sampled value of $D_{int,y} < 1 >$ acts as a select signal for the second divide-by-four block. Then the outputs $s < 0 : 3 >$ can be converted to binary by an encoder. This approach allows asynchronous sampling of the fine and coarse bits by evading the transitions in the signals.

3.2.4.5 Regenerative Buffer (BUFLAT)

The last output of the gated ROSC has to drive the capacitive load of the first frequency divider in the integer counter. This requires buffering since the ROSC size is optimized for speed and power dissipation. However, only the buffering functionality is not sufficient, and a regenerative characteristic is also desired, as the differential
voltage level at the input of the buffer can be at any value between the supplies since the ROSC can be at any voltage level when it is stopped. Furthermore, the output of the BUFLAT has to be stable at all times since a glitch at the output can clock the integer counter mistakenly. This can cause errors with the size of 16LSBs, which is detrimental to linearity. A multistage BUFLAT is appropriate here to reduce the loading on the last ROSC phase to lower the DNL error.

The designed BUFLAT has a clocked architecture, as shown in Fig. 3-20. The small inverters at the input reduce the loading on the gated ROSC. The transmission gate switches are on when the gated ROSC is active, and they are off when the gated ROSC is stopped. This cuts the connection of the integer counter with the gated ROSC and prevents any state change of the counter after the gated ROSC is stopped. The cross coupled inverter pair I, which is always active, regenerates the inputs to the rails. It also introduces hysteresis to reduce the possibility of metastability events. The second set of inverters is added for further buffering, and the cross coupled pair II ensures a 180 degrees phase shift between $V_{rop_{-buf}} < 0 >$ and $V_{ron_{-buf}} < 0 >$.

3.3 Cascaded Block Performance

The ADC consists of cascaded stages starting with the T&H block. At the output of each stage, the dynamic performance can be estimated through simulations in order to extract the performance degradation. The nominal corner annotated schematic transient noise simulation results for both operation cases are shown in Fig.3-21 which indicates the SNDR, SFDR, THD and ENOB values at the output of the sampling switch, BVTC, ZCD, and TDC (OUT refers to buffered TDC output). For 0.7V supply 1GS/s $f_s$ case, the operation starts with a sampled unquantized ENOB of 8.1. After the VTC block, the SNDR decreases by 6.6dB. Although the nonlinearity of the current sources in the VTC has an effect on the SNDR, a bigger portion of the degradation stems from the noise. The SNR value can be calculated at the output of this stage as 45.9dB compared to THD of -48.5dB. The integration noise is the dominant noise source at the output of the VTC stage. ZCD further contributes to
Figure 3-21: The SFDR, SNDR, ENOB and THD figures at the output of each stage for 0.7V supply at 1GS/s and 0.8V supply at 1.25GS/s.

The noise and decreases the SNDR down to 41.4dB while the THD at the output remains at -46.9dB level. The majority of the noise contributed at this stage is due to input differential pairs of the ZCD. The output SNDR is 39.9dB, which indicates 6.3 ENOBs. The reduction in the last stage is minor and mainly caused by the quantization noise. The TDC contributes negligibly to the overall noise performance. A similar behaviour can be observed for 0.8V 1.25GS/s case. For this case, the THD values are lower due to increased supply voltage improving linearity and SFDR. This gives a slight boost to the ENOB values, yet the noise from the VTC and ZCD still degrades the ENOB at the output. The results indicate that the ADC performance is mainly limited by the voltage-to-time conversion stages. Corner simulations show less than 0.5 ENOB variation at the output.

### 3.4 Measurement Results

The ADC implemented in 5nm CMOS is characterized using needle probes. A micrograph is shown in Fig. 3-22. The approximate placement of the blocks on the
Figure 3-22: Chip micrograph and the layout details of the single ADC.
top level is also shown. The chip has eight ADCs, and eight shift registers which can sample 256 samples each to capture the ADC outputs. The clock receiver (CLK) receives the clocks at $8f_s$ where $f_s$ is the single ADC sampling frequency. It includes a CML-to-CMOS converter and clock buffers while the sampling clocks for each ADC are generated locally. A bidirectional interface (BIDI interface) organizes the capture of the ADC results and shifts them out in a sequential manner. It is also used for the configuration bits for controlling the global current DAC that sets the VTC currents and the ones that set the differential voltage jumps in the BVTC.

The layout details of the single ADC are also shown in Fig. 3-22. The total area is $313\mu m^2$. The width is dominated by the sampling capacitors. The floor plan is done in a way to reduce the critical signal routings. The ROSC and the integer counter output routing are minimized by placing them next to the sampling block. The gated ROSC outputs are further balanced to reduce the DNL errors while interpolating. The only critical signal with long routing is the enable signal (ENB) for the gated ROSC, which is laid out carefully.

The measurement setup is given in Fig. 3-23. Both clock and input signals are
obtained from Agilent E8257D signal generators which are synchronized. Differential input signals are obtained by a balun, and bias-Ts are used for setting the input common mode voltage. In a similar manner, differential clocks are obtained. Supply voltages are set externally using DC sources. The test chip output is read out serially by an FPGA board, the outputs of which are captured and processed by a computer. The same board is also used for configuring the circuits on the chip.

3.4.1 Static Tests

DC measurements are carried out using a DC source at the input to create a piecewise linear ramp by sweeping the differential input in small voltage steps. This allows for finding the folding point for the ADC, characterizing the sign bit comparator, and characterizing the noise and the static performance.

3.4.1.1 Folding Point

As the ADC utilizes the folding technique, the exact folding point must be found. In the case of ideal operation, the folding location exactly corresponds to zero differential input. However, the mismatch and noise shift this folding point, and the location must be determined to map the output values to the correct scale. This requirement for folding ADCs either necessitates an auxiliary tuning circuitry to set it at a desired position or an algorithm run by the DSP to find the fold point. The process is demonstrated in Fig. 3-24. Input is swept between -400mV and 400mV, with 1.25mV voltage steps. At each point, ADC is clocked 256 times, and the outputs are shown in Fig. 3-24. The output of the sign bit comparator is also shown. It shows the distribution of the sign bit decisions depending on the noise and offset.

The output of the sign bit comparator can be used to find the location of the fold point. The input voltage, which gives an equal amount of 1s and 0s at the sign bit comparator output, indicates the location of the fold. The correct output at the folding point must be found such that the combined output shows a linear characteristic. For this case, the output value at the fold point is found by a linear fit. Then this value can
Figure 3-24: Output code of the ADC and the sign bit comparator for a DC ramp input between -400mV and 400mV.

Figure 3-25: Folded output code of the ADC and the sign bit comparator for a DC ramp input between -400mV and 400mV.
be used to map the raw data to the folded output waveform according to the equation
\[ D_{out} = (D_{raw} - D_{fold})D_{zx} \]
where \( D_{raw} \) is raw output data shown in Fig. 3-24. \( D_{fold} \) is the folding point output and \( D_{zx} \) is the signed output of the sign bit comparator (\( D_{zx} = 1 \) if \( D_{sign} = 1 \) and \( D_{zx} = -1 \) if \( D_{sign} = 0 \)). The resulting folded waveform is shown in Fig. 3-25.

The folding procedure must be carried out before the operation. Once the fold point is fixed, then the output can be mapped correctly. Although it is found by a foreground procedure here, in a complete receiver, the DSP can run a routine to extract the fold point from the input signal statistics in the background.

### 3.4.1.2 ADC Static ENOB

The same input ramp can be used to estimate the static ENOB of the ADC. The corresponding figure is given in Fig. 3-26. The estimated ENOB values range between 6.3 to 6.8. Since there is no degradation due to dynamic effects, the resulting waveform shows the effect of noise. At differential input levels close to zero, the noise has a weaker effect. ENOB peaks at zero differential input since the integration noise of the BVTC becomes minimum at this point due to the shortest ramp duration and
integration window. The ADC noise is dominated by the BVTC and ZCD.

