HIGH-SPEED SAR ADC DESIGN IN SIGE BiCMOS TECHNOLOGY

by

Peter Hermansen

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Graduate Department of Electrical Engineering
University of Toronto

© Copyright 2019 by Peter Hermansen
Abstract

High-Speed SAR ADC Design in SiGe BiCMOS Technology

Peter Hermansen
Master of Applied Science
Graduate Department of Electrical Engineering
University of Toronto
2019

This thesis investigates architectures and design techniques for high-bandwidth SAR ADCs in SiGe BiCMOS technology to be used in 64 GBAud fibre-optic systems. First, a 6-bit SAR ADC operating with the highest sampling rate of 3 GS/s was designed and characterized in a 55nm SiGe BiCMOS process. The circuit achieved an ENOB larger than 3.8-bits at 3.0 GS/s, and larger than 4.3-bits at 2.5 GS/s. The ADC consumes 70mW resulting in a FOM of 1.326 pJ/conversion at 2.5 GS/s. Next, a novel SAR architecture was developed and integrated in a time-interleaved, 6-bit, 64GS/s SAR ADC. It uses the same 55nm process, and a total of 16 SAR lanes. Simulation results demonstrate over 4-bits ENOB up to a 19GHz input signal, and up to 3.6-bits ENOB at 31GHz. The ADC power consumption is 1.57W, resulting in a FOM of 2.07 mW/bit/GHz.
Acknowledgements

First and foremost I would like to thank my supervisor, Professor Sorin P. Voinigescu. I am grateful for his guidance and support over the past two years. I would also like to thank Professor David Johns, Professor Antonio Liscidini, and Professor Amr Helmy for their participation in my examination committee.

I would like to thank Alireza Zandieh, and Naftali Weiss. Without their countless technical discussions and advice, these past couple years would have not gone as smoothly. I would also like to thank my colleagues from BA4182 Robert Baker, Shai Bonen, Utku Alakusu, Jashva Rafique, Mecca Gong and Sadegh Dadash. I would like to acknowledge Professor Eran Socher for his contributions to the project. Thank you to Dan Case, Steve Troyer, and The’ Linh Nguyen from Finisar Corporation for the opportunity work on this project, and their support during it.

Thank you to the Finisar Corporation for their financial support, and STMicroelectronics for providing design kits and chip fabrication. I would also like to acknowledge CMC and Jaro Pristupa for CAD support.

Finally I would like to thank my friends and family, especially my wife Linda for joining me as I moved across the country for graduate school, and as I make my next career steps.
Contents

1 Introduction 1
  1.1 Motivation .............................................. 1
  1.2 Objective .............................................. 2
  1.3 Thesis Organization ...................................... 3

2 Background 4
  2.1 ADC Performance Metrics .................................. 4
  2.2 High-Speed ADC Architectures ............................. 5
    2.2.1 Flash ADCs ........................................... 5
    2.2.2 SAR ADCs ............................................. 6
    2.2.3 Pipelined ADCs ........................................ 7
  2.3 SAR ADC Architectures .................................... 8
    2.3.1 Synchronous SAR ADC ................................ 8
    2.3.2 Asynchronous SAR ADC ................................ 9
  2.4 Time-Interleaving ....................................... 10
    2.4.1 Direct Interleaving Architecture ..................... 10
    2.4.2 Sub-Sampling Switch Architecture ..................... 11
    2.4.3 In-line Demultiplexer Architecture .................... 12

3 SAR ADC Lane 14
  3.1 Overview ............................................... 14
  3.2 High-Speed Logic ........................................ 15
    3.2.1 MOS CML vs Bipolar CML ............................. 15
    3.2.2 Quasi-CML Logic ..................................... 16
    3.2.3 Inductive Peaking CML Logic ........................ 17
    3.2.4 Set and Reset Latch Design .......................... 20
  3.3 Track and Hold Amplifier ................................. 21
    3.3.1 Input Buffer ......................................... 22
    3.3.2 Switched Emitter Follower ............................ 23
    3.3.3 Output Buffer ........................................ 25
  3.4 R2R DAC Design .......................................... 26
  3.5 High-Speed Comparator ................................... 27
  3.6 Sequencer ............................................... 29
  3.7 SAR ADC Lane Design Summary ............................. 31
4 Experimental Results

4.1 Sequencer .................................................. 33
  4.1.1 Test Setup ........................................... 33
  4.1.2 Measurement Results ................................. 35
  4.1.3 Performance Summary and Comparison ............... 35

4.2 SAR ADC Lane ............................................. 38
  4.2.1 Test Setup ........................................... 38
  4.2.2 Measurement Results ................................. 40
  4.2.3 Performance Summary and Comparison ............... 46

5 Time-Interleaved SAR ADC ................................. 49

5.1 System Overview .......................................... 49
5.2 SubADC Architecture ..................................... 51
5.3 Clock Generation and Distribution ...................... 52
5.4 Linear Sampled Signal Distribution ....................... 58
5.5 Layout and Post Layout Simulation Results ............ 60

6 Conclusion .................................................. 66

6.1 Summary .................................................. 66
6.2 Future Work ............................................... 66

Bibliography .................................................. 68
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Comparison with state of the art high-speed clock generation circuits</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparison with state of the art SAR ADCs</td>
<td>48</td>
</tr>
<tr>
<td>5.1</td>
<td>Breakdown of time-interleaved ADC Power Consumption</td>
<td>62</td>
</tr>
<tr>
<td>5.2</td>
<td>Comparison with state-of-the-art high-speed ADCs</td>
<td>65</td>
</tr>
</tbody>
</table>
## List of Figures

1.1 Key technologies and baud rates in coherent optical systems [1]. .................................. 2

2.1 A Basic 2 bit flash ADC. ................................................................. 6
2.2 Flow chart for SAR conversion Algorithm. .................................................. 7
2.3 Block diagram for a basic SAR ADC. ..................................................... 7
2.4 Block diagram for a basic pipeline ADC. ................................................... 8
2.5 Timing diagram for a 4-bit synchronous SAR ADC. ...................................... 9
2.6 Timing diagram for a 4-bit asynchronous SAR ADC. ................................. 10
2.7 Example of direct interleaving architecture. ................................................ 11
2.8 Example of sub-sampling switch interleaver architecture. ............................ 12
2.9 Example of an in-line demultiplexing interleaver architecture. ..................... 13

3.1 SAR ADC lane block diagram. ............................................................... 14
3.2 CML based inverter circuits using (a) MOSFETs and (b) HBTs. ....................... 15
3.3 CML based latch circuit using (a) a standard current tail and (b) a quasi-CML tail. 16
3.4 CML based inverter circuits using (a) passive peaking and (b) active peaking. 17
3.5 Passive model for active inductor circuit.................................................... 18
3.6 CML latches with 3 or more inputs (a) standard CML asynchronous reset latch (b) quasi-CML synchronous BiCMOS SR latch (c) asynchronous reset function achieved through cascaded gates. ........................................... 20
3.7 Quasi-CML latches with (a) synchronous set and reset and (b) asynchronous reset. 21
3.8 Schematic of THA operating in track mode.................................................. 21
3.9 Schematic of THA used in the ADC Lane................................................. 22
3.10 Schematic of THA operating in track mode................................................ 23
3.11 Schematic of THA operating in hold mode.............................................. 24
3.12 R2R type current DAC. ....................................................................... 26
3.13 Schematic of the R2R DAC. .................................................................. 27
3.14 Schematic of the high-speed comparator and latch. ........................................ 28
3.15 Sequencer circuit for SAR ADC lane clock generation. .............................. 29
3.16 Clock generation circuit, all blocks employ active peaking CML Logic. ........... 30
3.17 SAR ADC lane block diagram. .................................................................. 31
3.18 Block diagram of complete chip. ................................................................ 32
3.19 Output buffer used to read out DAC value. .............................................. 32

4.1 Sequencer die micrograph. ........................................................................ 34
4.2 Measurement set up for the sequencer. ........................................... 34
4.3 Measurement of single output phase using a 42GHz input clock. ................. 35
4.4 Frequency domain measurement of single output phase using a 42GHz input clock. The attenuation of tones at higher frequencies is partly due to cable and probe loss which has not been calibrated out. ....................................................... 36
4.5 Measurement of single output phase using a 5GHz input clock. ...................... 36
4.6 Measurement of duty cycle vs. input clock frequency. ............................... 37
4.7 SAR ADC die micrograph. ............................................................... 38
4.8 Measurement set up for the SAR ADC. .............................................. 39
4.9 Photo of measurement set up for the SAR ADC. ..................................... 39
4.10 Measurement of SFDR vs. input clock frequency for different sampling rates. ... 40
4.11 Measurement of SNDR vs. input clock frequency for different sampling rates. ... 41
4.12 Measurement of ENOB vs. input clock frequency for different sampling rates. ... 41
4.13 Measured output spectrum at 2.5GS/s sampling rate and a 821MHz input tone. The first 11 harmonics are marked, the resolution bandwidth is 500 kHz. ......................... 42
4.14 Measured output spectrum at 3GS/s sampling rate and a 821MHz input tone. The first 11 harmonics are marked, the resolution bandwidth is 500 kHz. ......................... 42
4.15 Measured output spectrum at 3.5GS/s sampling rate and a 821MHz input tone. The first 11 harmonics are marked, the resolution bandwidth is 500 kHz. ......................... 43
4.16 Measured output amplitude normalized to full scale amplitude vs. input frequency. . 44
4.17 Measured output amplitude, SFDR, and SNDR at 2.5 GS/s and 3.0 GS/s, for tones down-converted to 311 MHz in odd Nyquist bands. .......................... 44
4.18 Measured SNDR vs input amplitude normalized to the 600mV differential full scale using a 311 MHz tone. ............................................................. 45
4.19 Measured integral non-linearity. ...................................................... 45
4.20 Measured differential non-linearity. ................................................... 46
4.21 Large metastability error observed in the DAC output signal at 3.5 GS/s. .......... 47

5.1 The time-interleaved ADC block diagram. With the exception of the input clock, input reset, and output data line, all other connections shown here are differential. ............... 50
5.2 The proposed modified SAR ADC lane, and corresponding clocking diagram. The R2R DAC uses a 1.8V supply, all other blocks shown here use a 2.5V supply. .......... 52
5.3 Proposed clock distribution to synchronous SAR subADCs. ......................... 53
5.4 Clock buffers used between sequencer and subADC lanes. The final stage is terminated and the end of the subADC lanes. ........................................... 54
5.5 Circuits used to distribute the clock to the sequencers, comparator clock generation blocks and MTHAs. All circuits shown here are implemented with passive peaking from a 1.8V supply. .................................................. 55
5.6 Clock delay cells used to vary the clock delay between the MTHAs, sequencer A and sequencer B. ................................................................. 56
5.7 Circuits used for ADC Lane clock generation. All circuits shown here are implemented with active peaking, and use a 2.5V supply. .......................... 57
5.8 Schematic of half of the time-interleaved front end. .................................. 59
5.9 Layout of the time-interleaved SAR ADC. ......................................... 61
5.10 Simulated ADC performance at 72 GS/s. The SFDR and SNDR are presented after the
MTHAs, STHAs and after conversion. ................................................................. 63
5.11 Simulated ADC performance at 64 GS/s. The SFDR and SNDR are presented after the
MTHAs, STHAs and after conversion. ................................................................. 63
5.12 Simulated ADC ENOB at 64 GS/s and 72 GS/s. ........................................... 64
5.13 Simulated full scale output amplitude for varying input frequencies at 72 GS/s. ........ 64
List of Acronyms

ADC  Analog-to-Digital Converter
ASG  Analog Signal Generator
AWG  Arbitrary Waveform Generator
BiCMOS  Bipolar and Complementary Metal-Oxide-Semiconductor
CML  Current-Mode Logic
CMOS  Complementary Metal-Oxide-Semiconductor
DAC  Digital-to-Analog Converter
DNL  Differential Nonlinearity
DR  Droop Rate
DSP  Digital Signal Processing
ENOB  Effective Number of Bits
FinFET  Fin Field Effect Transistor
HBT  Heterojunction Bipolar Transistor
INL  Integral Nonlinearity
LSB  Least Significant Bit
MOSFET  Metal-Oxide-Semiconductor Field-Effect Transistor
MSB  Most Significant Bit
MTHA  Master Track and Hold Amplifier
QAM  Quadrature Amplitude Modulation
QPSK  Quadrature Phase Shift Keying
RTOS  Real Time Oscilloscope
SAR  Successive Approximation Register
SFDR  Spurious-Free Dynamic Range
SiGe  Silicon Germanium

SNDR  Signal-to-Noise and Distortion Ratio

SNR  Signal-to-Noise Ratio

SR  Slew Rate

STHA  Slave Track and Hold Amplifier

THA  Track and Hold Amplifier
Chapter 1

Introduction

1.1 Motivation

In 2010, the first commercial coherent optical systems were deployed targeting 100Gb/s data rates. These first generation systems operating at 28-32 GBaud require substantial digital signal processing (DSP) to realize dual-polarized QPSK optical links [2]. This modulation format was key in realizing these high data rate systems. To support these modulation schemes, high sampling rate analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) become critical, in some implementations requiring double sampling to provide sufficient equalization for long haul links [3]. Since then, coherent optical technologies have continued to drive data rates higher, while reducing the cost per Gigabit with 200Gb/s systems being deployed in 2014, and with the roll-out of the most advanced third generation of coherent optics currently underway. Third generation coherent 400Gb/s optical systems require high symbol rates, 56-69 GBaud, and complex modulation formats, such as 64-QAM [1].

Coherent technology requires substantial DSP to remove linear noise from the channel and correct for other non-idealities. In order to implement these systems, high-speed data converters are needed with a economically viable power consumption. Early implementation of these ADC based receiver fiber-optic systems targeted 10 to 40Gb/s data rates, and demonstrated for the first time a practical solution for coherent optical receivers [4, 5]. This solution utilized a highly parallelized, time-interleaved CMOS successive approximation register (SAR) ADCs, and has since become a very common approach to implement receivers in ultra broadband communication systems.

