Millimeter-Wave Analog to Digital Converters: Technology Challenges and Architectures

by

Shahriar Shahramian

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2011 by Shahriar Shahramian
Abstract

Millimeter-Wave Analog to Digital Converters: Technology Challenges and Architectures

Shahriar Shahramian
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto
2011

While data converters have been around for nearly ninety years, mm-wave data converters are still in their infancy. Only recently the 40-GHz sampling barrier was broken with the introduction of the next generation high-speed sampling oscilloscopes. Meanwhile, data communication is the main driving force behind mm-wave data converter development. As with any mm-wave circuit, designers must go beyond simply relying on technology advancement to achieve acceptable performance. Careful device and passive modeling is critical and systematic design methodology may offer repeatable and scalable mm-wave designs.

In this thesis the design methodology and architectural challenges of mm-wave ADCs are explored. Some of the fundamental mm-wave ADC building blocks such as track and hold amplifiers, data distribution networks and flip-flops are implemented in SiGe BiCMOS and CMOS technologies and characterized. Several record breaking circuits are
presented along with systematic design methodology. The impact of these circuit blocks on the performance of the next generation ADCs is studied and experimentally verified using a 35-GS/s, 4-bit ADC-DAC chain implemented in a SiGe BiCMOS technology.
Dedication

To my father and brother.
Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. First and foremost, I’d like to thank my supervisors Professor Tony Chan Carusone and Professor Sorin P. Voinigescu. Throughout the six years of my graduate studies I have had the privilege of studying under two great experts in this field. I am deeply indebted to you for your patience, guidance and opportunities you have provided me.

My father and brother, I dedicate this thesis to you. Everything that I am today, all my accomplishments and endeavors, I owe to you.

I have had great friends who continue to support me in all dimensions of my life. My roommate and dear friend Ricardo, you are akin to my own family. Farsheed, I know I can always count on you and I thank you for that. Collin, with your kindness and warm personality, it is always a pleasure to have you around. Alex, you have made these last couple of years of my graduate studies really fun! Mike, I was lucky enough to see you get married and even luckier to have known you. Adam, it was a pleasure working with you, you are greatly missed. Tod, I thank you for all your support. Bert, your presence is surely to bring a good time! Micheal, I always look forward to your visits, you are always welcomed. Ken, without you we all would have missed every deadline! Sean, you are amongst my most disciplined friend, I have always enjoyed our chats. Dustin, your relaxed and cheerful approach to everything is always refreshing. Clifford, you’ve been our athletic role model!

I would also like to acknowledge CMC, Gennum, Nortel and STMicroelectronics for
their financial support which made this work possible.

Ana, you are my love and my partner in this journey. I look forward to the future because you are by my side.
# Contents

## 1 Introduction

1.1 A Brief History of Data Converters .............................................. 1

1.2 mm-Wave Data Converter Applications ........................................ 2

## 2 SiGe Millimeter-Wave Track and Hold Amplifiers

2.1 Introduction and State of the Art .............................................. 5

2.2 Track and Hold Amplifier Design .............................................. 7

2.2.1 Broadband Low-Noise Amplifier (LNA) .................................. 7

2.2.2 Emitter Follower/Differential Pair and Track and Hold Stages (EF/DIFF and T/H) .................................................. 11

2.3 Conclusion ................................................................. 27

## 3 SiGe Millimeter-Wave Analog to Digital Converters

3.1 State of the Art mm-Wave ADCs and Architectures ......................... 28

3.2 Impact of the Data Distribution On ADC Performance ..................... 31

3.2.1 Direct Driving Method .............................................. 31

3.2.2 Passive Data Distribution Using Power Splitters ....................... 33

3.2.3 Passive Data Distribution Using Transmission Lines .................. 33

3.2.4 Active Data Distribution ........................................... 34

3.2.5 Data Distribution Selection ......................................... 35

3.2.6 Impact of Data Distribution on THA ................................ 36
3.2.7 THA and Data Tree Breakout ........................................ 40
3.2.8 THA and Data Tree Measurement Results ......................... 47
3.3 The 4-bit FLASH ADC-DAC CHAIN .................................. 51
  3.3.1 ADC-DAC Circuit Description .................................... 53
  3.3.2 ADC-DAC Measurement Results .................................. 61
3.4 Conclusions ............................................................... 64