### 3.4.1.3 Sign Bit Comparator Characterization

Fig. 3-27 shows the comparator input referred noise and offset. The results are obtained by assuming that the noise is Gaussian. The outputs of the sign bit comparator are fitted into an erf function to obtain the standard deviations. The results show -19.57mV offset and 1.5$V_{rms}$ noise.

### 3.4.1.4 INL and DNL

The INL and DNL are found by sinusoidal histogram tests as explained in section 2.2.1.2. The results obtained for 0.7V supply 1GS/s $f_s$ and for 0.8V supply 1.25GS/s $f_s$ are summarized in Fig. 3-28 and in Fig. 3-29.

### 3.4.2 Dynamic Tests

Sinusoidal input tests are carried out to capture the dynamic performance of the ADC. The SFDR, SNDR, HD2, HD3, and THD values are obtained across the second
Figure 3-28: INL and DNL curves of the proposed ADC for 0.7V supply and 1GS/s $f_s$ case.
VDD=0.8V, \( f_s = 1.25\text{GS/s} \)

Figure 3-29: INL and DNL curves of the proposed ADC for 0.8V supply and 1.25GS/s \( f_s \) case.
Nyquist zone.

### 3.4.2.1 Input Spectra at Nyquist

ADC output spectra at Nyquist input frequency for both cases; 0.7V supply, 1GS/s $f_s$ and 0.8V supply 1.25GS/s $f_s$ are given in Fig. 3-30 and Fig. 3-31. The measurements are taken with full scale input of 0.75$V_{pp,diff}$. The ENOBs are 6.16 for the first case and 6.23 for the second case. THD is dominated by the second harmonic in the first case and the third harmonic in the second case. The second case shows a lower THD of -45.42dB compared to the first case -44.08dB, due to the relaxed headroom offered by the higher supply voltage.

### 3.4.2.2 Input Frequency Sweeps

The input frequency is swept in the second Nyquist zone, and the SFDR and SNDR values are calculated for a differential input of 0.75$V_{pp,diff}$. The results are demonstrated in Fig. 3-32 and Fig. 3-33 for 0.7V supply 1GS/s sampling rate and 0.8V supply 1.25GS/s sampling rate respectively. The SNDR has a flat profile with a variation of less than 1.1dB across the whole Nyquist zone for both cases. SFDR values stay above 47dBc for the first case and 49dBc for the second case.

In the second Nyquist zone, distortion values are also calculated. The THD, HD2, and HD3 curves are given in Fig. 3-34 and Fig. 3-35. The frequency sweeps show that THD stays below -44dB at 1GS/s and below -45.4dB at 1.25GS/s. THD is mainly dominated by the third harmonic.

### 3.4.3 Power Dissipation

The ADC dissipates 1.18mW at 1GS/s and 1.9mW at 1.25GS/s. The numbers correspond to 16.6fJ/conv-step and 20.3fJ/conv-step Walden FoM. The gated ROSC and the remaining parts are powered by different supplies with the same voltage values. The power breakdown is shown in Fig. 3-36. 33% of the total power is dissipated by the gated ROSC, and the remaining 67% by the rest of the blocks.
Figure 3-30: Output spectra at Nyquist frequency for 0.7V supply and 1GS/s $f_s$. The first ten harmonics are shown in red, and other spectral contents and noise are in blue.

Figure 3-31: Output spectra at Nyquist frequency for 0.8V supply and 1.25GS/s $f_s$. The first ten harmonics are shown in red, and other spectral contents and noise are in blue.
Figure 3-32: SNDR and SFDR values across the second Nyquist zone for 0.7V supply and 1GS/s $f_s$.

Figure 3-33: SNDR and SFDR values across the second Nyquist zone for 0.8V supply and 1.25GS/s $f_s$. 
Figure 3-34: THD, HD2 and HD3 values across the second Nyquist zone for 0.7V supply and 1GS/s $f_s$.

Figure 3-35: THD, HD2 and HD3 values across the second Nyquist zone for 0.8V supply and 1.25GS/s $f_s$.

Figure 3-36: The power dissipated at 0.7V supply and 1GS/s $f_s$ and 0.8V supply and 1.25GS/s $f_s$. 
3.5 Conclusions and Comparison with Prior Art

A comparison of the ADC performance with the state-of-the-art voltage and time domain ADCs running at GS/s sampling speeds can be found in Fig. 3-37.

In summary, the proposed ADC achieves competitive Walden FoM with the smallest area and a high input swing. In addition, it provides above 1GS/s sampling speed which can scale up with the new CMOS technology nodes thanks to the digital intensive design. Thus, it is a promising sub-ADC candidate for the next generation wireline communication systems targeting above 100GS/s as it is highly compact, power efficient, runs at GS/s, and has a digital intensive design.
<table>
<thead>
<tr>
<th>Resolution (bits)</th>
<th>FOM @ Nyquist</th>
<th>SNDR @ Nyquist</th>
<th>SFDR @ Nyquist</th>
<th>DNL (min,max)</th>
<th>INL (min,max)</th>
<th>Supply (V)</th>
<th>fs (GS/s)</th>
<th>Fs (GS/s)</th>
<th>Active Area (μm²)</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>7000</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>49.6</td>
<td>1.8</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>5</td>
<td>6.0</td>
<td>1425</td>
<td>7.4</td>
</tr>
</tbody>
</table>
Chapter 4

64x Interleaved SAR ADC

A 56GS/s 8-bit asynchronous SAR ADC fabricated in 4nm CMOS technology is demonstrated. The 16x4 interleaved ADC uses a novel bootstrapping technique and a class-AB follower in the 1st rank interleaver. It achieves a broad input common-mode ($V_{in,cm}$) range; from 0.3V to 0.6V, the total harmonic distortion stays below -52dB at 4.1GHz with -0.2dBFS amplitude at 0.8$V_{pp,diff}$ maximum full scale. The ADC includes analog foreground calibration means for offset, gain, skew, and bandwidth. The measured ENOB is 6.5 at low frequency and stays above 5.2 up to Nyquist frequency, which is limited by jitter. The bandwidth is higher than 27GHz. The ADC uses a single 0.8V supply voltage and achieves a Walden FoM of 47fJ/conv-step.

4.1 Motivation

Distributed computing has driven the aggressive data rate trend of wireline I/O links in the last couple of decades. The bandwidth is doubled every four years for a variety of communication standards. Emerging data intensive applications such as machine learning and AI will likely contribute to data traffic leading to the adoption of new modulation techniques that offer high spectral efficiency and impose more stringent latency requirements. Furthermore, Internet of Things devices lead to a
substantial increase in data movement in data centers causing a tight bandwidth demand. In order to fulfill the high throughput demand of the next generation data-intensive applications, the bandwidth efficiency has to be increased to overcome the channel limitations. This leads to the emergence of multi-level signaling such as PAM4, which is widely used targeting data rates for 112Gb/s and above [101, 105, 106, 107, 108, 109]. This trend will pave the way for more complex modulation formats that also utilize frequency domain such as OFDM. The degraded SNR due to complex modulation schemes, in addition to the flexibility required to run the SerDes at different conditions, favor the ADC-DSP based RX architectures, which allow complex digital equalization strategies and provide the ease of portability between different technology nodes.

Asynchronous SAR based ADC is a proven architecture that scales well with the technology and maintains a reasonable speed with high power efficiency [54]. An asynchronous SAR ADC design is adopted for this work which is the choice for the majority of the wireline receivers targeting 112Gb/s and beyond data rates [101, 105, 106, 107, 108, 109]. The per channel sampling speeds are around 1GS/s for 6-8 bits of nominal resolution. Assuming the internal timing does not pose a bottleneck for the speed of the asynchronous SAR architecture, the single channel sampling rate is limited by the Bit Error Rate (BER) degradation due to metastability [108] which necessitates a high interleaving factor.