High-speed and medium resolution applications have made SAR the dominant architecture targeting optical commutation applications. Digital logic scaling has allowed time-interleaved SAR ADCs to reduce footprint and power as CMOS technology nodes have progressed. Implementation in advanced 14nm FinFET technology has allowed for lower power than had been previously achieved before in an ADC targeting 56 GBaud systems [6]. The input bandwidth and speed of the SAR ADC lanes has not seen this same improvement. Since FinFET speed performance has saturated, it is unlikely that implementation in the most advanced 7nm node would result in individual SAR lane speed or ADC bandwidth improvements. Extending the ADC bandwidth beyond 22 GHz [6] has not yet been demonstrated in a high-speed CMOS SAR ADC.

As a result of this, receivers implementing these ADC for 56 - 64 GBaud systems must operate with a Nyquist frequency above the bandwidth of the ADC itself. Higher-speed systems have been
Table 1.1: Evolution of Coherent Technology

<table>
<thead>
<tr>
<th></th>
<th>1st Gen Coherent</th>
<th>2nd Gen Coherent</th>
<th>3rd Gen Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduced</td>
<td>2010</td>
<td>2013</td>
<td>2017</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>100Gb/s</td>
<td>200Gb/s</td>
<td>400Gb/s+</td>
</tr>
<tr>
<td>Baud Rate</td>
<td>28-32 GBAud</td>
<td>32-45 GBAud</td>
<td>56-69 GBAud</td>
</tr>
<tr>
<td>Key Technology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase Modulation (QPSK)</td>
<td></td>
<td>Multi-level</td>
<td>Intelligent</td>
</tr>
<tr>
<td>Coherent Receiver</td>
<td></td>
<td>modulation (16-QAM, 8-QAM)</td>
<td>Modulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nyquist Shaping</td>
<td>(PCS, Hybrid,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD-FEC</td>
<td>up to 64-QAM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Real-time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Telemetry</td>
</tr>
<tr>
<td>Si Process</td>
<td>35nm</td>
<td>28nm</td>
<td>16nm</td>
</tr>
<tr>
<td>Time to reach 20K units</td>
<td>3.75 years (3Q13)</td>
<td>2.25 years (3Q16)</td>
<td>1.75 years* (2Q19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(* Forecast)</td>
</tr>
<tr>
<td># of DSP sources</td>
<td>8</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 1.1: Key technologies and baud rates in coherent optical systems [1].

demonstrated using SiGe HBTs to achieve ADC input bandwidths above 32 GHz targeting 64GBaud systems [7], and using InP HBTs for transmitters targeting symbol rates over 100 GBAud [8, 9]. The speed of these devices are expected to continue to scale with further technology nodes, and using HBT Current-Mode Logic (CML), the speed of mixed signal systems will improve as well. However due to larger layout footprint of CML and lack of monolithic integration with the most aggressively scaled CMOS node, BiCMOS ADC solutions have lost ground to CMOS. In this thesis high-speed SAR ADC designs using CML will be investigated. Compact CML gates need to be developed to allow for the implementation of a high-speed, time-interleaved SAR ADCs in SiGe BiCMOS technology.

1.2 Objective

The objectives of this thesis are as follows:

1. Demonstrate the advantages and limitations of 55nm SiGe BiCMOS technology in high-speed mixed-signal systems.

2. Develop a small footprint high-speed SAR ADC lane using current mode logic.

3. Measure and characterize the fabricated SAR ADC design.

4. Use the SAR lane in the development of a high-speed 64GS/s time-interleaved SAR ADC.
1.3 Thesis Organization

These thesis objectives are addressed throughout the coming chapters. They are organized as follows:

- **Chapter 2: Background**
  - Information on general ADC performance metrics, high-speed ADC architectures, SAR ADC clocking schemes, and interleaver architectures is presented.

- **Chapter 3: SAR ADC Lane**
  - The design of a SAR ADC lane in 55nm SiGe BiCMOS technology is described.
  - Novel, small footprint CML circuits are presented.
  - Circuit designs are analyzed and design choices are justified block by block.

- **Chapter 4: Experimental Results**
  - A high-speed sequencer is measured to demonstrate the small footprint CML family.
  - The SAR ADC lane is measured, characterized, and compared with other high-speed designs.

- **Chapter 5: Time-Interleaved SAR ADC**
  - An alternative synchronous SAR ADC lane is presented.
  - The design of clock and data distribution systems and circuits are analyzed.
  - Simulation results of the time-interleaved SAR ADC based on the novel SAR ADC architecture are presented and compared with the state-of-the-art.
Chapter 2

Background

2.1 ADC Performance Metrics

There are a number of important metrics and terms used to characterize ADCs. They will be used in this thesis to analyze and compare ADC topologies[10]:

- **Resolution, N, Bits**
  - Defines the number of quantization levels which the ADC can resolve. There are a total of $2^N$ quantization levels in a N-Bit ADC.

- **Sample Rate, $f_s$, Samples/s (S/s)**
  - Defines the number of samples the ADC converts per second.

- **Most Significant Bit, MSB**
  - The bit corresponding to the largest binary weighting in a N-Bit ADC.

- **Least Significant Bit, LSB**
  - The bit corresponding to the smallest binary weighting in a N-Bit ADC.

- **Integral Nonlinearity, INL, LSBs**
  - Defined as the deviation of the ADC measured output value from the ideal output value. It can be defined in terms of unit LSBs as the difference between the best fit line of the digital output codes and the measured value.

- **Differential Nonlinearity, DNL, LSBs**
  - Defined as the deviation of the difference between two adjacent quantization levels from the ideal difference. It can be defined in terms of unit LSBs as:

\[
DNL(i) = \frac{V_{out}(i + 1) - V_{out}(i)}{\text{ideal LSB step width}} - 1
\]  

where $i$ is referring to the $i^{th}$ quantization level.
• Signal-to-noise and distortion ratio, SNDR, dB

  – The ratio of the output fundamental tone power to the sum of output noise and distortion powers.

• Signal-to-noise ratio, SNR, dB

  – The ratio of the output fundamental tone power to the output noise power.

• Spurious free dynamic range, SFDR, dB

  – The ratio of the output fundamental tone power to the output power of the highest spur.

• Effective number of bits, ENOB, Bits

  – The measured performance of the ADC in bits.

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

(2.2)

2.2 High-Speed ADC Architectures

2.2.1 Flash ADCs

Historically, flash ADCs have been the standard high-speed ADC architecture [10]. Modern flash ADCs have been shown to achieve sampling rates above 100 GS/s through the use of time-interleaving [7]. A flash ADC operates by generating a reference voltage for each quantization level boundary and comparing this to the input analog signal. A basic two bit flash ADC can be seen in Figure 2.1. The flash ADC operates by first sampling an input signal using a track and hold circuit (T&H). This held voltage is then compared to each of the generated reference voltages. For an N bit ADC there are \(2^N - 1\) comparators.

The comparators with reference voltages below the held input signal output a “high” digital level. The comparators above the input voltage level output a “low” digital level. This generates a thermometer coded output which typically requires decoding to convert to a binary weighted output.

The reference voltages in a flash ADC can be generated in a number of ways. The most common is through the use of a resistor ladder which has been implemented numerous times in SiGe and CMOS flash ADCs [11, 12, 7]. However, flash ADCs that generate quantization levels through comparator offset currents [13], an on chip DAC [14, 15, 16], or off chip reference voltages [17] have been published as well.

Due to the fact that one reference voltage is needed for each quantization level boundary, the number of comparisons that must be done each conversion cycle increases exponentially with the resolution of the ADC. This means that the area and power consumption of the flash ADC increases exponentially with increased resolution. For this reason flash ADCs are typically used for resolutions of up to 8 bits.

In SiGe BiCMOS technologies the flash has been the most common architecture[13, 18, 19, 20, 7]. The reason for this is that the high-speed of the HBTs in SiGe BiCMOS processes allows for fast comparators and T&H amplifiers which are the primary components of a flash ADC. Converter topologies that rely on primarily digital logic, typically require compact, low power logic circuits which are easier to implement in CMOS, albeit, at much lower speed.
2.2.2 SAR ADCs

The SAR ADC is widely used for medium resolution applications. It was first implemented in an integrated MOS process in 1975 [21, 22]. The SAR converts a signal by performing a binary search algorithm. The algorithm is outlined in the flow chart in Figure 2.2. After sampling the input signal using the T&H circuit, the held input voltage is compared to one half of the full scale voltage of the ADC. This corresponds to the MSB. If the held voltage is higher than the MSB voltage, the MSB remains high and the next most significant bit turns on. If the held voltage is lower, the MSB turns off and the next most significant bit turns on. After this, one half of the MSB voltage is added to the current reference voltage generated by the DAC. This new reference voltage is then compared to the held input signal. This comparison cycle continues until the LSB is compared. The resulting stored bit sequence is read out and bits are reset so the next cycle can begin.

A simple block diagram of a SAR ADC can be seen in Figure 2.3. The DAC used to generate the reference voltage is typically capacitive, operating to both store the sampled voltage from the T&H and to perform the binary search algorithm. The advantage of a capacitive DAC, in addition to functioning as the hold capacitor, is that it typically does not consume static current. In fact, with the exception of a buffer or pre-amplifier which is commonly found before the comparator, all power consumption in a typical CMOS SAR ADC is dynamic. For this reason SAR ADC performance has improved drastically with technology scaling [23]. Due to this and when incorporated into time-interleaved ADCs, it is a very popular choice for medium resolution, high-speed applications [24, 25, 26, 6].

Unlike the flash ADC, in a SAR ADC, area and power consumption do not increase exponentially with increasing resolution. To increase the resolution of the ADC, the DAC resolution must be increased and the logic required for additional conversion cycles must be increased. As a result, the power consumption of a SAR ADC increases roughly linearly with increases in resolution. For this reason, in medium resolution applications, 6 - 10 bits, SAR ADCs are typically more energy efficient than other ADC topologies. In addition to this the digital outputs are binary weighted so there is no need for a decoder to convert from thermometer code as there is in a flash ADC architecture.
2.2.3 Pipelined ADCs

A pipelined ADC consists of multiple lower resolution ADCs cascaded to create one single higher resolution ADC. An example of a pipelined ADC can be seen in Figure 2.4. By splitting the conversion of the analog signal into multiple stages that run simultaneously, higher resolutions can be achieved while maintaining high sampling rates. Typically there is an amplifier between stages due to comparator noise limitations as resolutions increase. The pipelined ADC in Figure 2.4 would operate by first sampling the analog input signal using the T&H circuit to store the voltage. The ADC conversion cycles of the first ADC stage would leave a voltage residue by subtracting the converted analog signal from the held input signal. The results of the N-bit ADC conversion are stored. The amplifier stage is then activated, amplifying the residue and passing the signal on to the next stage in the pipeline. The second stage then begins its conversion. The first ADC stage samples the input signal and begins conversion of the next sample simultaneously.

This simultaneous operation of multiple ADCs in series is what allows for higher conversion rates. The latency of this converter is approximately the same as a single SAR ADC with the same resolution, the difference being that the resulting digital bits are output in multiple stages. The performance
requirements of the first stage are stringent as any noise or non-linearities added in this stage will be amplified to subsequent stages. As a result, comparator offsets must be carefully calibrated out and the gain of interstage amplifiers must be well controlled. Pipeline ADCs have been implemented with both SAR, and flash ADC stages [27, 28, 29]. When used in conjunction with time interleaving in modern CMOS technologies, they show promise as a path to sampling speeds of over 100GS/s.

2.3 SAR ADC Architectures

2.3.1 Synchronous SAR ADC

The synchronous SAR ADC operation is depicted for a 4-bit ADC in Figure 2.5. Here, the time allocation for the conversion of a single bit is determined by an external clock. Each bit is allocated the same time to resolve. This allocated conversion time must be sufficient for the DAC to settle, the comparator to make a decision, and the SAR logic to switch. The time required for the comparator to make a decision is the most variable. If the difference between the held sample and the DAC voltage is very small, the conversion time can be substantially larger than when the difference between the two voltages is large. This is due to the exponential nature of the comparator latch settling. Although these long comparisons may only happen once or twice per cycle, the time allocated for each bit to make a decision must be equal. If the conversion time for one bit ends while the comparator is metastable, a bit error may be observed. To avoid this, the time allocated for each bit conversion must be equal to the worst case conversion time, or where the risk of a metastability error becomes sufficiently statistically improbable.

The downside to this architecture is the time wasted during easier comparisons. If the difference between the held sample and the DAC voltage is large, the comparator will make a decision quickly. This leaves a time period after the decision is made where nothing is being done but waiting for the next bit comparison to start. An example of this can be seen in Figure 2.5, labeled as an easy comparison. The advantage to this architecture is the simplicity of the clock generation logic. As everything is timed off of a external clock, a synchronous SAR ADC can be implemented with little more than a shift register.
2.3.2 Asynchronous SAR ADC

Unlike a synchronous SAR ADC, an asynchronous clocking architecture only uses an external clock to define the end of one conversion cycle, and the start of another. It was first presented in [30], and has become increasingly popular in many publications in recent years [6, 31]. An example of a 4-bit SAR ADC with an asynchronous clocking architecture can be seen in Figure 2.6. Once the conversion cycle is started by the external clock all other clocking is done internally. After the comparison of a particular bit is completed, the start of the next bit comparison is triggered. This eliminates the problem with wasted time that comes with a synchronous SAR clocking architecture. The result is a variable conversion time for each bit. As can be seen in Figure 2.6, the conversion of the second bit takes longer time due to the small difference between reference voltage and held voltage. Unlike with a synchronous SAR ADC, when the easy conversions are completed, the next conversion starts earlier allowing for a faster overall conversion time.

The variable conversion time means that the next bit conversion will not start while the comparator latch is metastable. Instead it is possible that the external clock will stop the conversion cycle before all bits have been compared. Normally this will only be the least significant bit, but it is possible that an extremely long, random metastability event will occur. Known as sparkle codes, these rare large errors occur due to multiple bits being missed due to the abnormally long comparison time. Numerous techniques have been proposed to eliminate them: either detecting and stopping the metastability event [32], or a back-end correction algorithm [33].
2.4 Time-Interleaving

Time-interleaving consists of running multiple slower ADCs in parallel to create one single higher-speed ADC. Time-interleaved ADCs have become the standard in ADC based wire-line and optical receivers. Due to the high data rates required in modern standards, ADCs with sampling rates over 10 GS/s are required. To achieve these high-sampling rates with 6-10 bit resolution, time-interleaving is needed. In a time-interleaved ADC as shown in Figure 2.7, the first ADC initially samples the input voltage, followed by the second, third, fourth and finally the first again. It is critical that the T&H circuits in each lane are clocked equally out of phase with each other, otherwise timing mismatch will degrade the signal performance. In the following subsections three standard time-interleaving front-end architectures will be described.