With increased interleaving factors, the design of the interleaver becomes critical. Using an NMOS switch as the 1st rank sampler followed by a PMOS source follower is a simple and efficient design approach [70], but this combination can limit the ADC performance due to several factors. NMOS or a Transmission Gate sampling switch resolution heavily depends on input signal characteristics. A bootstrapped switch can resolve this issue, yet the tracking bandwidth of the conventional bootstrapped switches can limit the input bandwidth. The PMOS buffer noise and linearity performance depend on the input common mode voltage \( V_{cm} \), and the current driving capability is low. Furthermore, the slew rate is signal dependent. To address these
limitations, the reported 8-bit 56GS/s ADC uses a first rank T&H circuit comprising a new bootstrapped sampling NMOS switch architecture. A switch-capacitor biased Class-AB buffer is introduced and used as a first rank buffer which offers high current driving capability and shows a good linearity performance.

Inverter based analog front ends are scaling friendly alternatives to the current-mode-logic style front ends. They can offer area and power advantages due to the elimination of passive inductors, T-coils, and static currents. Furthermore, they can provide the necessary gain boost at lower supply voltages \[10\]. The output common mode voltage of the inverter based front ends are close to the mid-rail voltage defined by the device sizes. Different process corners can lead to the variation of the output common mode. In this work, the ADC is designed for an inverter based front end targeting 0.4V mid-rail \(V_{cm}\) with a high tolerance to common mode variation.

Interleaving spurs create significant harmonic distortion limiting the resolution of the high-speed ADCs designed for wireline applications. To target offset, gain, skew, and bandwidth mismatches, the calibration knobs are added that allow a full foreground calibration. This is achieved by offset and gain resistive ladders in the sub-ADC level, skew and bandwidth programmable capacitors in the interleaver level, and a resistive ladder for duty cycle correction in the clock receiver.

### 4.2 Implementation

#### 4.2.1 Overall architecture and clocking

The ADC utilizes a two-rank interleaver architecture as shown in Fig \[4-1\]. A 16x4 interleaving factor is chosen for optimal bandwidth \[79\]. The first rank consists of 16 T&H samplers and buffers. Each buffer drives the S&Hs of 4 sub-ADCs. In total, 64 sub-ADCs are used to reach 56GS/s.

Input is terminated with differential 50Ω termination resistance and a common mode capacitor to prevent static current to the ground. T-coils are used for input matching
and capacitive load cancellation that help increase the input bandwidth. ESD diodes are used for protection.

The sub-ADC outputs are captured by 256 sample shift registers which allows a total of 16384 samples to be captured. The sampling clocks of each shift register, which is obtained from the corresponding sub-ADC, are staggered to prevent synchronous switching activity at the C64 rate \(56\text{GHz}/64=875\text{MHz}\). This eliminates the switching glitches in the supply voltage.

The clocking approach is shown in Fig. 4-2. The ADC receives differential in-phase and quadrature-phase CML clocks from external RF signal generators (CK[0:4]). The clock receiver consists of CML to CMOS converter with 8-bit duty cycle correction (DCC) control and differential phase buffers for driving the heavy capacitive load. The duty cycle can be corrected within the 46-54% range within the CML to CMOS converter. The CMOS level quadrature clocks obtained at the output of the clock receiver are inputs to the PLS MAIN block. The PLS MAIN divides C4 clocks to obtain C8 and C16 clocks and creates 4 Unit Interval (4UI) wide enable pulses (ENB16) from them. ENB16 is used by PLSGEN blocks to mask the 2UI wide

Figure 4-1: Interleaved ADC top level block diagram.
sampling and reset pulses locally, as shown in Fig. 4-3. The PLSGEN blocks are chain connected, and the ENB16 pulse is retimed and forwarded on this chain such that each PLSGEN creates the reset and sampling pulses within proximity of the T&H and the buffer, reducing the routing effort and any possible signal coupling. The most critical signals are CK4 clocks that have to be routed to each PLSGEN block (connections not shown in Fig. 4-2 for clarity) within this scheme, so their routing must be optimized. The first rank interleaver is grouped as even and odd such that only CK4\[1,3\] is routed to the odd section (shown on the left of Fig. 4-2) and CK4\[2,4\] is routed to the even section (shown on the right of Fig. 4-2) to minimize the high speed CK4 routing.

After the first rank sampling and reset pulses are obtained, the second rank ADC
Figure 4-4: The first rank \((CK_16_{\text{sam}})\) and the second rank \((CK_{64_{\text{adc}}})\) sampling clocks and the quarter rate clock CK4.

The first rank sampling pulse \((CK_{64_{\text{adc}}})\) can be created using first rank sampling pulses, which are 8UI apart. The obtained waveforms for the first and second rank sampling pulses are given in Fig. 4-4. The first rank clock period, which is 16UI long, is composed of 2UI tracking and 2UI reset times that leaves 12UI hold time which can be used for the 2nd rank sampling. The second rank S&H utilizes 8UI of this interval which leaves 2UI margins before and after the sampling pulse. This provides a timing buffer against PVT variations and eliminates the need for an aligner [106]. Furthermore, it relaxes the ENB16 enable pulse rise-fall time requirement since the jitter is defined by CK4 clocks used for retiming the sampling pulse. This significantly reduces the clocking power, and the majority of the power is dissipated by buffering the CK4 clocks, which define the jitter performance.

### 4.2.2 First Rank T&H

The first rank T&H is a crucial block for achieving the target data rates since it has to support the full input signal bandwidth. Rising data rates shrink the sampling pulses requiring higher bandwidths from the first rank T&Hs. Increasing the first rank interleaving factor allows a wider tracking pulse. However, this reduces the input pole frequency formed by the input routing capacitance and the termination resistor due to increased routing capacitance. Furthermore, it also increases the number of
critical clocks to be routed, inflating the power dissipation and the effort for skew correction.

Several T&H architectures find wide use in interleaved SAR ADCs: single NMOS or PMOS, transmission gate (TG), and the bootstrapped switch \cite{70,108,101}. Additional modifications like cross-coupling for feed-through cancellation and charge injection cancellation can be added to existing topologies to increase the resolution. NMOS or PMOS T&Hs do not offer mid-rail input common mode ($V_{in,cm}$) operation. Their performance is sensitive to the $V_{in,cm}$ variation, and the input dependent on resistance ($R_{on}$) dominates the harmonic distortion. Thus, they are not promising candidates for this application targeting 8-bit resolution. Simple transmission gate (TG) T&Hs are generally used in 6-8 bit ADC interleaver architectures due to their compact layout and simplicity \cite{108}. They can also accommodate mid-rail $V_{in,cm}$, which makes them suitable candidates for the current application. A possible implementation with cross coupled devices for improved isolation during off condition is shown in Fig. 4-5. One problem with the TG switches is the clock routing overhead.

Figure 4-5: CMOS transmission gate sampling switch with feed-through cancellation.
which requires differential high speed clocks to be routed to each first rank interleaver T&H reducing the routing resources and increasing the power dissipation. The skew between the P and N clocks is also of concern to reach the desired performance. Furthermore, the cross-coupling for signal feed-through cancellation requires the same driving conditions at the gates of feedthrough transistors $TN_3 - TP_3, TN_4 - TP_4$ and the main switches $TN_1 - TP_1, TN_2 - TP_2$ which can reduce the area advantage. However, the fundamental limitation comes from the sensitivity of the resolution to the clocking conditions, which will be explained later in this section.

This work utilizes a constant $V_{gd}$ bootstrapping technique where the voltage between the sampling capacitor and the gate of the switch $TN_{1P}$ is fixed as shown in Fig. 4-6. Contrary to conventional bootstrapped switch topologies, the constant voltage here is generated using a source follower. When the $CKP_{samp}$ becomes high $TN_6$ turns off and $TP_2$ turns on. Then $TP_2$ biases $TP_1$ with the current provided by $C_{boost}$ capacitor. Assuming that $C_{boost}$ is big enough, the current is constant, and the source-gate voltage of $TP_1$ provides the necessary voltage shift from the sampling capacitor to the gate of the switch fixing the $V_{gd}$. Since the source of $TP_1$ can rise above VDD depending on the input swing, an extra protection device $TN_5$ is used to limit the drain-source voltage swing of the $TP_1$. This allows all the device terminal voltages to remain below $V_{max}$ specifications. When $CKP_{samp}$ goes down, $TN_4$ pulls down the gate node of $TN_{1P}$ with a constant slew rate, and since the $V_{gd}$ of $TN_{1P}$ is also constant, the turn off time of the $TN_{1P}$ is signal independent.

Apart from the benefits of the bootstrapping, such as less input dependent sampling time and weaker input dependent charge injection, this topology provides several advantages. First, the gate capacitance of $TP_1$ is embedded in the sampling capacitor. Considering that the first rank of the interleaver has 16 T&H switches, this eliminates any extra capacitive loading at the input, increasing the input pole frequency and the bandwidth. To settle within short tracking pulses, the T&H must have a short tracking delay between the tracked node and the gate of the switch. This is accomplished here by having a single device, $TP_1$, on the tracking path, which offers reduced delay...
Figure 4-6: The proposed bootstrapped switch.

compared to the conventional bootstrapping architectures [111]. Since the purpose of the $C_{boost}$ here is not to define the fixed voltage but to provide the bias current for $TP_1$, it can be sized independently of the switch $TN_{1P}$. A shorter tracking time provides a more power efficient operation by reducing the charge subtracted from $C_{boost}$ and allows a smaller capacitor to be used. Furthermore, this topology allows a single clock phase to be used, which reduces routing efforts for first rank critical sampling clocks.

In terms of loss, both the transmission gate and the proposed bootstrapped architecture perform similarly. The annotated schematic simulation results showing the loss at 28GHz input frequency against clock rise and fall times can be found in Fig. 4-7. The sampling clock rise and fall times are swept in order to obtain the sensitivity of the loss on different clocking conditions. The results show that the difference between the loss at different clocking conditions remains below 1.3dB. This can be explained by the lower on resistance offered by this technology, even without bootstrapping.

However, the TG T&H resolution starts to degrade when the clock buffer strength
is decreased. The comparison results are shown in Fig. 4-8 for annotated schematic simulations where the output of the T&Hs are quantized by ideal 8-bit quantizers for both architectures. Even if both architectures perform similarly in terms of linearity at low clock rise/fall times, the degradation of the TG becomes evident when the clock rise/fall time increases. This indicates the resolution degradation due to input signal dependent aperture effects such as input dependent sampling time and charge injection. Since the bootstrapped T&H architecture is robust against these effects, quantized ENOB remains within the range 7.3-8 at clock rise/fall times from 2\(\text{ps}\) to 14\(\text{ps}\). On the other hand, CMOS T&H is vulnerable to increasing clock rise/fall times, and the ENOB falls down to 5.3 at 14\(\text{ps}\). This increase in harmonic distortion can be detrimental for ADCs targeting higher order modulation formats.

In summary, the proposed first rank T&H increases the input bandwidth and offers better linearity and robustness against different clocking conditions at high speeds.
It also relaxes the clock buffering requirements contributing to power efficiency.

4.2.3 First Rank Buffer

The objective of the first rank buffer of an interleaved SAR ADC is to provide enough strength to drive multiple second rank sub-ADCs and to shield the input from the load of the following stages for a higher input bandwidth. Sub-ADCs introduce a heavy capacitive load that must be driven at the target 2nd rank sampling rates. Furthermore, due to high interleaving factors, the length of wires used for routing S&Hs inputs to the sub-ADCs increases, contributing to the load capacitance further and introducing a high resistance on the signal path. This requires buffers with a high driving capability and efficiency. The target buffer architecture must have a solid linearity performance with low harmonic distortion up to the Nyquist frequency, not to impact the resolution of the ADC. Moreover, the desired bandwidth has to be achieved in a power efficient manner.

The two commonly used buffer topologies are source followers and push-pull followers as shown in Fig. 4-9. The problem with the single transistor source followers (Fig. 4-9(a)) is the sensitivity of the buffer performance to the input common mode.
variations. Noise and linearity performance is sensitive to $V_{in,cm}$, which can affect the performance in case of a common mode shift from the AFE. Furthermore, the current sourcing capability is limited by the bias current. For large input swings, the slew rate is signal dependent since for low input voltages, the discharging current is defined by $T_{P1}$ while for high input voltages, the charging current is defined by the bias transistor $T_{P2}$. This introduces an input dependent settling behaviour which, especially at high sampling rates where the output does not settle fully, causes harmonic distortion.

The push-pull follower can reduce the signal dependent slew rate by utilizing both NMOS and PMOS, as shown in Fig. 4-9(b). This also increases the efficiency for the required current driving capability by turning on each transistor during a fraction of the operation by adjusting the bias voltages. However, the biasing requires resistors that occupy a significant area compared to active devices in the recent CMOS nodes.

At DC input, the biasing network introduces a resistive division from the input to the gates of the push-pull source follower. Assuming $R_{N1}=R_{P1}$ and $R_{N2}=R_{P2}$, the transfer function can be written as:

$$\frac{V_O(s)}{V_{IP}(s)} = \frac{R_{N1}(1 + sC_{N1}R_{N2})}{(R_{N1} + R_{N2})(1 + \frac{sR_{N1}R_{N2}C_{N1}}{R_{N1}+R_{N2}})} \times H_{pps}(s) \quad (4.1)$$

where $H_{pps}(s)$ is the inherent large signal frequency response of the push-pull source follower. It can be observed that at high frequencies, the biasing network passes the signal with a gain of 1; however, at DC input, the resistive division defines the transfer function. At DC input, the gate voltages of $T_{N1}$ and $T_{P1}$ also have voltage components stemming from $V_{BH}$ and $V_{BL}$. Thus, the DC input is not buffered correctly. The introduced pole and zero pair from the biasing network also change the low frequency gain behaviour. This difference between the high and low frequency behaviours and the failure at DC operation may affect the BER. The effective frequency seen at the receiver input depends on the sparsity of the input data stream. To illustrate, an NRZ waveform toggling between 1 and 0 has a higher frequency compared with the same bit repeated over multiple periods. Hence, different buffer gains are applied for different scenarios, which increases BER. Baseline wander, which is caused by the
effective low frequency at the input of the RX, can lead to eye closures, especially in higher order PAM signals where the signal margins are lower. To address this, a new biasing approach is introduced.

The proposed buffer utilizes a switched capacitor biasing network added to the push-pull source follower, as shown in Fig. 4-10. During the reset mode, the differential inputs are shorted with $T_{P4}$ and $T_{N4}$. The input CM is sampled on $C_{B1}$ and $C_{B2}$ while $C_A$, which was charged to $V_{BH} - V_{BL}$, is parallel connected to the $C_{B1}$ and $C_{B2}$. In the buffer mode, when $CKP_{RST}$ is low, the differential sampled input is connected to the buffer input. $C_{B1}$ and $C_{B2}$ provide the necessary voltage shifts to bias $T_{P1}$ and $T_{N1}$ while $C_A$ is charged back to $V_{BH} - V_{BL}$. The signal at the gates of $T_{P1}$ and $T_{N1}$ is defined by the capacitive voltage division of the corresponding $C_B$ and the capacitance seen at the gate of $T_{P1}$ and $T_{N1}$. Assuming $C_{B1} = C_{B2}$ and equivalent gate capacitance for $T_{P1}$ and $T_{N1}$, the transfer function can be written as:

$$\frac{V_o(s)}{V_{samp}(s)} = \frac{C_{B1}}{C_{N1} + C_{B1}} \times H_{pps}(s)$$

where $C_{N1}$ is the gate capacitance of $T_{N1}$ and $H_{pps}(s)$ is the inherent large signal frequency response of the push-pull source follower. Since the biasing network transfer function is frequency independent, DC operation is possible, which resolves the DC
baseline wander problem and prevents eye closure. The gain profile over frequency is defined by the push-pull source follower itself. The simulated gain in nominal conditions, which is shown in Fig. 4-11, is flat inside the 3dB frequency band with only 30 mDb of variation, and the 3dB frequency is around 46GHz. This configuration also allows the bottom plate parasitics of the $C_{B1}$ and $C_{B2}$ to be embedded in the sampling capacitor, eliminating excess capacitance on the signal path. However, the top plate parasitics must be decreased to reduce the loss due to capacitive division, as can be observed from equation 4.2. For this purpose, custom Metal-Oxide-Metal capacitors are used for $C_{B1}$ and $C_{B2}$, and the bottom plate is laid out encapsulating the top plate minimizing the coupling of the top plate as shown in Fig. 4-12. To reduce the resistance on the signal path, the transistors $T_{N1}$ and $T_{P1}$ are composed of unit size devices placed in an array structure. This allows the terminals of the devices to be connected to the higher level metals within close distance of the channel, which avoids routing with high resistance lower metal layers.