2.4.1 Direct Interleaving Architecture

Direct interleaving is the most basic time-interleaved front end implementation. In this implementation the input signal is fed directly to each subADC without any additional circuitry between the input and the T&H circuit in each subADC lane. An example of this can be seen in Figure 2.7. This has been successfully implemented using different ADC architectures [7, 34]. The main problem with this architecture is the limitation on the number of subADCs that can be time-interleaved. As the number of subADCs is increased the total capacitive load on the input node is increased as well. This has a detrimental impact on the bandwidth. In addition, with a direct interleaving architecture the bandwidth of each subADC front end sets the bandwidth of the entire time-interleaved ADC. This makes this architecture not practical for time-interleaved ADC with a large number of subADCs. However due to its simplicity, the ADC with the highest reported input bandwidth and sampling rate made use of
2.4.2 Sub-Sampling Switch Architecture

The sub-sampling switch interleaver architecture uses T&H stages to sample the input that are separate from the subADC T&H stages. To prevent loading the input with the subADCs, an input buffer is used between the subADC T&H stage and the input T&H stage. An example of the sub-sampling switch architecture can be seen in Figure 2.8. The input T&H first samples the input signal, storing it on the sampling capacitor. The held signal is buffered and is sampled by the subADC T&Hs. It is possible to cascade multiple sub-sampler switch interleavers together as presented in [25].

This architecture has a number of advantages over the direct interleaving architecture. First of all, the clock phases for sampling in each subADC, \( (\theta_1, \theta_2, ... \theta_N) \), have relaxed jitter and timing skew requirements in comparison to \( (\phi_1, \phi_2, ... \phi_N) \). This reduces the total number of critical clocks needed, simplifying systems with a large number of subADCs. Secondly, the input capacitance is reduced in comparison with the direct interleaving architecture. Since the subADCs do not load the input directly, larger numbers of subADCs can be used while maintaining a higher input bandwidth. However, there are downsides to this architecture in comparison to the direct sampling architecture. The additional T&H and buffer stages will increase the power consumption and the noise of the system. The clocking circuitry also becomes more complicated due to the addition of the \( \phi \) clocks.

The sub-sampling switch architecture has been used in a number of time-interleaved ADC designs [4, 5, 35, 24]. The ADCs presented [4, 5] used 160 individual SAR ADC lanes to achieve sampling rates of 24GS/s and 40GS/s respectively. In order to implement timing skew and bandwidth calibration, the
SAR ADC lanes were grouped together in blocks of 10 lanes reducing the total calibrations needed. Such a large number of subADCs would not be possible with adequate bandwidth in a direct sampling architecture. Due to the lower speed operation of the SAR ADC, in order to achieve high-sampling rates, a large number of subADCs are needed, making the sub-sampling switch architecture more practical.

2.4.3 In-line Demultiplexer Architecture

The in-line demultiplexing time-interleaver architecture was first presented in [36]. Instead of sampling the input signal on a dedicated sampling capacitor, the input is demultiplexed into T&H circuits located in each subADC. An example of this can be seen in Figure 2.9. When the switch driven with clock $\phi_1$ closes the input signal passes through to one of the subADCs. One subADC per $\phi_1$ clock cycle samples the input signal on its hold capacitor. The first $\phi_1$ clock cycle $\theta_1$ samples the input signal, the next $\theta_2$, this continues until $\theta_N$ after which the process is repeated. This clocking scheme depicted in Figure 2.9, ensures that the hold capacitor of only two subADCs is loading the input at any one time. This number can be reduced to one or increased to higher numbers. The fewer on at one time improves the bandwidth, however requires clocks with lower duty cycle that can be more difficult to generate and distribute with correct timing.

One of the advantages of this architecture over the sub-sampling switch architecture, is for the same number of subADCs the window for each subADC to sample is greater, and therefore the time alloted to T&H settling is greater. In addition to this, lower power consumption can be achieved due to the absence of buffers between switches in the interleaver. The downside of in-line demultiplexing is the reduction in bandwidth. This can be from two sources. First, since there are at least two switches in series, the input resistance of this architecture can be higher. Secondly, multiple hold capacitors may be
loading the input at any one time, and the capacitance of the switches in each of the subADCs is seen by the input. As a result, the bandwidth of this architecture is typically lower than a sub-sampling switch architecture for a given number of subADCs. This makes bandwidth mismatch due to switch mismatch a potential problem and more difficult to correct for with the absence of buffers.

It is possible to cascade the in-line demultiplexer architecture with a sub-sampling switch architecture as was first presented in [37]. Here, an initial interleaver stage consisted of four 1:4 in-line demultiplexer stages. The input voltage for each of these 16 channels is sampled on a hold capacitor and buffered to provide the held signal to 4 subADCs each with its own individual sampler. The resulting time-interleaved ADC consisted of 64 total subADC lanes and achieved a large signal bandwidth of 22GHz. Further analysis of interleavers combining in-line demultiplexing and sub-sampling switches was presented in [26]. From this it can be seen that there is a direct trade off between the input bandwidth of the time-interleaved ADC and the amount of in-line demultiplexing implemented. The more demultiplexing, the lower the bandwidth but the longer the sampling window. The hold time is the time allocated to transfer the signal from the hold capacitor to the subADC. From this analysis it was concluded that in order to achieve an input bandwidth over 30GHz using sub-sampling switches, in-line demultiplexing or a combination of the two with standard CMOS switches in 32nm CMOS technology, was to use only a sub-sampling switch interleaver architecture.
Chapter 3

SAR ADC Lane

3.1 Overview

This chapter provides an overview of the design of a high-speed SAR ADC lane utilizing a low power inductorless quasi-CML logic family. The goal of this ADC lane was later incorporation in a time interleaved SAR ADC. Design choices were made with this in mind. Minimizing SAR lane footprint and logic footprint drove many of the design choices. The ADC was designed to maximize lane speed with 6-bits of resolution with a target ENOB of over 4 bits. The full scale input amplitude is designed to be \(300\, mV_{pp}\) per side. With the target specification of 4-bits ENOB it is not clear if a flash architecture with 5-bits of resolution would be a more efficient approach. The SAR is chosen over the flash for two main reasons. First, the SAR ADC will output binary weighted data at lower speeds. This makes it easier integrate with DSP that would be needed for an optical receiver. Any demultiplexing needed would be lower speed and no thermometer coded to binary weighted conversion circuity is needed. Second, since the SAR ADC conversion sequence requires more complex digital circuitry than the flash, it better investigates the advantages and limitations of mixed signal SiGe BiCMOS circuits.

In this chapter, first high-speed CML logic families are analyzed. A high speed quasi-CML logic family with active inductive peaking is chosen for use in the lane due to its high-speed operation and compact design. Design of other blocks in the SAR ADC are outlined, the clock generation scheme described, and an overview is given of circuits added for testing. A block diagram of the final SAR Lane can be found in Figure 3.1.

![Figure 3.1: SAR ADC lane block diagram.](image-url)
3.2 High-Speed Logic

In order to maximize the speed of the SAR ADC lane, a high speed logic family must be chosen. CML is a logic family that provides higher-speed operation than CMOS logic. Unlike CMOS, CML logic consumes static power by design. Even in the most advanced CMOS nodes higher speeds are possible with CML logic. This is because at higher frequencies the voltage gain is reduced making rail-to-rail logic levels difficult to achieve. CML logic operates with a reduced output swing, based on a resistive load making higher-speed operation possible. To maximize speed, transistors in CML gates operate in active region. Operating in regions such as triode reduces the maximum switching speeds. In addition to this, standard CML logic is differential. This is advantageous in systems where delays in complementary clock signals can degrade performance. At higher speed operation the dynamic power consumption of CMOS logic can surpass the static power consumption of CML. For example, high-speed non-overlapping clock generation circuits have been presented with both CMOS logic [38] and CML [39]. Here the CML implementation consumes around the same total power at about double the speed, resulting in two-times better efficiency per bit. This shows that in a very high-speed mixed signal system CML circuits can demonstrate higher efficiencies than their CMOS logic counterparts. Figure 3.2 depicts two variations of a CML inverter gate, one with a MOS differential pair and one with HBT.

3.2.1 MOS CML vs Bipolar CML

CML based logic operates with logic levels chosen to provide enough swing to fully switch the tail current from one side to the other of a differential pair. As can be seen from the CML inverter gates in Figure 3.2, the output swing is from $V_{DD}$ to $V_{DD} - I_{Tail}R_L$. The required voltage swing to completely switch a differential pair over process is dependent on the technology. For a bipolar circuit the swing required is primarily dependent on the thermal voltage. To account for process and temperature variation, a good choice for this is $250mV_{pp}$ [40]. To maximize switching of HBT CML gates devices are biased at 1.5 times peak $f_t$ current density when the current is switched to one side. For a MOS CML gate, the required swing is dependent on the current density that the differential pair is biased. It has been shown that MOS CML gate delay is minimized when operating with a current density of about 0.3mA/$\mu$m [41]. This corresponds to peak $f_t$ current density for a n-MOSFET. For a 65nm process bias at this current
density a swing of $320mV_{pp}$ is required, with newer technology nodes not improving beyond $300mV_{pp}$ [40]. In addition to looking at the required swing, when comparing a MOS CML gate to a HBT CML, gate the difference in delay is critical. For the same power consumption it has been shown that a HBT logic gate has smaller delay [42]. Delay expressions for a MOS CML gate and a HBT CML gate are as follows [41]:

$$
\tau_{HBT} = \frac{\Delta V}{I_{Tail}} \left[ \frac{C_{bc} + C_{ca} + C_{int}}{I_{Tail}} + \left( k + \frac{R_b}{R_L} \right) \frac{C_{bc} + (1 + A_v)C_{bc}}{I_{Tail}} \right]
$$

(3.1)

$$
\tau_{MOS} = \frac{\Delta V}{I_{Tail}} \left[ C_{gs} + C_{db} + C_{int} + \left( k + \frac{R_g}{R_L} \right) (C_{gs} + (1 + A_v)C_{gd}) \right]
$$

(3.2)

where $k$ is the fanout factor, $C_{int}$ is the layout interconnect capacitance, $R_g$ is the gate resistance, and $R_b$ is the base resistance. These two delay constants were analyzed in depth in [41]. It was concluded that although the output time constant was smaller for the HBT differential pair, the input time constant of the MOS pair is smaller due to the smaller gain of the MOSFET. This results in a smaller Miller capacitance. This is utilized in CML latches using a MOS-HBT cascode on the clock path as is described later. Throughout the ADC design, HBT-based CML logic was used instead of MOS CML logic for inverter cells due to the higher switching speed. HBTs were sized such that, when switched fully, the device is biased at as close to 1.5 times peak $f_t$ current density as possible to maximize speed. In CML latch gates a BiCMOS quasi-CML approach was used instead as will be explained in the following subsection.

### 3.2.2 Quasi-CML Logic

To take advantage of the small input time constant of a MOS differential pair, and small output time constant of an HBT differential pair, a MOS-HBT cascode was utilized in all CML latches. Examples of standard and quasi-CML latches can be seen in Figure 3.3. This combination was first presented in a CML gate in [41]. However, the quasi-CML topology which was presented in [43] provides similarly
Figure 3.4: CML based inverter circuits using (a) passive peaking and (b) active peaking.

high-speed operation but without a dedicated current tail. This allows for the supply to be reduced and latches to be operated using only a 1.8V supply. The primary downside to this topology is the lack of common mode rejection. To make it possible to switch off the bottom MOS pair, high-threshold voltage MOSFET devices are used. The clock driving the bottom MOSFET pair can be either AC or DC coupled. AC coupling provides a bias level that is more consistent over PVT variation. However, it requires a large capacitor and can be problematic when used with a clock that does not have 50% duty cycle. DC coupling can be achieved through the use of an emitter follower to level shift the common mode level. As an added benefit, this emitter follower further improves speed by acting as a high-bandwidth buffer as in emitter coupled logic, however it is not as robust over PVT variation.

### 3.2.3 Inductive Peaking CML Logic

In order to further improve the speed of CML logic gates, inductive peaking can be used. This works by tuning out the capacitance on the output node using an inductor. Also known as shunt peaking this added inductor improves bandwidth due to the zero it creates in the frequency response, however this also creates a second pole associated with the output node. The effect this inductor has on the load impedance of the circuit in Figure 3.4 (a) can be found to be:

$$Z_L(s) = \frac{R_L + sL_{peak}}{1 + sR_LC_{out} + s^2C_{out}L_{peak}}$$  \(3.3\)

Where \(C_{out}\) is the value of the capacitance at the collector. Using the second order load impedance expression 3.3 the gain of the CML inverter can cast in the form \([40]\):

$$A_v(s) = A_0 \frac{1 + s/\omega_z}{1 + s/Q\omega_0 + s^2\omega_0^2}$$  \(3.4\)

where:

$$A_0 = -g_mR_L, \omega_z = \frac{R_L}{L_{peak}}, Q = \frac{1}{R_L} \sqrt{\frac{L_{peak}}{C_{out}}}, \omega_0 = \frac{1}{\sqrt{L_{peak}C_{out}}}$$
For a given output capacitance, load resistor, and desired Q, the inductor can be found to be:

\[ L_{\text{peak}} = Q^2 R_L C_{\text{out}} \]  

(3.5)

For a maximally flat group delay response, \( Q^2 \) should be chosen equal to 0.32 [15]. This provides a bandwidth increase of approximately 60\%. For a given fanout of CML inverters with HBTs biased at 0.75 peak \( f_t \) current density, the product \( R_L C_{\text{out}} \) is approximately constant and independent of the chosen current tail value. So, to a first order approximation, the value of the peak inductor increases linearly with the load resistance. In addition to this, at low current tail values, process restrictions on the minimum emitter length prevent optimal biasing of the HBT pair. As a result the \( R_L C_{\text{out}} \) product increases instead of remaining constant. When designing CML blocks with low current tails (1mA and lower) in BiCMOS 55nm technology this can become problematic as the inductor value increases further which results in inductors with larger area. In a SAR ADC lane contained within a time-interleaved ADC, minimizing the area is critical, so using digital blocks with large inductors is not practical. A solution to this problem is the use of an active inductor to provide the peaking, as shown in Figure 3.4 (b).