The bias voltage is set by a replica biasing scheme with $I_{bias}$ current defined by a current DAC as shown in Fig. 4-10. The output common mode voltage can be set independently from the input common mode voltage, which provides flexibility for compensating effects like negative charge injection from the T&H switch. This provides a well defined input common mode voltage for the asynchronous SAR sub-ADC and mainly defines the noise and speed performance of the comparator.

Figure 4-11: The gain profile of the proposed buffer obtained by periodic AC simulations in the nominal corner.
Figure 4-12: The layout techniques used for reducing the top plate parasitic capacitance of $C_{B1}$ and $C_{B2}$ and the resistance on the signal path.

4.2.4 Asynchronous SAR (ASAR) Sub-ADC

CMOS scaling leads to an increase of single channel SAR ADC sampling speeds by offering faster devices and better switches. However, higher interconnect resistance of lower metals threatens the speed advantage by increasing the time constants and settling times. Speed enhancement techniques like loop unrolling [56], especially for resolutions around 8-bit, do not provide the full speed advantage since the ADC is mainly limited by CDAC settling instead of comparator reset time. Furthermore, it increases the area due to additional comparators and offset calibration circuits for canceling the offsets of the comparators. This leads to a more significant penalty from the parasitics due to longer distance routings. The multi-bit per conversion approach [112] is also avoided for similar reasons since it necessities separate CDACs, comparators and increases the calibration cost. Thus, this work adopts a single comparator ASAR that offers compact area, power efficiency, and minimum calibration effort and maintains a good balance between maximum speed and metastability.

The top level block diagram of the ASAR is given in Fig. 4-13. The internal timing diagram is shown in Fig. 4-14. Three control signals related to the top level sub-ADC operation are shown. $CK_{64_{ADC}}$ is the sampling clock, $CK_{COMP}$ is the
Figure 4-13: Sub-ADC top level block diagram.

Figure 4-14: Internal timing of the sub-ADC.
comparator clock, and SHORT_CDAC is the control signal to short the CDAC capacitors for a reset before the subsequent conversion. The falling edge of the sampling clock $C_{K64_{ADC}}$ initiates the conversion process by raising the $C_{K_{COMP}}$ where the first decision decides on the polarity of the signal. Then the operation is organized asynchronously by sensing the outputs of the comparator. When the decision is detected, $C_{K_{COMP}}$ falls to GND to start the reset of the comparator, and when the reset is detected, $C_{K_{COMP}}$ rises to VDD to start another decision process. During the $C_{K_{COMP}}$ generation, the decision is captured by one of the memory cells which drives the CDAC switches. To prevent any distortion, the CDAC must settle before the next decision, which is controlled by careful design of the comparator clocking circuit and by reducing the delay on the CDAC path. In other words, the comparator clocking circuit can not be arbitrarily fast since the CDAC settling can limit the accuracy. In fact, there are two feedback loops during the conversion process: comparator-ckgen-comparator and comparator-memory cell-CDAC-comparator, and the slower one dictates the speed. An alternative clocking approach can be controlling the $C_{K_{COMP}}$ by sensing the outputs of the memory cells to allocate enough time for CDAC settling, yet this reduces the sampling speed considerably. At the end of the last decision, the SHORT_CDAC signal rises to short the CDAC before the following conversion to prevent any memory effect. The time allocated for SHORT_CDAC also serves as a timing margin for metastability resolution. As long as the minimum SHORT_CDAC duration does not fall below approximately 25ps, which is required for the shorting operation, the extra time can be used for comparator decisions, which means the longer the SHORT_CDAC for a regular operation, the lower the error probability for that conversion. The width of SHORT_CDAC is eventually defined by the input sample.

The S&H has a classical bootstrapped architecture given in [111]. The bandwidth is not the primary consideration at this stage since the input to the ADC is sampled and buffered by the first stage of the interleaver. Still, the linearity has to support 8-bit ADC resolution with an input swing around 0.8V which favors a bootstrapped design.
The comparator has a classical StrongArm architecture with a double differential input pair, as shown in Fig. 4-15 with output buffers and $G_M$ cells. The second input pair is connected to a resistive ladder VDAC ($V_{CALP}, V_{CALN}$) for offset calibration. The size of the calibration pair is one-fourth of the main pair in the signal path to reduce the injected noise for a given calibration range. $I_1$ and $I_2$ buffer the output. They also stop the propagation of a wrong decision in the case of metastability by sizing the NMOS and PMOS of the inverter such that the trip point is always lower than the outputs of the StrongArm when it is metastable. $G_M$ cells provide the current to drive the differential memory cells, which are connected to IGPO and IGN0. GN and GP are connected to the clock generation circuit to detect a decision or reset.

The CDAC has a differential architecture as shown in Fig. 4-16. Although the differential architecture doubles the area of CDAC, it serves a critical purpose - maintaining the input common mode of the comparator. This is desired since the
noise, speed, and offset performance of the comparator depends on the input common mode. An input common mode voltage varying with the CDAC output can lead to SNR variations for different decisions. Furthermore, it leads to decision dependent input referred offset, which leads to a less effective offset calibration. For the last three decisions, the CDAC operates in a single ended manner since the common mode variation is lower and does not affect the comparator operation significantly. To compensate for this area increase, the reference voltage is divided by four and two by a resistive divider and used for the last two decisions. This allows the unit capacitor to be chosen as C instead of C/4 and leads to a smaller total capacitance \[113\]. The unit capacitor size, defined by matching and input referred noise, is 0.25fF for this design. The capacitors are custom designed using metal layers, which create fringe capacitance.

It is critical to reduce the RC time constant of the decision path to the CDAC since the settling accuracy affects the resolution of the ADC. For this purpose, the CDAC is rotated such that the MSB decisions have the shortest path. The switching of MSBs leads to the highest voltage changes, hence requiring the longest time to settle. The decision wiring from the memory cells is reinforced with multiple metal layers to reduce the wiring resistance since the settling time constant is dominated by the

![Diagram](image_url)
Figure 4-17: The schematic of the super source follower reference buffer.

The CDAC is implemented with upper metal layers such that lower metal layers can be used for routing the inverted decisions (e.g. \( D_{7P}, D_{7N} \)) to the other half of the CDAC. The dummy capacitors are placed on the edges to reduce the effect of mismatch by imitating the same boundary conditions with the other capacitors.

For the reference buffer, a super source follower topology is adopted as shown in Fig. 4-17. This offers low output resistance and fast settling time with relatively low current consumption. \( T_{P2} \) and \( T_{N2} \) mirror the currents obtained from current DACs and set the bias currents for \( T_{P1} \) and \( T_{N1} \). The gate of \( T_{P1} \) (\( V_{REFSET} \)) is connected to a 16-bit resistive ladder that sets the reference voltage and the full scale voltage for the ADC. The feedback loop provides a low output resistance. The sinking current is set by \( T_{N1} \), which can be sized to sink a considerable amount to reduce the settling time. The sourcing current is set by \( T_{P2} \) and \( T_{N2} \) which allows biasing of \( T_{N1} \) independent of \( T_{P1} \).