The use of a HBT as an active inductor in this way was first presented in [7]. However, CMOS active inductors have been demonstrated earlier as an alternative to on chip inductors in CMOS CML gates [7]. Intuitively as the frequency of the signal at the output increases, more and more of the \( R_{\text{peak}} \) resistor is seen in series with the \( R_L \) resistor due to the coupling through the base-emitter capacitance of the HBT. This response of increased impedance at higher frequencies is similar to the effect of an inductor. The impedance looking into the emitter of the active inductor can be found to be:

\[
Z_{\text{peak}} \approx \frac{(r_o \| r_e) + (r_o \| r_e)(C_{\text{be}} + C_{\text{cb}})R_{\text{peak}} s}{1 + (R_{\text{peak}} C_{\text{cb}} + (r_o \| r_e) \frac{R_{\text{peak}}}{r_o} C_{\text{be}}) s + (r_o \| r_e) R_{\text{peak}} C_{\text{be}}^2 s^2}
\]  

(3.6)

This can be roughly modeled by the RLC network in Figure 3.5. The values of the passive components can be found to be approximately:

\[
L = (r_o \| r_e)(C_{\text{be}} + C_{\text{cb}})R_{\text{peak}} \approx \frac{R_{\text{peak}}}{\omega_T}
\]  

(3.7)

\[
R_p = r_o \| r_e \approx r_e = \frac{1}{g_m}
\]  

(3.8)

\[
C_p = \frac{C_{\text{be}}^2}{C_{\text{be}} + C_{\text{cb}}} = \omega_T \frac{C_{\text{be}}^2}{g_m}
\]  

(3.9)
It can be seen here that for a given HBT size and current bias, the value of the inductance provided increases linearly with respect to $R_{\text{peak}}$. Unlike with a passive inductor where the area needed to provide the optimal inductance value increases substantially when reducing the current tail of CML gates, the resistor area does not. Although the required value of the $R_{\text{peak}}$ increases, since the current flowing through it to the base of the HBT is very small, increases in resistance can be realized by reducing the width of the resistor. The result of this is that when reducing the current on a CML gate with this active inductor topology the total layout area decreases. This makes it ideal for use in the SAR ADC lane.

The primary difference in frequency response of the circuit is the additional pole created by the parasitic capacitance of the HBT. Using open-circuit time constant analysis the pole can be found to be at:

$$\omega_{\text{parasitic}} \approx \frac{1}{R_{\text{peak}} \ || \ R_L \ \frac{C_{be}^2}{C_{cb}}} \quad (3.10)$$

The pole associated with the output of the CML gate should be dominant, however to further push this pole to higher frequencies, the parasitic capacitance should be minimized. This is done by sizing the peaking HBT for 1.5 time peak $f_t$ current density when fully switched. So long as this is the case, the circuit can still be cast in form of equation 3.4 where:

$$A_0 = -g_m (R_L + r_o \ || \ r_e) \quad (3.11)$$

$$\omega_z = \frac{R_L + r_o \ || \ r_e}{(r_o \ || \ r_e)R_{\text{peak}}(C_{be} + C_{cb})} \quad (3.12)$$

$$Q = \frac{1}{R_L + r_o \ || \ r_e} \sqrt{\frac{(r_o \ || \ r_e)R_{\text{peak}}(C_{be} + C_{cb})}{C_{\text{out}}}} \quad (3.13)$$

$$\omega_0 = \frac{1}{\sqrt{(r_o \ || \ r_e)R_{\text{peak}}(C_{be} + C_{cb})C_{\text{out}}}} \quad (3.14)$$

There are some downsides to using this active inductor instead of a passive one. The first is linearity. The load impedance due to the active inductor was derived using small signal approximations. In reality, the use of this active inductor will be far more non-linear than a passive one. This is not a problem in a CML gate as the behavior is inherently non-linear, however it could be problematic if it were used in a circuit providing peaking to an analog signal. The second problem is noise. Noise from the HBT and the peaking resistor contributes to added output noise. As a result, it can cause a problem if used on a clock path that is sensitive to jitter. An example of this would be the sampling clock in a master track and hold amplifier as used in a time-interleaved ADC. The third problem is the required headroom to accommodate the extra HBT in the transistor stack. To implement the same logic functions, a quasi-CML logic family with a passive inductor that requires a 1.8V supply would require a 2.5V supply with an active inductor. This increases the power consumption for a given logic gate by 39%.
3.2.4 Set and Reset Latch Design

One of the problems with CML design is as the number of gate inputs increases, the number of levels of transistors stacked also increases. For two input gates such as latches and AND gates, the use of the quasi-CML topology allows for supply voltages to be maintained by replacing the tail with a switching pair. This solution does not scale to CML gates with 3 or more inputs such as set or reset latches. The functionality provided by these gates have been implemented in numerous ways. The additional inputs can be added in series with the CML logic stack, or by cascading two CML gates to provide the functionality [44]. Both of these options are not ideal. The first requires an increase in supply for each transistor in the stack. For only reset or set functionality, implementation with HBT switches and an active inductor requires a supply of 3.3V. This can be seen in Figure 3.6a. If set and reset functionality were implemented a supply of over 4V would be required. In addition, these would need additional emitter follower stages to provide level shifting for the 3 or 4 levels. This further increases the power consumption. Set and reset can be implemented with MOS switches reducing the required supply to 3.3V with active peaking, however this would still have multiple levels making bias points challenging to set correctly. The BiCMOS SR latch implementation is depicted in Figure 3.6b.

The second option is to implement the set and reset by cascading CML gates as seen in Figure 3.6c [44]. This allows the same supply to be used, however the power consumption still increases due to the additional gates. In addition, the total number of gates cascaded increases as the number of required inputs increase. If set and reset functionality were to be both implemented, three cascaded CML gates would be required. A problem with this approach is the increase in delay. Since the data must pass through additional stages for each additional input, for a 4 input gate the delay will be 3 times that of a single stage.

Due to the increases in supply or increases in delay neither of these multi-input latch solutions would work well in the SAR register within the ADC lane. A compact alternative is used instead, which does not require an increase in supply, however requires set and reset functionality to be implemented in a single ended fashion. An example of a synchronous SR latch as used in the SAR register is depicted in Figure 3.7a. Resistors are added between outputs and the base of Q3,4 to increase the reset and set speed. The value of $R_{pd}$ was found to maximize set and reset speed when it causes about a 200mV drop when reset or set is on. An asynchronous alternative reset latch is presented in Figure 3.7b. This is used along with a set version in the sequencer circuit used to generate clocks. As the sequencer reset does not need to be as fast as in the SAR register no $R_{pd}$ resistors are added. This allows the layout to be
more compact.

3.3 Track and Hold Amplifier

The THA utilized here was first presented in [45]. The design consists of three stages, a linear input buffer, a switched emitter follower based track-and-hold which uses a quasi CML MOS-HBT switch, and a output buffer. The switched emitter follower topology, first presented in [46] allows for very high speed operation. A block diagram of the THA can be found in Figure 3.8. The design of the blocks are approached with the following important design considerations:

- Minimum SNDR higher than the resolution of the ADC Lane.
- Fast settling time to maximize switching speed.
- Reduce power once other specifications are met.

As the ADC lane provides a maximum 6-bits of resolution, the THA must provide sufficient performance such that it is not limited by this stage. For this reason it is designed for a SNDR of over...
44dB. Secondly the speed of the THA amplifier must be sufficient such that the held voltage settles within $1/2 \times f_{\text{clk}}$. Finally once these previous two specifications are met the power of the THA should be minimized. The design of each individual block in the THA is outlined in the following subsections. A circuit schematic of the entire THA can be found in Figure 3.9. The total power consumption of these circuits is 17mW.

### 3.3.1 Input Buffer

The design of input buffer is critical to the overall performance of the THA. In designing this stage a number of factors need to be taken into account:

- High linearity is required. Any distortion added here can directly affect the accuracy of the held voltage.
- Bandwidth must be sufficiently larger than the maximum input signal frequency.
- Sufficient voltage headroom is needed to allow for the emitter follower switching of the following stage.

The final point listed above provides limitations on the number of buffer topologies to be used. In order to provide proper biasing for the switched emitter follower following this stage, a 2.5V supply must be used. During track mode, current is pulled from the load resistors in the buffer to achieve enough voltage drop to turn the emitter follower in the next stage off. This eliminates the use of a HBT-HBT cascode topology as it would cause the cascode device to enter saturation region during hold mode, substantially reducing the maximum sampling speed of the THA. This leaves three possible topologies: an emitter degenerated differential pair, a MOS-MOS cascode differential pair, or a MOS differential pair. Neither the MOS differential pair, or the MOS-MOS cascode provide the bandwidth required by
Chapter 3. SAR ADC Lane

Figure 3.10: Schematic of THA operating in track mode.

this THA. A more complete analysis of different input buffer topologies can be found in [47]. For this reason an emitter degenerated differential pair is chosen.

In order to maximize linearity and bandwidth, the buffer is chosen to have a gain of 1. The tail current is chosen in order to bias the differential pair at as close to peak $I_f$ current density as possible. The voltage gain of an emitter degenerated differential pair is approximated by:

$$A_v = \frac{g_m R_L}{1 + \frac{r_E}{r_E}}$$

where $r_E \approx \frac{1}{g_m}$ is the small signal parasitic emitter resistance and $g_m = \frac{I_C}{V_T}$. Knowing that the desired gain of the buffer is 1, and that the desired full scale voltage for the ADC is 300mVpp per side the maximum peak-to-peak voltage for linear operation can be referred to the input of the amplifier as shown in [48] as:

$$\text{MaxSwing} = \text{MIN} \left( \frac{I_{\text{Tail}} R_L}{A_v}, \frac{2(V_{CEQ1,2} - V_{CE-SAT})}{A_v}, 2V_T(1 + g_m R_E) \right)$$

The first condition is to ensure that the output swing at the full scale voltage does not clip. The second ensures that the input pair does not enter saturation. The third gives the range over which the linear small signal approximation is valid.

3.3.2 Switched Emitter Follower

The switched emitter follower design used here provides a high bandwidth with sufficient linearity [45]. It does not utilize a switching output buffer as in [46] which would require higher power consumption. A lower supply voltage is possible in comparison to the design presented in [48]. As the requirements of the switch when tracking the input signal and when holding the sampled voltage differ, the operation of this circuit in both these regimes will be analyzed separately.
During track mode the circuit operates as shown in Figure 3.10. The two main requirements of the circuit in track mode are sufficient linearity, and the ability of the circuit to slew the hold capacitor voltage to the full scale voltage. The primary cause of distortion in this stage is due to the voltage swing on the input of the emitter follower. As a result of this the third order harmonic can be found to be [46]:

$$HD_3 = \frac{V_T (2\pi f_{in} A_{CH})^3}{12 I_{Tail_2}^2 (A_{Tail_2} + V_T 2\pi f_{in} A_{CH})}$$ (3.17)

where $f_{in}$ is the input frequency and $A$ is the amplitude of the input signal. As can be seen with this topology, there are two options to improve linearity. The first is to reduce the size of the hold capacitor value. The second is to increase the current through the switch.

The other major consideration when the circuit is in track mode is the slew rate (SR) of the hold capacitor. This must be large enough so that the switch can track the input signal at its maximum amplitude and maximum speed. The slew rate is as follows:

$$SR = \frac{I_{Tail_2}}{C_H}$$ (3.18)

As can be seen, both increasing tail current and reducing the value of the hold capacitor increase both linearity and the slew rate of the switch. However, in addition to increasing power consumption in order to maintain linearity, Q3,4 must be increased in size. This will increase loading of the previous stage reducing the bandwidth of the THA.

During hold mode the circuit operates as shown in Figure 3.11. Here the MOS-HBT switch directs current through the load resistor of the input buffer lowering the voltage at the base of Q3,4 until these devices turn off. The major design considerations in hold mode include:

- The rate at which the held voltage degrades due to leakage currents or droop rate (DR).
- The feed-through of the input signal to the hold capacitor.
- The clock kick-back from stages following the THA.
- The voltage error caused during switching, known as the sampling pedestal.

The sampling pedestal is caused by the non-zero switching time of the MOS-HBT switch. It should theoretically be canceled in a differential circuit [46]. However, perfect cancellation cannot be achieved due to large signal non-linearities and mismatch. Methods to alleviate it would be to reduce the size of $I_{\text{Tail2}}$ of the switch or increase the size of $C_H$.

The droop rate is primarily due to the sub-threshold current of the MOSFET in the MOS-HBT switch. It can be equated to be:

$$DR \approx \frac{I_{D0} \frac{W_{M2,3}}{L_{M2,3}} \frac{V^2_G}{kT} V_{\text{hold}}}{C_H}$$  \hspace{1cm} (3.19)$$

Where $I_{D0}$ is the sub-threshold characteristic current constant, $n$ is the sub-threshold slope factor, and $V_G$ is the gate source voltage on the MOSFETs in hold mode. To reduce it, $I_{\text{Tail2}}$ can be reduced, $C_H$ can be increased, or the amplitude of the clock driving the switch can be increased to reduce the value of $V_{GS}$.

To reduce the effect of clock kick-back, cascaded buffer stages should be used between the switch and the clocked circuits in the ADC lane. The output buffer and emitter follower following the switch as seen in Figure 3.1 was found to be sufficient.

There can be substantial input signal feed-through from the previous stage through the base emitter capacitance of Q1,2. This can be canceled out completely with the addition of the feed forward capacitance $C_{ff}$. This capacitance value should be approximately equal to $C_{bcQ9,10}$. The impact this capacitor has on bandwidth can be significant. In order to improve bandwidth at the cost of a loss in linearity, $C_{ff}$ can be sized for less then perfect cancellation.

As can be seen, the track mode and the hold mode put different requirements on $I_{\text{Tail2}}$ and $C_H$. As a result, the design of the quasi-CML switch and hold capacitor must be done iteratively. The optimal value for $C_H$ was found to be 40 fF, after parasitic extraction $C_{ff}$ was found to be 3 fF. M1,2,3,4 were sized to provide 1.8 mA of current with Q5,6,7,8 sized to have close to 1.5 times the peak $f_t$ current density when fully turned on.