The comparator outputs at each decision cycle have to be captured by memory cells that drive the CDAC switches. A total of eight memory cells work in a domino sequence: after a decision is captured by a memory cell, it enables the next memory cell. The operation starts with a global reset of all the memory cells. The memory cell schematic is illustrated in Fig. 4-18. There are three control signals: ENB_REG,
KEEP, and RES, which are generated by the memory cell clocking logic. When RES is low, TP₃-TP₄ and TP₅-TP₆ reset the memory cell to its initial state. ENB_REG is asserted to enable the corresponding memory cell. TN₁ and TN₂ act as switches. Depending on the comparator decision, either node X or node Y is discharged. This activates the cross coupled inverter pair TP₁-TP₂, TN₃-TN₄ that regenerates and stores the decision. After the decision is taken ENB_REG is deasserted, and the memory cell is disconnected from the $G_M$ cell outputs IGPO and IGNO. The KEEP signal is asserted to keep the decision until the next ADC conversion cycle.

Differential architecture introduces a secondary input dependent regeneration. In essence, it operates similarly to a StrongArm latch. This reduces the metastability event propagation to the CDAC. This point is illustrated in Fig. 4-19. In the case of a single ended memory cell, when the differential input is close to zero, the outputs of the comparator can lead to the capture of a high state by both cells, which is a state that must not exist. Taking another input dependent decision in the memory cell prevents this state. In addition, it reduces the time during which the CDAC switch inputs remain at a metastable state since either the comparator or the memory cell eventually regenerates out of the metastable state due to noise. The measurement results prove the effectiveness of this technique where no sparkle codes stemming from
Figure 4-19: Single ended vs. differential memory cell architecture and the possible output states.

the metastability are detected.

4.3 Calibration

Time-Interleaved ADCs are plagued by interleaving spurs that stem from the mismatch between sub-channels. The main contributors are offset, gain, skew, and bandwidth mismatches which create significant harmonic distortion. To alleviate this problem, the ADC in this work utilizes full foreground calibration. This is realized by the following control knobs placed at various circuit blocks:

1. Sub-ADC

   - Offset: 8-bit R-2R ladder
   - Gain: 16-bit R-2.5R ladder
2. Interleaver

- Skew: 6-bit MOS capacitor array
- Bandwidth: 5-bit MOS capacitor array

3. Clock Receiver

- Duty Cycle: 2x8-bit R-2R ladder

The calibration procedure includes three steps where the input is a DC signal, a low frequency sinusoidal signal, and a high frequency sinusoidal signal. The steps are explained below:

1. **DC input**: Offset is calibrated with a DC input using the 8-bit resistive ladder and the double differential input pair of the comparator in the sub-ADC level when the input is at 0V differential. The R-2R ladder is programmed for each sub-ADC to drive the corresponding output to 128, which is the middle code. The distribution of the output values after the offset calibration is shown in Fig. 4-20 (a). From the histogram, the estimation of the noise performance can also be found by extracting the standard deviation. The higher the noise, the wider the spread of the output values with 0V differential input. From the standard deviation, the SNR can be calculated, which shows around 7.08 ENOB. Fig. 4-20 (b) shows the values of 1024 samples collected from the shift registers, and Fig. 4-20 (c) illustrates the codes applied to each of the R-2R ladder of the 64 sub-ADCs in order to obtain 128 at the outputs. Similarly, the gain is calibrated with an input 0.2dB backed off from the full scale 0.8V_{pp, diff} (corresponds to 252 at the output) using the 16-bit R-2.5R resistive ladder connected to the reference buffer of the corresponding sub-ADC. 4-20 (d) (e) and (f) illustrate the output distribution, output code values, and the codes applied to the corresponding R-2.5R ladder.

2. **Low Frequency Sine Input**: With the configuration values from the previous step serving as a starting point, the following calibration steps are carried out with a sinusoidal input signal. First, a low frequency 4.1GHz sine signal is
applied to the input, and the offset and gain calibrations are repeated. The output amplitude of the signal generator is adjusted to obtain the same values at the output of the ADC as in the DC test to compensate for the losses from the measurement setup. Then the duty cycle is adjusted at the clock receiver with two 8-bit R-2R ladders. The duty cycle setting is swept to find the maximum ENOB when the ADCs are divided into two groups clocked by I and Q phases. The output can be perceived as two 32x interleaved ADCs which have I and Q clocks. The aggregate ENOBs vs. the duty cycle R-2R value is shown in Fig. 4-21 (a). After the duty cycle correction, the timing skews between the 16 first rank T&H blocks must be corrected. This is done by the 6-bit programmable capacitors on the first rank interleaver clock path. They are programmed such that each T&H sampling pulse has a time difference of 17.86ps from the previous pulse, which corresponds to 1UI. The output of each T&H can be observed by the aggregate output of the four sub-ADCs which are driven by it. The output phases of these 16 groups of 4 can be obtained by sine fitting, and a search algorithm finds the code to be applied to the programmable capacitor to fix the time skews between them to 17.86ps. The results, together with the code applied to the programmable capacitor, are shown in Fig. 4-21 (b) and (c).

3. **High Frequency Sine Input:** In the final step, a high frequency sine, which is at 26.1GHz, is applied to the input. In the high frequency region, the bandwidth difference between different sub-channels introduces distortion. To prevent this, the bandwidth calibration is done using the 5-bit capacitor array in parallel with the sampling capacitors of each of the 16 first rank T&H. The sampled and buffered amplitude at each of the 16 branches of the first rank of the interleaver can be observed in a similar manner to the timing skew by capturing the combined outputs of each 4 sub-ADCs groups. The code applied to the programmable capacitor is found by searching the value of 125 at the output, which corresponds to 0.2dB back-off from $0.8V_{pp,diff}$ at the input. The results are illustrated in 4-21 (d) and (e). After the bandwidth calibration, the relative skew information between different sub-ADCs is lost; hence the skew calibration
Figure 4-20: Gain and offset calibration steps. (a) The distribution of the output due to noise after offset calibration. (b) The output values corresponding to 1024 samples collected. (c) The digital code applied to each ADC to calibrate the offset. (d) The distribution of the output after gain calibration. (e) The output values corresponding to 1024 samples collected after gain calibration. (f) The digital code applied to each ADC to correct gain.

is repeated again, and the result is shown in (f). The duty cycle of the I and Q clocks is also corrected again. With this final step, all the calibration knobs are configured, and the measurements can be taken.

4.4 Measurement Results

The ADC test chip is fabricated in Samsung 4nm FinFET technology. Fig. 4-22 shows the die photo and the ADC layout details. The active circuit occupies 0.078 $mm^2$. The ADC test chip is placed on a GL102F 4-2-4 substrate and mounted to a Nelco4000-13 test card for characterization. Reported measurement results are taken without any post-processing after analog foreground calibrations are performed and stored on the chip. A single 0.8V supply and $0.8V_{pp,diff}$ input at 56GS/s are used during these measurements.
Figure 4-21: I-Q duty cycle, skew, and bandwidth calibration steps. (a) The code applied to tune duty cycle of the I-Q clocks vs. ENOB (b) The skew values between the first rank T&Hs after timing skew correction. (c) The digital codes applied for skew correction. (d) The amplitude at the output of each 16 first rank channels after bandwidth correction. (e) The digital codes applied to bandwidth correction programmable capacitors. (f) The timing skew values after the final skew calibration.

Figure 4-22: Chip micrograph and the layout of the proposed ADC.
4.4.1 Sinusoidal input tests

The ADC is tested with a slow sinusoidal input to characterize the static linearity. The INL and DNL curves are extracted using a histogram test with a single tone sine input at 4.1GHz. A full scale input of $0.8\text{V}_{\text{pp,diff}}$ is applied, and 128000 data points are collected. The obtained curves are given in Fig. 4-23. The results indicate that the DNL values remain between $-0.77$LSBs and $1.69$LSBs, which shows that there are no missing codes. INL values remain between $-1.14$LSBs and $1.13$LSBs.

An important point to note is the occurrence of the DNL spikes located at 16, 32, 64, 240 (256-16), 224 (256-32), 192 (256-64). These indicate the settling errors in the CDAC when the MSB bits are switched. Hence, they prove that for this design implemented in 4nm technology, the sampling rate is limited by the CDAC settling and not by the clock generation loop of the comparator.

Then the dynamic characterization is carried out. For these measurements, a sinusoidal signal with a swing 0.2dB backed off from the $0.8\text{V}_{\text{pp,diff}}$ is applied to the ADC. The input frequency is swept up to the Nyquist frequency, and the SFDR and the SNDR values are captured. The results are shown in Fig. 4-24. The SNDR starts around 42dB and drops to approximately 33dB at the Nyquist frequency, while the SFDR stays above 42dB over the Nyquist range. Again using a similar procedure, the input frequency is swept to obtain the 3dB bandwidth. Including the losses from the substrate, the cables, and the test card, the bandwidth is found as 27GHz. The frequency response is shown in Fig. 4-25.
Figure 4-24: SNDR and SFDR values over the first Nyquist zone.

Figure 4-25: SAR frequency profile including the cable, the test card, and the substrate losses.
Figure 4-26: SAR frequency output spectrum corresponding to $f_{\text{in}}$=4.1GHz with input amplitude 0.2dB backed off from 0.8$V_{pp,\text{diff}}$.

- $F_{\text{sine}}$ = 4.105 GHz
- $A_{\text{mp}}$ = -0.2dBFS
- $S_{\text{FDR}}$ = 61.9dBc
- $H_{\text{D2}}$ = -68.7dBc
- $H_{\text{D3}}$ = -66.1dBc
- $T_{\text{HD}}$ = -60.6dB
- $S_{\text{NR}}$ = 41.0dB
- $S_{\text{NDR}}$ = 41.0dB
- $E_{\text{NOB}}$ = 6.5

Figure 4-27: SAR frequency output spectrum corresponding to $f_{\text{in}}$=25.1GHz with input amplitude 0.2dB backed off from 0.8$V_{pp,\text{diff}}$.

- $F_{\text{sine}}$ = 25.119GHz
- $A_{\text{mp}}$ = -0.2dBFS
- $S_{\text{FDR}}$ = 41.7dBc
- $H_{\text{D2}}$ = -44.1dBc
- $H_{\text{D3}}$ = -41.6dBc
- $T_{\text{HD}}$ = -39.7dB
- $S_{\text{NR}}$ = 34.9dB
- $S_{\text{NDR}}$ = 33.6dB
- $E_{\text{NOB}}$ = 5.3
The output spectra of the ADC are obtained for low and high frequency input. The spectrum corresponding to 4.1GHz input frequency is shown in Fig. 4-26. The input swing is again 0.2dB backed off from 0.8V_{pp\_diff}. At 4.1GHz, total harmonic distortion (THD) is less than -60.6dB, and the SFDR is at 61.9dBc. SNDR is around 41dB which corresponds to an ENOB of 6.5. As can be seen from the spectrum, the harmonics and interleaving spurs are well below -60dBc level indicating the effectiveness of the calibration procedure. The output spectrum at 25.1GHz input frequency is shown in Fig. 4-27. THD increases to -39.7dB while the SFDR is at 41.7dBc. The SNDR decreases to 33.6dB, which indicates an ENOB around 5.3. The resolution degrading effects of harmonics and the interleaving spurs become more dominant. Especially interleaving spurs due to timing skew and bandwidth mismatch arise at high frequencies since they are frequency dependent. Some portion of the even order harmonics are contributed by the input balun due to common mode coupling. However, the dominant effect of the resolution degradation is due to the jitter at this input frequency.

To evaluate the jitter performance, 16000 samples are collected and folded in one period of the sinusoidal input signal as shown in Fig. 4-28 (a) to obtain a modulo time plot. Then a sine fit is applied to the samples as marked by the red line. The error between the samples and the fitted sine creates a residue distribution at each time point which is illustrated in Fig. 4-28 (b). This residue distribution is an indicator of noise at each point. The errors stem from the jitter and the thermal noise when the signal derivative is the highest and mainly from the thermal noise when the signal derivative is the lowest. Thus the standard deviations of the residues create a cosine wave, and from the amplitude and the offset of this cosine, the jitter can be extracted. For this ADC, the jitter is found as 98fs at 26.9GHz. According to equation 2.24, the ENOB for an ideal 8-bit equalizer with the calculated jitter standard deviation can be found as approximately 5.6, which indicates the design is limited by jitter close to Nyquist frequency. In practice, the effect of the jitter on resolution degradation starts earlier. The same technique to extract the jitter can be applied to find the rms voltage error due to thermal noise and jitter over the input frequencies across the
Figure 4-28: Jitter calculation steps. (a) The modulo time plot with the collected samples. (b) The distribution of voltage error at each time point. (c) The standard deviation of the error voltages fitted to a cosine wave.

first Nyquist zone. Fig. 4-29 illustrates the results. The crossover point is around 12GHz, and when the input frequency is higher than this, the errors due to jitter dominate.

In order to demonstrate the linearity of the interleaver against different $V_{in,cm}$ conditions stemming from the AFE, $V_{in,cm}$ is swept between 0.3V to 0.6V, and the THD is measured for two input frequencies $f_{in} = 19.8GHz$ and $f_{in} = 4.1GHz$ as shown in Fig. 4-30. The input is 0.2dB backed off from the full scale $0.8V_{pp,diff}$. The THD stays below -53dB inside the sweep range for 4.1GHz input frequency and below -41dB for 19.8GHz input frequency. These results demonstrate the robustness of the ADC against common mode variations and prove the linearity of the interleaver.

### 4.4.2 Modulated Data Experiments

The ADC is designed to support a high speed wireline RX. As a result, it is also tested with modulated data. Although the equalization capability is missing on the RX side due to a lack of DSP, the FFE equalized data can be transmitted on the TX side. The measurements are taken without a channel, and only signal integrity degrading effects arise due to high speed cable, connectors, test board, and substrate, so only FFE equalization is adequate. The results are analyzed by a statistical post processing tool, and FEC is applied. The error correction code and the signaling are not within the scope of this thesis. This subsection demonstrates the capability of the proposed high speed ADC in supporting higher order modulation formats.
Figure 4-29: The rms error voltages due to jitter and thermal noise across the first Nyquist zone.

Figure 4-30: The THD values corresponding to the input common mode $V_{in,cm}$ sweep from 0.3V to 0.6V when the input is backed off 0.2dB from $0.8V_{pp,\text{diff}}$ at input frequencies 4.1GHz and 19.8GHz.
As a first step, a pseudorandom binary sequence (PRBS) is generated and loaded to an arbitrary waveform generator (AWG) that acts as a TX, which drives the ADC. The channel impulse response can be found from the deconvolution of the transmitted and received data. The channel in this case shows 8dB loss at Nyquist input frequency. After the frequency profile of the channel is obtained, the FFE equalized data can be generated and transmitted from the AWG. The eye diagrams for the 112Gb/s PAM4 modulation, and 168Gb/s PAM8 modulation at the RX are shown in Fig. 4-31. PAM4 raw data with 396506 bits is captured error free. PAM8 raw data with 503118 bits has a BER of $2.4 \times 10^{-3}$. However, after the FEC is applied with block coded modulation with an inner binary product code and outer Reed Solomon KR4 code, the received bits are error free. The SNDR is measured as 22dB for PAM4 and 22.3dB for PAM8.