### 3.3.3 Output Buffer

The output buffer design has similar requirements to the input buffer. High bandwidth and linearity are both desired, however unlike the input buffer, any current drawn at the input from the previous MOS-HBT switch stage directly increases the DR of the held voltage. As a result the base current that would result from a HBT input pair make it impractical when a relatively low duty cycle clock is used in the THA as is the case in the SAR ADC lane. As a result, three topologies were considered: a MOSFET differential pair, a MOS-MOS cascode, and a MOS-HBT cascode. The first two allow operation from a 1.8V supply. However, the bandwidth of the MOS-HBT cascode is superior [41] to other cascode topologies due to the reduced input time constant associated with the input MOSFET Miller Capacitor, and the output time constant associated with the HBT. The overall time constant can be approximated to be [40]:

$$\tau_{\text{MOS-HBT}} \approx \Delta V \left( \frac{C_{bcQ9,10} + C_{cgQ9,10}}{I_{\text{Tail3}}} \right) + \left( k + \frac{R_{gM5,6}}{R_L} \right) \Delta V \left( \frac{C_{gsM5,6} + C_{gdM5,6}}{I_{\text{Tail3}}} \right)$$  \hspace{1cm} (3.20)$$
where $\Delta V$ is the maximum swing of the stage, $k$ is the fanout factor. In order to maximize bandwidth and linearity, the gain of the stage is designed to be 1, and inductive peaking is used to improve bandwidth. It is critical that the bandwidth of this stage is large enough to allow for the held voltage output to settle before the first comparison in the SAR ADC conversion algorithm. The tail current value was chosen to be 1 mA which allows for adequate bandwidth while minimizing power consumption. Transistors sizes were chosen to maximize the bandwidth of the stage.

3.4 R2R DAC Design

Although capacitive based DACs are popular in recent CMOS high-speed SAR ADC designs [26, 28, 6, 24], to achieve a fast settling time capacitor values must be so small that parasitic capacitances will dominate and models become unreliable. In addition a capacitive DAC requires a reliable voltage reference further complicating the overall ADC design. Even with the unit size capacitor being small, the layout of a binary weighted capacitive DAC can be large. Capacitor area increases by a factor of $2^N$ as the number of bits increases and dummy capacitors are added to improve matching. The R2R DAC provides a compact alternative that has been successfully implemented at high speeds using HBT technology [49, 50, 51]. The R2R DAC concept is illustrated in Figure 3.12.

The operation of the R2R DAC can be understood using basic circuit analysis. It can be seen in Figure 3.12 when the MSB switch is turned on and all others are turned off, the impedance at the output is equal to $R$ ignoring the switch impedance. As a result the output voltage will be equal to $V_{DD} - IR$ which defines the voltage of the MSB. If however, the next most significant bit is switched on and all others are off the resistive ladder behaves as a resistive divider and the resulting output voltage will be equal to $V_{DD} - IR_2$. This trend continues as switches further down the ladder are switched on resulting in the binary nature of the DAC.

By using SiGe HBT based CML switches and SiGe HBT current tails, very high speed operation can be achieved that would not be possible in a pure CMOS technology. The accuracy of this design depends on how significant the mismatch between resistors and between current tails are. As the 2R resistors in the DAC can be implemented by using two unit sized resistors in series, the DAC design can be made more robust against resistor mismatch. By using HBT degenerated current tails, more accurate currents can be generated than possible with simple CMOS, or cascode CMOS current tails. There are a number
of design choices to consider when designing this DAC:

- The full scale of the DAC must be chosen to match the full scale of the ADC.
- The settling time of the DAC must be fast enough for operation in the ADC Lane.
- The layout area must be compact enough to minimize the ADC Lane size.

Based on this it can be shown that the most critical time constant associated with the MSB switching can be approximated to be:

$$\tau_{MSB} \approx R \left( C_{bc} + C_{cs} + C_L \right) + R_{in} \left[ 1 + \frac{I_{tail}}{2V_T} R \right] \left( C_{bc} + C_{be} \right)$$  \hspace{1cm} (3.21)

where $C_L$ is the input capacitance of the next stage, and $R_{in}$ is its input resistance. The choice of the resistors and value of current tail used can be related to the full scale voltage as:

$$[2 + 2^{-(N)}] I_{tail} R = V_{FS}$$  \hspace{1cm} (3.22)

It can be seen here that for a given full scale voltage, the product of the current and resistance is fixed. This means that the only way to increase the speed of the DAC itself is to increase the power consumption by increasing the current tail and decreasing the resistance value.

Given the desired full scale voltage of $300mV_{pp}$, the current tail was chosen to be $0.5mA$ setting the resistance to be $298\Omega$. The HBT switches were sized to achieve as close to 1.5 times peak $f_t$ current density as possible.

### 3.5 High-Speed Comparator

The high-speed comparator consists of two stages: the comparison stage, and the latch stage as seen in Figure 3.14. The comparison stage takes a differential input from the THA, the reference input generated
from the R2R DAC, and amplifies the difference. It provides this function by using two different HBT input pairs with separate but well matched current tails. In order to provide superior matching, HBT current tails were used instead of MOS based current tails. One pair amplifies the difference of the positive THA and DAC signals, the other the negative signals. The tail currents are steered by the operation of these two pairs and are summed on the load resistors.

In addition to the comparison function performed by the comparator block, it also functions as a preamplifier before the latch stage. There are a number of constraints on the gain provided by this stage. The first is the delay of the comparator block. The delay must be small enough such that it does not limit the speed of the ADC by increasing the time needed around the SAR loop. Unfortunately, increasing the gain corresponds directly with a decrease in speed. The only ways to provide an increase in gain are to either increase the value of the current tail, the load resistor, or add another amplification stage, the latter two both increase the delay around the SAR loop. Another problem with increasing the resistive load is the output swing. When the two current tails are fully switched, a total of $2I_{\text{tail1}}$ flows through the resistive load. The voltage drop when this occurs must not push the input HBTs into saturation mode. It was found that a gain of 5.5 dB provided the optimal trade-off between the delay and gain of the comparator stage. In addition the comparator block also provides isolation from kickback due to the clocking latch in the following stage.

Following the comparator stage is a quasi-CML latch with active peaking. During the tracking phase of the latch, the clock is low and the latch functions like a standard CML inverter stage. This provides about 5 dB of amplification. During latching mode the clock is high and positive feedback is employed
Chapter 3. SAR ADC Lane

Figure 3.15: Sequencer circuit for SAR ADC lane clock generation.

to provide regenerative gain. As the latch mode time constant is a function of the transconductance of the latch and the capacitance at the output node, \( \frac{C}{g_m} \), for the fixed output load from the SR latches, increasing the bias current while maintaining current density only increases capacitance of the CML latch itself. This increases the maximum speed of the latch at the cost of extra power consumption. The value of the current tail in this case was chosen to be 2 mA. To maximize speed the HBTs are sized to fully switch to 1.5 peak \( f_t \) current density, and the MOSFETs in the tail are sized to provide the correct current when fully switched. The quasi-CML clock pair, \( M_{1,2} \), is driven by an emitter follower that allows for a DC coupled clock with the correct bias levels.

### 3.6 Sequencer

The timing diagram of the SAR ADC lane is shown in Figure 3.15. One clock cycle is allocated for sampling the input and resetting the circuit, one clock cycle for each bit comparison and another to store the digital output. Since the comparator uses a CML latch as opposed to a dynamic latch it is not edge triggered. As a result the phase shifted clocks (P-signals) setting and clocking the SR latches must
operate out of phase with the comparator clocks. This allows the combination of comparator latch and SR latch to operate like a flip-flop. For each comparison, one of the P-signals sets the SR latch output high which turns the corresponding bit on in the R2R DAC. This P-signal then turns off and the change in DAC value has half a clock cycle to pass through the comparator and the comparator latch which will be tracking the comparator signal. When the comparator latch transitions to latching mode, the next P-signal turns on, clocking the same SR latch and storing the resulting bit. The same P-signal which clocks the SR latch sets the next most significant bit and the cycle continues. This is outlined in the clock diagram in Figure 3.15.

In order to generate the clock phases, a sequencer must be used as shown in Figure 3.16. This consists of a shift register made of quasi-CML flip-flops in order to generate the 8 phases needed for the SAR sequence. Once these phase shifted pulses are generated, they pass to an AND gate whose other input is the master clock with a small fixed delay. This results in phase shifted signals with half the duty cycle of those in the shift register. For a 40GHz master clock, the circuit in Figure 3.16 generates 5GHz clock signals with 6.25% duty cycle. Although the smaller duty cycle makes the clocks more challenging to distribute due to the short pulse width, this is needed for proper operation of the SAR lane. By looking at the clock sequence in Figure 3.15, it can be seen why. Since the comparator requires a high clock to hold its output value, the P-signals must only sample during the hold period. It should be noted that all blocks shown in Figure 3.16 use active peaking and operate off a 2.5V supply. This leads to a 40mW power consumption to generate the signals themselves, with an additional 20mW for the clock buffers.

Clock amplifiers are used to boost an external single ended clock and distribute differential clocks to the flip-flops and AND gates in the sequencer. The block diagram of the full chip can be seen in Figure 3.18. The clocks and delayed clocks which drive the sequencer are AC coupled in order to reduce the power consumption that would come from emitter followers for level shifting. As area is not a limiting factor for the clock amplifiers, CML stages with passive inductors for peaking are used with a 1.8V supply.
3.7 SAR ADC Lane Design Summary

The complete SAR lane block diagram can be seen in Figure 3.17. An emitter follower is added between the THA buffer and the comparator block. In addition to providing the correct DC bias levels to the comparator, it further isolates the THA from the comparator, preventing kick back from the latch and R2R DAC. The latter causes errors in the hold signal reducing the performance of the ADC. Additional emitter followers were placed on the THA clock path and within the SR latches for DC coupled level shifting.

The block diagram of the full chip can be seen in Figure 3.18. The input to the SAR ADC is biased with a resistive divider with an impedance, as seen from the input of 50 Ω allowing for inputs to be AC coupled. The resistive divider consumes 3mA from a 1.8V supply. The clock path input is terminated on 50 Ω on chip. The clock signal passes through a number of CML clock buffers forking into five paths to distribute the clock to the sequencer and comparator as can be seen in Figure 3.18. Four different clock paths are distributed to the sequencer where they are AC coupled. Two provide clocks for the shift register. The other two have an additional clock buffer to provide a fixed delay and drive the AND gates in the sequencer. The comparator clock path branches off early on in the clock tree. This passes through emitter follower coupled CML buffers operating off a 2.5V supply. The current through the emitter followers can be adjusted to provide a variable delay.

Once the ADC conversion cycle is complete the resulting output bits are stored in a register. The output on this register are used as inputs for a DAC. The DAC employs the same R2R architecture that is used in the lane itself. The signal from this DAC is then passed to a 50 Ω output buffer. The schematic of the 50 Ω output buffer can be seen in Figure 3.19. A common emitter topology with emitter degeneration is used. It is designed to have a gain of 1 and sufficient linearity in order to not degrade the output signal. For correct operation of the output buffer a bias tee is used to provide a DC impedance of 50 Ω from a 1.8V supply and 50Ω AC termination off chip. The result is a 25 Ω load impedance. The current tail used is 16 mA, and the transistor pair is sized for maximum linearity.
Figure 3.18: Block diagram of complete chip.

Figure 3.19: Output buffer used to read out DAC value.
Chapter 4

Experimental Results

In the following chapter the experimental results of two circuits are outlined. The first is the sequencer circuit, which generates eight 6.25% clock signals. This circuit allows for testing of the inductorless quasi-CML logic family, and the novel set and reset latches. The second circuit tested is the SAR ADC lane. This incorporates the sequencer for clock generation, and demonstrates the use of the logic family in a small footprint SAR ADC. All circuits outlined here were fabricated in STMicroelectronics’ 55nnm SiGe BiCMOS process. This features 9 metal layers and a few different flavors of MOSFETS and HBTs. The highest speed HBTs, which are the primary flavor of HBT used here, have a minimum 100nm emitter width, minimum 350nm emitter length, and a 330/[f_T]/f_MAX [52].

4.1 Sequencer

4.1.1 Test Setup

The die micrograph of the sequencer can be seen in Figure 4.1. The die occupies an area of 750\(\mu\)m\(\times\)870\(\mu\)m. The sequencer core itself, which here is considered to be the shift register and AND gates required for clock generation, occupies an area of 50\(\mu\)m\(\times\)105\(\mu\)m. The setup used to test this circuit is depicted in Figure 4.2. Power supplies, and I/O signals were interfaced with the circuit using 100\(\mu\)m pitch Picoprobes from GGB Industries. The probes used on the signal paths have a 67GHz bandwidth.

The sequencer was measured with input clock frequencies ranging from 3 GHz to 42 GHz. The clock was generated using an Agilent E8257D analog signal generator (ASG). This clock signal is AC coupled through the use of an Anritsu SC7287 bias tee, which has 67 GHz of bandwidth. This provided a bias of 1.8V. A reset pulse was generated using a Keysight M9502A arbitrary waveform generator (AWG). This too was AC coupled using an Anritsu SC7287 bias tee, which provided a bias of 1.6V. In order to reset the sequencer, a 160kHz signal with a 0.5% duty cycle was applied. Although the speed of the reset was not critical, the rise and fall times were set to the minimum allowed by the AWG. The output was measured using a Keysight URX1002A real time oscilloscope (RTOS). This has 110 GHz of input bandwidth and a sampling rate of 256 GS/s.
Figure 4.1: Sequencer die micrograph.

Figure 4.2: Measurement set up for the sequencer.
4.1.2 Measurement Results

The output of the sequencer using a 42GHz input clock is demonstrated in Figure 4.3. As can be seen from the information on this figure, the measured frequency is about 5.25 GHz, the period about 190 ps, and the duty cycle is 6.07%. The Fourier transform of the 42GHz measured pulse can be seen in Figure 4.4. One thing to note is the apparent amplitude modulation of the measured output pulse. Since it is at the limit of the operation speed of the sequencer, the output is only at the level of a logical high for a very short period of time. The oscilloscope samples as a result do not always capture the peak of the pulse. Due to this, small variations in measured amplitude between pulses are observed.

The measured pulse which is generated using a 5GHz input clock signal is displayed in Figure 4.5. As can be seen here the measured amplitude is higher due to less cable and probe loss comparison to the 42GHz clock signal. The frequency of the pulse generated by this signal is 625 MHz. It can be seen here that the amplitude modulation is not observed. On the lower end, the minimum input frequency that the sequencer still operates at is 3 GHz. The reason for this is the on chip DC blocking capacitors on the clock path. As a result the total output frequency range is 375 MHz to 5.25 GHz. The measured duty cycle at these frequencies can be seen in Figure 4.6.

4.1.3 Performance Summary and Comparison

The sequencer operates with up to a 42 GHz input clock to generate eight 6.25% duty cycle clock signals. The power consumption of the sequencer core by itself, including the flip-flops in the shift register and
Figure 4.4: Frequency domain measurement of single output phase using a 42GHz input clock. The attenuation of tones at higher frequencies is partly due to cable and probe loss which has not been calibrated out.

Figure 4.5: Measurement of single output phase using a 5GHz input clock.
Chapter 4. Experimental Results

![Graph showing duty cycle vs. input clock frequency]

Figure 4.6: Measurement of duty cycle vs. input clock frequency.

Table 4.1: Comparison with state of the art high-speed clock generation circuits

<table>
<thead>
<tr>
<th>Clock Generation</th>
<th>This work</th>
<th>[39]</th>
<th>[38]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>55nm SiGe BiCMOS</td>
<td>55nm SiGe BiCMOS</td>
<td>22nm FDSOI CMOS</td>
</tr>
<tr>
<td>Architecture</td>
<td>ANDing multiphase clock</td>
<td>ANDing multiphase clock</td>
<td>ANDing multiphase clock</td>
</tr>
<tr>
<td>Logic</td>
<td>quasi-CML active peaking</td>
<td>CML passive peaking</td>
<td>CMOS</td>
</tr>
<tr>
<td>Bandwidth (GHz)</td>
<td>5.25</td>
<td>62</td>
<td>32</td>
</tr>
<tr>
<td>Duty Cycle (%)</td>
<td>6.25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Number of Phases</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Power (mW)</td>
<td>40</td>
<td>176</td>
<td>200</td>
</tr>
<tr>
<td>Efficiency (fJ/bit)</td>
<td>59.5</td>
<td>177</td>
<td>391</td>
</tr>
</tbody>
</table>

The AND gates, is 40 mW. Two other high-speed multiphase clock generation circuits using ANDing to reduce the duty cycle have been published [39, 38]. The first generates 25% duty cycle I and Q signals using standard CML logic with passive peaking in 55nm SiGe BiCMOS. The second generates the same, instead using 22nm FDSOI CMOS technology and CMOS logic. Although these both operate with higher speed input clocks the power consumption is far higher. A comparison can be seen in Table 4.1. The efficiency of each of these circuits is calculated using the following equation:

\[ Efficiency = \frac{Power \times DutyCycle}{\#ofPhases \times Bandwidth} \]  

(4.1)

here bandwidth is the maximum output frequency of the clock generation circuit.
4.2 SAR ADC Lane

4.2.1 Test Setup

The die micrograph of the SAR ADC can be seen in Figure 4.7. The die occupies an area of 1.07mm × 0.985mm. As can be seen from the micrograph, the area is limited by the pads, not the by circuits themselves. The SAR ADC lane not including the sequencer or the output DAC and buffer, occupies an area of 200µm × 40µm. The setup used to test this circuit is depicted in Figure 4.8. Power supplies, and signals were interfaced with the circuit using 100µm pitch Picoprobes from GGB industries. The probes used on the signal path have a 67GHz bandwidth.

The ADC was measured with sampling rates ranging from 2.5 GS/s to 3.5 GS/s. The required input clock frequencies required to generate these sampling rates are from 20 to 28 GHz. The clock was generated using the ASG. This clock signal is AC coupled through the use of a Picosecond 5342 bias tee, which has 50GHz bandwidth. This provided a bias of 1.8V. A reset pulse was generated using the AWG. This too was AC coupled using a Picosecond 5342 bias tee, which provided a bias of 1.6V. In order to reset the sequencer, a 160kHz signal with a 0.5% duty cycle was applied. The output was measured differentially using the RTOS. In order to bias the output properly, Anritsu SC7287 bias tees were used to set the bias point of the output buffer to be 1.6 V. The differential input signal was generated using the AWG. The input signals were AC coupled to the chip using SHF DCB67A DC blocking capacitors, which have 67GHz bandwidth.

In order to compensate the cable and probe losses offset-short calibration was performed using a Keysight N5227A PNA Microwave Network Analyzer. To account for the inherent sinc function due to the output DAC, all measurements shown here have the effect of the sinc function corrected for. No other corrections were made for DAC and output buffer non-linearity. A photo of the test setup can be seen in Figure 4.9. The RTOS in the photo is displaying the differential output from the circuit using a 311 MHz input tone and a 3.5 GS/s sample rate.
Figure 4.8: Measurement set up for the SAR ADC.

Figure 4.9: Photo of measurement set up for the SAR ADC.


4.2.2 Measurement Results

There are a number of measurement results on the following pages. When observing these measurements a few things should be taken into account. The minimum frequency measured is 101 MHz. The reason for this is the high-pass nature of the output bias tee. Below 100 MHz signal attenuation is observed, this causes substantial decrease in measured performance which is not indicative of the ADC performance. For this reason linearity measurements of frequencies below 101 MHz are not presented here.

Measurements of three different sampling rates are presented here: 2.5 GS/s, 3.0 GS/s and 3.5 GS/s. The best performance was observed at 2.5 GS/s. Further decreases in conversion rate only resulted in marginal improvements in performance so no lower sampling rates are presented. The first plot displays the measured SFDR at different sampling rates and input frequencies. This can be seen in Figure 4.10. As can be seen here, the measured SFDR is found to be above 34 dB for the entire first Nyquist band at a sample rate of 2.5 GS/s, above 31 dB at a sample rate of 3 GS/s, and above 28 dB at a sample rate of 3.5 GS/s.

Figure 4.11, and Figure 4.12 depict the measured SNDR and ENOB respectively. As can be seen, the SNDR is above 28 dB (4.3-bits ENOB) at a sample rate of 2.5 GS/s, it is above 25 dB (3.8-bits ENOB) at a sample rate of 3 GS/s, and above 19 dB (2.9-bits ENOB) at 3.5 GS/s. The measured output signal spectra can be seen for 821 MHz input tones in: Figure 4.13 for 2.5 GS/s, Figure 4.14 for 3.0 GS/s, and Figure 4.15 for 3.5 GS/s.

Figure 4.10: Measurement of SFDR vs. input clock frequency for different sampling rates.
Figure 4.11: Measurement of SNDR vs. input clock frequency for different sampling rates.

Figure 4.12: Measurement of ENOB vs. input clock frequency for different sampling rates.
Figure 4.13: Measured output spectrum at 2.5GS/s sampling rate and a 821MHz input tone. The first 11 harmonics are marked, the resolution bandwidth is 500 kHz.

Figure 4.14: Measured output spectrum at 3GS/s sampling rate and a 821MHz input tone. The first 11 harmonics are marked, the resolution bandwidth is 500 kHz.
The large signal input bandwidth of the SAR ADC was found by applying input tones of increasing frequency, and measuring the resulting down-converted tone amplitude. The measured output amplitude at increasing input frequency is presented in Figure 4.16. From here it can be seen that the input bandwidth is about 17.5 GHz. The ADC performance in higher Nyquist bands was also measured. This can be seen in Figure 4.17. This plot displays the measured SFDR and SNDR for higher order odd Nyquist bands at 2.5 GS/s and 3.0 GS/s. At a sample rate of 2.5 GS/s, the SFDR is above 34 dB and the SNDR is above 26 dB within the input bandwidth. At a sample rate of 3.0 GS/s the SFDR is above 31 dB and the SNDR is above 23 dB within the input bandwidth.

Additional plots presenting the measured SNDR vs input amplitude at 2.5 GS/s, INL, and DNL can be found in Figure 4.18, Figure 4.19, and Figure 4.20. Due to the attenuation of low frequency signals at the output due to the DC blocking of the bias tee, INL and DNL could not be measured by a single slow input voltage ramp. Instead a higher speed input ramp was applied to the ADC. The period of the ramp was chosen such that it was non periodic within a large number of conversion cycles. The measured levels of voltage steps from the output DAC are then found. All these measured voltages are plotted and clear divisions are found between quantization levels. The measured output voltage levels associated with each quantization level are then grouped and averaged. These are used to find the presented INL and DNL measurements. It should be noted that these INL and DNL measurements include the non-linearity of the ADC, DAC, and output buffer. As a result the measured INL and DNL are worse than that of the ADC itself.
Figure 4.16: Measured output amplitude normalized to full scale amplitude vs. input frequency.

Figure 4.17: Measured output amplitude, SFDR, and SNDR at 2.5 GS/s and 3.0 GS/s, for tones downconverted to 311 MHz in odd Nyquist bands.
Figure 4.18: Measured SNDR vs input amplitude normalized to the 600mV differential full scale using a 311 MHz tone.

Figure 4.19: Measured integral non-linearity.
4.2.3 Performance Summary and Comparison

There are a few measurement results that need further explanation. The first being the measured SFDR. As can be seen there is a substantial drop in SFDR in the first Nyquist band as input frequencies approach the Nyquist frequency. For example in Figure 4.10 the measured value drops from over 38 dB to only over 31 dB for 3.0 GS/s. In the higher Nyquist bands the measured SFDR values are above 31 dB for 3.0 GS/s up to the large signal bandwidth. The reason for this is the subthreshold current through the MOS-HBT switch in the THA. There is some signal feed-through due to this small bias current even when the THA is in hold mode. This is different from the feed-through due to the base-emitter capacitance which is canceled using a feed-forward capacitor. The reason that the effect of this feed-through decreases at higher frequencies is the low THA bandwidth under subthreshold bias conditions. The effects were observed in simulation as well.

There are some areas in which the ADC performed worse than expected. The first is large signal input bandwidth. This was measured to be 17.5 GHz. The expected bandwidth is over 32 GHz. Although the exact cause of this degradation cannot be conclusively determined, it is most likely due to the improper biasing of the THA. Due to insufficient metal to carry the current in the chip layout, the clock buffer driving the THA will see a lower supply voltage than 2.5V. This is due to the voltage drop on the surrounding metals. This lower supply reduces the on current of the THA MOS-HBT switch. Although this will not have a substantial impact on the low frequency gain of the circuit, since it is biasing an emitter follower, the bandwidth of the THA can be reduced substantially. In order to prevent this problem, layout should be improved to prevent significant voltage drop.

The ADC performance at higher sampling rates was also worse than expected. Once again, although the cause can not be conclusively determined, but experimental results show metastability errors at higher sampling rates. An example of this is shown in Figure 4.21. The cause of this is either insufficient time allocated for comparator regeneration, insufficient delay on the comparator clock path, or a combi-
nation of the two. It was found that the delay control only provided slightly over 1 ps of variable delay on the comparator clock path. When measuring the ADC at a sample rate of 3.5 GS/s the performance was maximized when the delay on the comparator clock path was set to the minimum allowed. This would indicate that further decreases in delay would have further increased performance. The reason for this is when not timed right, the SR latches in the SAR logic sample the comparator voltage before the total allocated time for regeneration has passed. The resulting sampled comparator voltage would be smaller than it should, increasing the likelihood of a metastability error. The other possible cause is that the allocated regeneration time was insufficient. The latter can be solved by decreasing the clock rate, which was done here, or using a clocking scheme that allows for more comparator regeneration time.

The power consumption of the SAR ADC lane is 70mW. This includes all blocks in the SAR ADC lane shown in Figure 3.1, as well as the clock buffer used for the comparator clock. The sequencer used for the lane, and clock buffers used for the sequencer consume 60 mW. The output DAC, 50Ω buffer, and input clock amplifiers consume an additional 102 mW. The lane performance is compared with other, both asynchronous [53, 54], and synchronous [24], SAR lanes in Table 4.2. The power and FOM of [24] are approximated for a single lane based on the time-interleaved performance and power and information provided in the publication. The Walden figure-of-merit (FOM) can calculated using:

$$FOM = \frac{Power}{f_s \times 2^{ENOB}}$$ (4.2)

This gives values of 1326 fJ/conversion, and 1675 fJ/conversion for 2.5 GS/s and 3.0 GS/s, respectively. Looking at Table 4.2, it can be seen that this is orders of magnitude higher than high-speed SAR lanes in CMOS. This is due to the static power consumption of CML, and the higher supply voltage required for HBT devices. With that being said, the SAR lane presented here demonstrates the highest sampling rate in a single SAR ADC lane to date in any technology. Once the problems discussed earlier are corrected, it shows promise for use in high-speed time-interleaved SAR ADCs with higher bandwidth than has been achieved in CMOS only technology.
Table 4.2: Comparison with state of the art SAR ADCs

<table>
<thead>
<tr>
<th>SAR ADCs</th>
<th>This work</th>
<th>[54]</th>
<th>[53]</th>
<th>[24]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>55nm SiGe BiCMOS</td>
<td>28nm CMOS</td>
<td>32nm SOI CMOS</td>
<td>28nm FDSOI</td>
</tr>
<tr>
<td>Architecture</td>
<td>Sync. SAR</td>
<td>Async. SAR</td>
<td>Async. SAR</td>
<td>Sync. SAR</td>
</tr>
<tr>
<td>Resolution (bit)</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Sampling (Rate GS/s)</td>
<td>3/2.5</td>
<td>2</td>
<td>1.2</td>
<td>1.25</td>
</tr>
<tr>
<td>ENOB @f_{Nyquist}</td>
<td>3.8/4.4</td>
<td>4.2</td>
<td>6.2</td>
<td>5.3</td>
</tr>
<tr>
<td>Power (mW)</td>
<td>70</td>
<td>2.12</td>
<td>3.06</td>
<td>2.5*</td>
</tr>
<tr>
<td>FOM (fJ/conversion)</td>
<td>1675/1326</td>
<td>55.4</td>
<td>34</td>
<td>51*</td>
</tr>
</tbody>
</table>

*Lane power and FOM was approximated based on information provided.
Chapter 5

Time-Interleaved SAR ADC

In previous sections the design of a small footprint SAR ADC lane was demonstrated. This lane was designed to be eventually included in a larger time interleaved SAR ADC. In this section, the design of the complete 16× time-interleaved SAR ADC is discussed. Performance limitations demonstrated through measurements in Chapter 4 are addressed, with the circuits being modified accordingly. First, the time-interleaved architecture used is discussed in Section 5.1. This is followed by the development of a novel synchronous clocking architecture to improve lane performance, which can be found in Section 5.2. After this, clock distribution methods are discussed in Section 5.3 and input signal distribution 5.4. Finally, the performance is analyzed and compared to state-of-the-art in Section 5.5.