In order to demonstrate the performance of the ADC for frequency domain modulation techniques, it is tested with 64QAM OFDM data. 255 sub-carriers are utilized in the first Nyquist zone. Using block coded modulation with an inner binary product code and outer Reed Solomon KR4 code results in a data rate of 184Gb/s. The received 64QAM constellation is demonstrated in Fig. 4-32(b). The SNDR profile for all the channels and the transmitted signal power is shown in Fig. 4-32(a). Power loading is applied to increase the SNDR of the sub-channels that suffer from low SNDR. These include mainly the channels close to the Nyquist frequency. The re-
sulting average sub-channel SNDR is 22.6dB. The received data is error free with block coded modulation with an inner binary product code and outer Reed Solomon KR4 code.

4.5 Conclusions and Comparison with the Prior Art

A 56GS/s 64x time interleaved asynchronous SAR ADC is discussed in this chapter. The interleaver provides excellent linearity thanks to the bootstrapped switch and push-pull follower architectures introduced. Combined with the power efficiency of the asynchronous SAR, the architecture is a promising candidate for the next generation high speed wireline links tailored for complex modulation schemes. To prove this point, in addition to the ADC characterization tests, the modulated data experiments are also carried out. A comparison table with the prior art is given in Fig. 4-33.

Figure 4-32: The signal power and the SNDR of the OFDM sub-channels and the received 64QAM constellation diagram.
<table>
<thead>
<tr>
<th>Technology</th>
<th>[8]*</th>
<th>[7]*</th>
<th>[2]</th>
<th>[6]</th>
<th>This Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>7nm</td>
<td>20nm</td>
<td>14nm</td>
<td>28nm</td>
<td>4nm</td>
</tr>
<tr>
<td>(bits)</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>fs (GS/s)</td>
<td>97</td>
<td>64</td>
<td>72</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>SNDR @ fin,low dB @ GHz</td>
<td>41@DC</td>
<td>37.6@8</td>
<td>39.3@2</td>
<td>40.5@7</td>
<td>41.0@4.1</td>
</tr>
<tr>
<td>SNDR @ fin,high dB @ GHz</td>
<td>32@36</td>
<td>33.7@16</td>
<td>32.7@27.5</td>
<td>33@27</td>
<td>32.8@26.9</td>
</tr>
<tr>
<td>INL (min,max)</td>
<td>NA</td>
<td>NA</td>
<td>(-2.37, 4.62)</td>
<td>NA</td>
<td>(-1.14, 1.13)</td>
</tr>
<tr>
<td>BW (GHz)</td>
<td>&gt;40</td>
<td>20</td>
<td>21</td>
<td>31.5</td>
<td>&gt;27</td>
</tr>
<tr>
<td>Power (mW)</td>
<td>305</td>
<td>950</td>
<td>235</td>
<td>702</td>
<td>240</td>
</tr>
<tr>
<td>FOM @ fin,low (fJ/conv.step)</td>
<td>34</td>
<td>240</td>
<td>43</td>
<td>145</td>
<td>47</td>
</tr>
<tr>
<td>FOM @ fin,high (fJ/conv.step)</td>
<td>96</td>
<td>375</td>
<td>121</td>
<td>344</td>
<td>117</td>
</tr>
<tr>
<td>Supply (V)</td>
<td>1st rank@1.2 2nd rank@0.9</td>
<td>1.5, 1</td>
<td>sub-ADC@0.8 Inerleaver@0.9</td>
<td>sub-ADC@0.95, Input Buffers @0.95/-0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>NA</td>
<td>~1.7</td>
<td>0.15 (active)</td>
<td>0.878 (active)</td>
<td>0.078(active)</td>
</tr>
</tbody>
</table>

*Includes AFE

Figure 4-33: Comparison with prior art.
Chapter 5

Conclusion

Increasing per lane data rates fueled by data centric applications requires spectrally efficient modulation techniques, which necessitate complex modulation and signal detection algorithms. This is enabled by ADC-DSP based receiver architectures. High speed and power efficient ADC architectures are key enablers for these receivers. Innovation in ADC architectures is necessary for reaching sufficient resolution while increasing the sampling rates above 100GS/s and reducing power dissipation. For this purpose, this thesis concentrates on the implementation of two ADC architectures: time-based and SAR.

There has been a growing interest in time-based ADC architectures in the last decade. The search for more efficient architectures with higher sampling rates leads to the adoption of innovative time-based techniques. The biggest reason behind this is the superior scaling properties of time-based ADCs, thanks to their digital-intensive design. In other words, they rarely utilize analog blocks, and most of the functionality can be accomplished by digital style circuits. Thus, they are promising candidates for the next generation wireline links implemented in FinFET technologies due to compactness, speed, and power efficiency. Future research directions can also include the possibility of hardware synthesis for these architectures. In this thesis, a novel 8-bit time-based ADC architecture is proposed that is implemented in 5nm CMOS tech-
Figure 5-1: The areas of single channel ADCs published in ISSCC and VLSI implemented in CMOS technology with sampling rates over 500MS/s. The time-based ADC presented in this thesis is labeled as ISSCC23.

nology. It is envisioned to be an excellent candidate for a sub-ADC in an interleaved architecture. Three techniques are introduced which can be listed as: 1) a bipolar ramp-based voltage-to-time converter to eliminate the reference voltage and to allow a wide input swing of 0.75\(V_{pp,\text{diff}}\); 2) \(2\times\) interpolating sense-amplifier-latches to reach a sub-gate delay of 2.3ps in a power and area efficient manner; and 3) redundancy to accommodate possible wrong decisions of the sign bit comparator allowing a minimum size design without any calibration. The designed single-stage, calibration-free, 8-bit, time-based ADC achieves 16.6fJ/conv-step FoM with 313\(\mu m^2\) area. The FoM is among the best published in this frequency range, while the ADC occupies the smallest area among previously demonstrated architectures in ISSCC and VLSI. The results are published in ISSCC23. A comparison of FoM against all the published papers in ISSCC and VLSI is given in Fig. 5-2 [115]. Fig. 5-1 shows the area comparison of single channel ADCs with sampling rates higher than 500MS/s published in ISSCC and VLSI. The closest design [68] has 6-bit resolution [115].

SAR ADC is a proven architecture that offers high power efficiency and scalability. With the help of techniques like asynchronous timing, it also offers sampling rates in the range of GS/s. This is the reason why the majority of interleaved architectures
today utilize SAR ADC. This thesis presents a 56GS/s interleaved SAR ADC suitable for wireline receivers targeting wireline channels modulated with higher order modulation formats. 56GS/s 8-bit asynchronous SAR ADC is fabricated in 4nm CMOS technology. The 16x4 interleaved ADC uses a novel bootstrapping technique and a class-AB follower in the 1st rank interleaver. It achieves a broad input common mode $V_{in,cm}$ range; from 0.3V to 0.6V, the total harmonic distortion stays below -52dB at 4.1 GHz with -0.2dBFS amplitude at $0.8V_{pp,diff}$ maximum full scale. The ADC includes analog foreground calibration means for offset, gain, skew, and bandwidth. The measured ENOB is 6.5 at low frequency and stays above 5.2 up to Nyquist frequency, which is limited by the sampling jitter. The bandwidth is higher than 27GHz. The ADC uses a single 0.8V supply voltage and achieves an efficiency of 47fJ/conv-step at 4.1GHz input frequency and 117fJ/conv-step at 26.9GHz input frequency. A Walden FoM comparison with previously published ADCs in VLSI and ISSCC is given in Fig. 5-2 [115].


160


[63] Ahmed ElShater, Praveen Kumar Venkatachala, Calvin Yoji Lee, Jason Muhlestein, Spencer Leuenberger, Kazuki Sobue, Koichi Hamashita, and Un-Ku Moon. A 10-mW 16-b 15-MS/s Two-Step SAR ADC With 95-dB DR Us-


[85] Shuang Zhu, Benwei Xu, Bo Wu, Kiran Soppimath, and Yun Chiu. A 0.073-mm2 10-GS/s 6-bit time-domain folding ADC in 65-nm CMOS with inherent DEM. In 2015 IEEE Custom Integrated Circuits Conference (CICC), pages 1–4, 2015.