5.1 System Overview

The block diagram of the overall time-interleaved architecture is shown in Figure 5.1. It should be noted that, with the exception of the input clock signal, input reset signal, and output data signals, all signals shown in this block diagram are differential. The ADC has 6-bits of resolution, with a target ENOB of over 4.0-bits, and a maximum input frequency of up to 32 GHz. This would require a minimum sample rate of 64 GS/s. The full scale input voltage is designed to be 600 mVpp differential. To maximize bandwidth and minimize clocking complexity a sub-sampling switch interleaver architecture is used. Two master track and hold amplifiers (MTHAs) sample the signal first, each providing a held signal to slave track and hold amplifiers (STHAs) in eight SAR ADC lanes for an overall total of 16 SAR ADC lanes. Eight SAR ADC lanes are used because 8 clock cycles are needed for 6-bit SAR conversion. One phase is needed for each bit comparison, and the other two are used for sampling and resetting each SAR lane.

The ADC is grouped into two smaller ADC blocks, “ADC A” and “ADC B”. To achieve a total sampling rate of at least 64 GS/s, ADC A and ADC B each operate with complementary 32 GHz clocks. Each MTHA operates at 32 GS/s using a 50% duty cycle clock, M-clock, provided by the “clock amplifiers” block. The clock amplifiers provide 32 GHz clock signals to the “clock generation” block. This block generates the comparator clocks, and the phase shifted clocks (P-signals) used for the SAR conversion. P-signals A and B are 90° out of phase with each other. The SAR subADCs use the same blocks outlined in Chapter 3, with an architecture modification that will be outlined in section 5.2. To achieve a total sampling rate of 64 GS/s, each lane must operate at 4 GS/s. All of the ADC output
Figure 5.1: The time-interleaved ADC block diagram. With the exception of the input clock, input reset, and output data line, all other connections shown here are differential.
data is converted from CML to CMOS signal levels and stored in memory on chip.

5.2 SubADC Architecture

The metastability errors identified in Chapter 4 are a result of insufficient time allowed for the comparator latch voltage regeneration. Although this may be due to clock timing issues which are addressed in Section 5.3, in this section a lane architecture modification is presented to substantially increase regeneration time of the comparator latch without reducing the overall conversion speed of the ADC lane. This lane architecture change and the required clock timing is depicted in Figure 5.2. The circuits shown in the SAR ADC lane are identical to those in the design presented in Chapter 3. All circuits use a 2.5V supply with the exception of the R2R DAC, which uses 1.8V. The difference is the two separate comparator latches operating with complementary clocks. It should be noted that each of the two latches use a 1mA tail current as opposed to 2mA, which was the case in the previous design. This results in the same fan-out for each latch and for the comparator. As a result, the overall lane power consumption remains 70mW, identical to that of the previous design.

An alternate comparator SAR ADC architecture has been presented before for use in an asynchronous clocking sequence [53]. This alternate architecture improves speed by eliminating the delay due to the time required to reset the comparator after each comparison. Once one comparison has completed, its decision triggers the other to begin the next comparison. This eliminates any down time during the reset phase. This is different from the advantages of the proposed two comparator synchronous architecture that is presented here. In this case, since a dynamic comparator is not used, there is no need for a reset phase between each comparison. The SAR lane performance is improved upon in two different ways. The first is the ease of clock distribution since the comparator clock is half the frequency (16GHz instead of 32GHz), and the P-signals are twice the duty cycle (12.5% instead of 6.25%). The second is the time allocated for comparator latch regeneration is doubled.

The SAR ADC conversion cycle starts with the P(0) signal that resets the SR latches and samples the input signal. The P(1) signal then sets the MSB to high. During this time, the comparator clock is high and latch 1 samples the comparator output for the MSB comparison. When the P(2) signal turns on, the comparator clock goes low putting latch 1 in regeneration mode. At the same time the MSB is clocked and the next bit is set high. Simultaneously latch 2 samples the comparator voltage. When P(3) turns on, latch 2 is sampled by the second bit in the SAR register. At the same time the third bit is set and sampled by latch 1. This process continues until the SAR algorithm has been completed.

The halving of comparator clock frequency and the doubling of the P-signal duty cycle simplifies the clock distribution. Although the low duty cycle and higher speed clock is not problematic when distributing the clocks to a single lane, it becomes much more difficult when clocks must be distributed to more. The inclusion of a sequencer in each lane would require substantially more area and power consumption, and distribution to eight lanes from one sequencer is much more manageable. In this case two sequencers are used: one for ADC A, referred to as sequencer A, and one as ADC B, referred to as sequencer B. The clock distribution will be discussed in more depth in a later section.

The regeneration time for each comparison using this clocking scheme doubles. This change addresses one of the possible causes of the identified metastability errors in the SAR ADC lane measurements. Since the voltage regeneration is exponentially dependent on time, and since the time constant of the CML latch in latching mode does not change, the probability of a metastability event occurring decreases
Figure 5.2: The proposed modified SAR ADC lane, and corresponding clocking diagram. The R2R DAC uses a 1.8V supply, all other blocks shown here use a 2.5V supply.

5.3 Clock Generation and Distribution

As previously discussed, a single sequencer for each lane is not practical due to area requirements and power consumption. The approach that is used is to have one sequencer per 8 subADC lanes. As a result, the P-signals must be distributed over the 8 lanes, each having a width of 40 µm, with P-signals shifting one position over for each individual lane. The proposed interconnect layout is depicted in Figure 5.3. Each of the clock phases must pass through eight total interconnect segments. The seven shorter segments have a length of 45 µm. One long segment is 115 µm long and must over or under pass line segments. This is depicted as the lighter gray line in Figure 5.3. The exception to this is the interconnect line for the P<7> signal, where all eight segments have a length of 45 µm.
Figure 5.3: Proposed clock distribution to synchronous SAR subADCs.
As a result of this required interconnect, two challenges arise. The first is the P-signal delay between lanes, and other P-signal phases caused by the 430µm long interconnect. The delay in the first P-signal in each subADC conversion sequence, during which the STHA samples the input, has the most impact on performance. This corresponds to P(0) in the first lane, P(1) in the second, and P(8) in the last. Although the input to each lane is sampling an already sampled signal, to maximize linearity the subADC must sample this signal when the held data is most settled. This is ensured in a few ways. First, the input data delay is matched with the sampling clock delay. Second, a variable delay is added to control the clock phase sent to the clock generation block with respect to the M-clock. Each of these points are discussed later in this section. Third, the STHA must sample the held data at the most settled point. This is achieved by ensuring that no P-signals pass through the 115µm length interconnect segment before being used to sample the STHA. This way all the P-signals used to sample the STHAs pass through only identical 45µm long segments. This can be seen in Figure 5.3. Although some of the other P-signal phases may experience unequal delay due to the 115µm segment, this just results in small variations in the timing difference between the comparator clock and the P-signals. This change in relative timing is minimized by increasing the comparator clock interconnect segment length to match the P-signal delay as best as possible.

The second challenge is the difficulty of driving 430µm long lines. This difficulty is diminished with to the doubling of the P-signal duty cycles with the new SAR clocking architecture, however it still poses a significant challenge. Due to layout constraints, interconnect between different P-signals are tightly packed causing substantial sidewall capacitance between differential lines, and between adjacent lines.
In order to drive the interconnect lines using standard CML clock buffers with active peaking, large current tails are needed. Since the sequencers are in the center of the ADC layout itself, these large currents are a challenge to provide a path for without substantial IR drop. In order to both reduce the power consumption of the clock buffer used and these parasitic voltage drops, the final clock buffer uses a distributed load with far-end termination as depicted in Figure 5.4. Three clock buffers are placed between the sequencer output and the ADC lanes. The first two use active peaking to provide a compact layout that is needed in this area of the chip. The final stage uses a 6mA tail, an 80Ω far-end termination, and line segments with 80Ω characteristic impedance. The load it is driving in each lane is approximately equal with the exception of when the P-signal is used for STHA sampling. In this case it drives 2mA clock buffer placed immediately before the 115µm segment. Although the distributed load segments are not perfectly matched, it was found that impedances were matched close enough to allow for distribution of the 4GHz, 12.5% duty cycle P-signals. At the far end of the ADC lanes the area constraints are not as critical. This allows for passive inductive peaking to be used here with a 1.8V supply. When compared to active peaking with a 2.5V supply, this reduces the power by about 67mW over the 16 total distributed clock buffers. The same three clock buffers are used to distribute the comparator clock.

The block diagram for the clock amplifiers can be seen in Figure 5.5. This block performs a number of functions. It takes the off chip clock single-extendedly and distributes it to the sequencer and comparator clock generation circuits. It re-times the off chip reset signal, distributes it to sequencer A, and distributes M-clock to sample the MTHAs. Finally, it has variable delay cells with independent timing control of the clock being delivered to ADC A, and ADC B. The clock buffers and flip-flops utilize passive inductive peaking. All circuits in this block operate with a 1.8V supply and consume a total power of 185 mW.

The DC currents in each block are labeled on the figure. The clock delay cells consume 4 mA each, and the resistive divider used to bias the input buffer and match it to 50 Ω, consumes 3 mA.

The input clock signal is provided single-extendedly off chip. The resistive divider provides a 1.65V
DC level, while also ensuring that the input impedance is 50 Ω. From here, M-clocks are provided to the MTHA for sampling, and clocks are provided to the sequencer and comparator clock generation blocks. Before sampling the MTHA, the M-clocks pass through two CML buffers and are AC coupled to the MTHA. To minimize added jitter, as few clock buffers as possible were used on the M-clock clock path. Any jitter on these clock signals degrades the linearity of the MTHAs. The sequencer clocks first pass through the variable delay cells. More detailed circuit circuit schematics can be seen in Figure 5.6. After passing through the two delay blocks, the sequencer clocks are used to re-time the input reset signal. This is done through the use of two flip-flops. Each side of the clock distribution tree clocks one of the flip-flops to provide equal loading for both sides. The flip-flops each consist of two 1mA quasi-CML latches. Emitter followers consuming 1mA of current, are used for level shifting. After this, four CML clock buffers are used to distribute the sequencer clocks to sequencers A and B. The output of the reset path is taken single extendedly and the output clocks to the sequencers are AC coupled so no DC level shifting is needed.

The variable delay cell is depicted in Figure 5.6. It consists of a current summer with a clock input and a delayed clock input. By adjusting the control voltage, the current can be steered from the clock input to the delayed clock input controlling the weight of each input. The total delay this block provides is 6.5ps. Two cascaded delay cells are used in the clock amplifier block to provide a total of 13 ps of variable delay. The maximum delay range equals the difference between the clock signal and the delayed clock signal. The current summer itself, including the mirror current, consumes a total of 2.4mA at any one time. The delay is provided by two 1mA buffers. To increase delay, no peaking inductors are added to these two circuits. It should be noted that only one current mirror is used for both of the two delay cells. As a result, two current mirrors are used, one for sequencer A clocks, and one for sequencer B clocks, each with separate control voltages.
Figure 5.7: Circuits used for ADC Lane clock generation. All circuits shown here are implemented with active peaking, and use a 2.5V supply.
A block diagram of the sequencer and comparator clock generation block can be seen in Figure 5.7. This block takes the clocks (32GHz 50% duty cycle), and reset signal provided by the clock amplifier block, and generates the P-signals (4GHz 12.5% duty cycle) needed to operate the SAR ADC lanes. The P-signals of the A sequencer are generated using a shift register with set/reset functionality. The reset signal is used to set the register sequence. Once the sequencer has been set it will repeat on its own due to the feedback of the final flip-flop. This final flip-flop also provides the input signal for the B side sequencer. An extra latch is added to ensure correct phase difference between the two shift registers. In order to drive the interconnect and circuit loading, a 1mA CML clock buffer is used at the output of the final flip-flop. Since the final flip-flop in sequencer A provides the signal for the input of sequencer B, no reset or set signals are needed for this sequencer. This prevents the two sequencers from running out of sync with each other.

For the SAR subADC lanes, the comparator clock must be half the frequency of the M-clock (16GHz instead of 32GHz). This is achieved using a static frequency divider. The frequency divider consists of two 1mA quasi-CML latches with active peaking. To ensure that the comparator clock phase is aligned with that of the P-signals, the same reset clock used to set the state of the sequencer is used to set the initial state of the divider. All the comparator clock generation circuits can also been seen in Figure 5.7. The output signal of the divider is distributed to a latch on the “B” side that is followed by a flip-flop. This results in two comparator clocks with “B” 90° out of phase with respect to “A”. An additional flip-flop is added to the output of the divider on the “A” comparator clock path to match the delays between the two sides. Delay blocks follow the flip-flops on the comparator clock path. Both operate off of the same control voltage. Unlike the delay circuit depicted in Figure 5.6, only one fixed delay buffer is used, consuming 0.5 mA and providing 4ps of delay. The current summer in this case consumes only 1 mA. Both these blocks use active peaking and a 2.5V supply like all other circuits shown in Figure 5.7.

5.4 Linear Sampled Signal Distribution

The distribution of data signals using interleavers is a challenge in high-speed, highly-parallelized ADCs. State-of-the-art time-interleaved SAR ADCs targeting sampling rates of 56GS/s communication standards have not yet demonstrated bandwidths greater than the 22GHz, not even in advanced 14nm FinFET technology [6]. The distribution of data to highly-parallel ADCs is an active area of research, with recent publications demonstrating over 32GHz bandwidth in a 55nm SiGe BiCMOS flash ADC [7], and 60GHz of in bandwidth in 22nm FDSOI using non-overlapping quadrature clocks [38]. In this section, the approach for clock and data distribution for the 16× interleaved SAR ADC is discussed. Design choices are made to ensure an input bandwidth of at least 32 GHz, and a sampling rate high enough to digitally convert this bandwidth.

The MTHA and input signal distribution scheme is depicted in Figure 5.8. The linear input buffer uses a differential common emitter topology with resistive degeneration. The same topology is described and analyzed in detail in Chapter 3. It has a total tail current of 6 mA. To sample the input signal, a switched emitter follower with MOS-HBT switch is used. Once again this is the same topology as used in the subADC lanes. The MTHAs are designed to have a 12mA current consumption. Unlike with the STHAs where the P-signals are DC coupled through emitter followers, the M-clock is AC coupled to sample the MTHA. This is possible due to the 50% duty cycle clock used to sample the input signals in the MTHA. Although the DC blocking capacitors have a larger footprint than emitter followers, they
Figure 5.8: Schematic of half of the time-interleaved front end.
save power. Since the DC level at the MOS-HBT switch in this case is set using a current mirror, it is also more robust over process variation. The M-clocks used to sample the input signal are provided by the clock amplifier block.

There are a number of design constraints on the linear signal distribution amplifier, which is depicted in Figure 5.8. First, it must provide the sampled input signal to eight subADC lanes with maximum linearity. Second, it must provide a DC bias level of about 1.65V to the input buffers in the STHAs. The linearity depends on a few parameters: the non-linearity of the buffer topology used, how settled the input signal is when sampled by the STHAs, and the gain and bandwidth mismatch between ADC lanes. Since the output DC voltage must be around 1.65 V, either a MOSFET based amplifier can be used, or HBT based amplifier can be used in conjunction with a level shifting stage. Since the addition of a level shifting stage for each subADC lane would substantially increase power consumption, a simple MOS stage is used with a gain of 1. This was found to provide adequate linearity and bandwidth.

In order for the subADCs to sample the signal when the linear signal distribution amplifier is most settled, the delay of the sampled signal to each lane should be matched to the delay of the STHA sampling clocks to each lane. This is achieved by using transmission line segments with the same 45µm length as the clock line segments. There is a problem however with providing identical 80Ω characteristic impedance segments as in the clock distribution lines. Although this would match the delay better than the 42Ω lines used, the IR drop along the interconnect lines themselves would cause gain and bandwidth mismatch between lanes. The 42Ω lines with matched load were chosen because they were found to minimize non-linearity caused by both delay mismatch and gain/bandwidth mismatch. An option to eliminate the systematic gain and bandwidth mismatch would be to use a separate load for each subADC. The problem with the latter approach is that it would cause a more significant mismatch due to variations in the resistors values used in each separate load. The total power consumption of the two MTHAs is 161 mW.

5.5 Layout and Post Layout Simulation Results

The layout of the time-interleaved SAR ADC can be seen in Figure 5.9. This depicts the ADC lanes, clock generation circuits, clock amplifier circuits, and the digital interface circuits. The digital interface circuits include CML to CMOS conversion for each of the 6-bits at 2Gb/s in each ADC lane, and clocks generated for the CMOS logic blocks following the ADC. The layout of the entire chip is not shown here as access to it was not available. The ADC lanes, from input to CMOS output, are 300µm by 40µm each. The eight lanes in each ADC half occupy an area of 300 µm by 320 µm. With the addition of the two sequencers this becomes 300µm by 1050µm. Including the clock amplifiers and MTHAs, the total ADC footprint is 520µm by 1050µm. The ADC was fabricated in the same STMicroelectronics 55nm SiGe BiCMOS process, like the ADC lane. Power supplies of 2.5, 1.8, and 1.1 are used for the chip. Power supplies of 2.5, 1.8 are used for the ADC core, and 1.1 is used for the digital CMOS circuits. Following the ADC interface, circuits designed by Finisar Corp. are used to demultiplex the 2 Gb/s data from each lane and store it in the on-chip memory.

The ADC power breakdown can be seen in Table 5.1. Each ADC lane consumes a total of 70 mW of power. The CML to CMOS converters, buffers needed to drive them, and the digital circuits used to generate 2.0GHz, CMOS level, quadrature clocks consume a total of 286 mW. Due to the static current consumption of the CML circuits, power consumption is significant. The clock amplifiers
Figure 5.9: Layout of the time-interleaved SAR ADC.
Table 5.1: Breakdown of time-interleaved ADC Power Consumption

<table>
<thead>
<tr>
<th>Circuit Block</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTHAs</td>
<td>161</td>
</tr>
<tr>
<td>Clock Amplifiers</td>
<td>185</td>
</tr>
<tr>
<td>Clock Generation</td>
<td>316</td>
</tr>
<tr>
<td>ADC Core</td>
<td>1125</td>
</tr>
<tr>
<td>Output Data Interface</td>
<td>286</td>
</tr>
<tr>
<td>Total Chip Power</td>
<td>2073</td>
</tr>
</tbody>
</table>

Consume 185 mW, while the clock generation circuits consume 316 mW. This leads to a total chip power consumption of 2.07 W.

All of the simulation results presented here are obtained after parasitic RC extraction, in the typical corner and with a ambient temperature of 65°C. The linearity results are presented for a sampling rate of 72 GS/s in Figure 5.10, and a sampling rate of 64 GS/s in Figure 5.11, which require input clocks of 32, and 36 GHz respectively. In each figure, the SFDR and SNDR of the reconstructed signal after the MTHA stages, STHA stages, and the entire ADC are shown. As can be seen, the front-end nonlinearity limits the ADC performance. For input signal frequencies below 10 GHz, the ADC is limited by the STHAs, and above 10GHz, the performance is limited by the MTHA. It can be seen that at both sampling rates, the SNDR of the MTHA is only slightly above 30 dB at higher input frequencies. Since the MTHA signal is reconstructed before the output buffer, and since the input buffer provides over 44dB SNDR, it can be concluded that the linearity is primarily limited by the switched emitter follower and the quality of the clock signals used to sample it.

The simulated ENOB for the ADC at 64 GS/s, and 72 GS/s can be seen in Figure 5.12. At 72GS/s, 4-bits ENOB was simulated up to an input frequency of 17GHz. At 64GS/s, the ENOB of above 4-bits was simulated up to a 19 GHz input signal. At 64GS/s, 3.6-bits ENOB was simulated up to a 31GHz input signal. At 72GS/s, over 3.3-bits ENOB was simulated up to 36GHz signal. The large signal bandwidth of the ADC was also simulated to be about 33 GHz. The simulated output amplitude referenced to the full scale input can seen in Figure 5.13. This was found to be primarily limited by the MTHA.

The ADC performance is compared to that of ADCs with sampling rates of 40GS/s and above in Table 5.2. It is compared with SiGe designs, all of which employ flash architecture, and CMOS designs which are all SAR-type. To date, there are no published SiGe SAR ADCs with sampling rates above 1GS/s. The power consumption used to calculate the FOM for each of the SiGe flash ADCs excludes DAC and/or clock amplifier power consumption. The ADC power consumption presented in [3] includes an entire channel in a 4-channel receiver, although what blocks this includes are not stated in the publication. In the FOM calculation, for this work the output digital interface and clock amplifiers are not included. In order to properly compare the $2\times$ oversampling ADCs with the others, the FOM used was calculated with:

$$FOM = \frac{Power}{2 \times f_{in} \times 2^{ENOB}}$$ (5.1)
Figure 5.10: Simulated ADC performance at 72 GS/s. The SFDR and SNDR are presented after the MTHAs, STHAs and after conversion.

Figure 5.11: Simulated ADC performance at 64 GS/s. The SFDR and SNDR are presented after the MTHAs, STHAs and after conversion.
Figure 5.12: Simulated ADC ENOB at 64 GS/s and 72 GS/s.

Figure 5.13: Simulated full scale output amplitude for varying input frequencies at 72 GS/s.
Table 5.2: Comparison with state-of-the-art high-speed ADCs

<table>
<thead>
<tr>
<th>ADCs</th>
<th>This work</th>
<th>[55]</th>
<th>[20]</th>
<th>[7]</th>
<th>[3]</th>
<th>[6]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology</strong></td>
<td>55nm SiGe BiCMOS</td>
<td>180nm SiGe BiCMOS</td>
<td>130nm SiGe BiCMOS</td>
<td>55nm SiGe BiCMOS</td>
<td>20nm CMOS</td>
<td>14nm CMOS</td>
</tr>
<tr>
<td><strong>Architecture</strong></td>
<td>16× TI SAR</td>
<td>2× TI Flash</td>
<td>Flash</td>
<td>2× TI Flash</td>
<td>128× TI SAR</td>
<td>64× TI SAR</td>
</tr>
<tr>
<td><strong>Resolution (bit)</strong></td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td><strong>Sampling (Rate GS/s)</strong></td>
<td>72—64</td>
<td>50</td>
<td>40</td>
<td>128</td>
<td>64</td>
<td>72</td>
</tr>
<tr>
<td><strong>Input Bandwidth (GHz)</strong></td>
<td>33</td>
<td>20</td>
<td>Information Not Available</td>
<td>32</td>
<td>Information Not Available</td>
<td>21</td>
</tr>
<tr>
<td><strong>ENOB @f_{in} (GHz)</strong></td>
<td>4.3—4.4@1</td>
<td>4.0@1</td>
<td>3.7@1</td>
<td>4.6@1</td>
<td>6.3†@1</td>
<td>6.2@1</td>
</tr>
<tr>
<td></td>
<td>4.0—4.1@19</td>
<td>3.7@12</td>
<td>3.0@12</td>
<td>4.4@21</td>
<td>@1</td>
<td>4.8@36</td>
</tr>
<tr>
<td></td>
<td>3.3—3.6@35—@31</td>
<td>3.4@22</td>
<td>2.8@20</td>
<td>4.0@32</td>
<td>@19</td>
<td>@19</td>
</tr>
<tr>
<td><em><em>ERBW</em> (GHz)</em>*</td>
<td>23—23</td>
<td>18</td>
<td>12</td>
<td>32</td>
<td>12†</td>
<td>14</td>
</tr>
<tr>
<td><strong>Power (W)</strong></td>
<td>1.57†/2.07</td>
<td>5.4</td>
<td>2.3</td>
<td>1.25*</td>
<td>0.950†</td>
<td>0.235</td>
</tr>
<tr>
<td><strong>FOM (mW/bit/GHz)</strong></td>
<td>2.28—2.08</td>
<td>11.65</td>
<td>8.25</td>
<td>1.22</td>
<td>0.73</td>
<td>0.12</td>
</tr>
</tbody>
</table>

*Effective resolution bandwidth, the input frequency where the SNDR drops 3dB below its low frequency value †Power without output data interface, and clock amplifiers *Performance presented includes entire receiver channel ‡Power consumption does not include multiplexers needed for 128Gb/s output data, and CML to CMOS conversion
Chapter 6

Conclusion

6.1 Summary

The analysis and design of a SAR ADC utilizing novel small-footprint CML gates was presented. The novel set and reset latches were demonstrated by the measurements of a high-speed sequencer circuit. This circuit generates eight 6.25% duty cycle clock signals up to an output frequency of 5.25 GHz. This requires an input clock of 42 GHz and consumes 40mW of power, demonstrating an efficiency of 59.5 fJ/bit. This sequencer is used to clock a 6-bit SAR ADC lane which relies on active peaking CML logic, and novel set/reset latches. The SAR lane achieves 3.8-bits ENOB in the first Nyquist band at 3 GS/s, about 4.4-bits ENOB in the first Nyquist band at 2.5 GS/s, consumes a total power of 70mW, and demonstrates an input bandwidth greater than 17 GHz. The SFDR in the first Nyquist band was measured to be over 34 dB at 2.5 GS/s, over 31 dB at 3.0 GS/s, and over 28 dB at 3.5 GS/s. Measurements of higher frequency bands demonstrated over 26dB SNDR and over 34dB SFDR at 2.5 GS/s within the input bandwidth. For a 3.0GS/s sampling rate, the SNDR was measured to be over 23 dB and SFDR over 31 dB within the input bandwidth. This results in a Walden FOM of 1.326 pJ/conversion for the first Nyquist band. This demonstrates a higher sampling rate for a single SAR ADC lane than has been previously demonstrated based on the authors knowledge.

Using the already designed SAR ADC, a novel synchronous lane was developed to improve SAR lane speed. This is incorporated in a time-interleaved SAR ADC designed to operate at a sampling rate of 64-72 GS/s. Based on simulation results, this 6-bit time-interleaved SAR ADC demonstrates about 33GHz of input bandwidth, higher than any previously published SAR ADCs. The simulated results show over 4-bits ENOB operation at all sampling rates up to a 19GHz input frequency. At 72GS/s, over 3.3-bits ENOB was simulated up to a 35GHz input frequency. At 64GS/s, over 3.6-bits ENOB was simulated up to a 31GHz input frequency. This resulted in a FOM of 2.28 mW/bit/GHz at 72 GS/s and 2.08 mW/bit/GHz at 64 GS/s.

6.2 Future Work

It is certainly possible to create interleavers and ADCs targeting 100+ GBaud symbol rates with aggressive clocking in BiCMOS. A flash ADC was demonstrated with a 128GS/s sampling rate in 55nm SiGe BiCMOS [7]. The time-interleaved SAR could be scaled to reach this sampling rate with a total of 32
SAR ADC lanes and a 25% duty cycle quadrature clock generator [39]. This could result in a sampling rate of 128-144 GS/s. A different interleaver topology would likely need to be used to provide sufficient input bandwidth. Using this non-overlapping quadrature clock approach, a ADC frontend with over 60GHz of bandwidth has been demonstrated [38]. An alternative to improve speed, and potentially performance, would be to pipeline the SAR and create two or more conversion steps. This has been shown to increase the performance and speed of a single SAR lane in CMOS [28] and a similar architecture could be adopted for a multi-step SAR ADC in BiCMOS. The HBTs in BiCMOS would be well suited for implementing well-controlled high-performance amplifiers as required by this topology.

It is not clear if the baud rate will continue to increase with the development of future fourth-generation coherent optical communication systems. The reason for this is limitations in CMOS high-speed performance, and the power consumption of BiCMOS technology. Although performance improvement of HBTs in BiCMOS technology is likely with further scaling, the power consumption in comparison with advanced CMOS processes means that ADCs and DACs using current BiCMOS technology may not be economically viable. However, it may be the only way to realize the required bandwidth. Although BiCMOS with monolithically integrated advanced FDSOI technology has been proposed [56], it is not commercially available. In addition to this it is likely that the most advanced FDSOI technology would need to be incorporated in order to be competitive with ADCs fabricated in the most advanced 14nm, [6] and 7nm FinFET technologies.