4 Toward CMOS Millimeter-Wave Data Converters ...................... 67
  4.1 Introduction ............................................................. 67
  4.2 A 30-GS/s THA in 0.13-μm CMOS Technology ....................... 67
    4.2.1 Introduction and State of the Art ................................ 67
    4.2.2 Track and Hold Amplifier Design ................................ 68
    4.2.3 Transimpedance Amplifier (TIA) .................................. 70
    4.2.4 Track and Hold Stage .............................................. 70
    4.2.5 Clock Distribution Network ...................................... 72
    4.2.6 Fabrication and Testing .......................................... 72
    4.2.7 Conclusions ....................................................... 75
  4.3 An 81Gb/s, 1.2V TIALA-Retimer in Standard 65nm CMOS .......... 77
    4.3.1 Introduction ....................................................... 77
    4.3.2 Circuit Design .................................................... 80
    4.3.3 Fabrication ....................................................... 86
    4.3.4 Measurement Results ............................................. 86
    4.3.5 Conclusions and Future Work .................................... 95

5 Conclusions and The Future of mm-Wave ADCs ...................... 98
  5.1 Contributions .......................................................... 98
  5.2 Future of mm-Wave ADCs ............................................. 99
List of Tables

1.1 Time line of major development in data converters. [1] . . . . . . . . . . 4

2.1 Table of comparison of state of the art published THAs. . . . . . . . . . 26

3.1 Table of comparison of state of the art published ADCs. . . . . . . . . . 66

4.1 Table of comparison of state of the art published THAs. . . . . . . . . . 96

4.2 Table of comparison of state of the art published retimers. . . . . . . . . 97
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Simplified system block diagram of a DSP-based fiber optic equalizer.</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Simplified block diagram of the 40-GSamples/s THA.</td>
<td>8</td>
</tr>
<tr>
<td>2.3</td>
<td>Schematic of an emitter-follower/differential pair (EF/DIFF).</td>
<td>9</td>
</tr>
<tr>
<td>2.4</td>
<td>Schematic of a transimpedance amplifier (TIA).</td>
<td>10</td>
</tr>
<tr>
<td>2.5</td>
<td>Circuit diagram of the emitter-follower/differential pair followed by the</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>track and hold block. Signals $I_{n_A}$ and $I_{n_B}$ are provided by the TIA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>output.</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Equivalent T/H half-circuit in track mode.</td>
<td>12</td>
</tr>
<tr>
<td>2.7</td>
<td>Equivalent T/H half-circuit in hold mode.</td>
<td>15</td>
</tr>
<tr>
<td>2.8</td>
<td>Schematic of the emitter-follower inverter and double emitter-followers</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>used in the clock network.</td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>Simulated (Top) single-ended and (Bottom) differential outputs of a 10 GHz</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>sinusoid sampled at 40 GHz.</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>Chip micrograph of the 40-GSamples/s THA.</td>
<td>20</td>
</tr>
<tr>
<td>2.11</td>
<td>Simulated and measured single-ended THA input return loss ($S_{11}$),output</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>return loss ($S_{22}$), and transmission ($S_{21}$).</td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>Measured (A) single-ended and (B) differential outputs of a 10 GHz sinusoid</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>sampled at 40 GHz.</td>
<td></td>
</tr>
<tr>
<td>2.13</td>
<td>Measured P1dB versus input frequency sampled at 40 GSamples/s.</td>
<td>23</td>
</tr>
<tr>
<td>2.14</td>
<td>Spectrum of IIP3 measurement at 11 GHz.</td>
<td>23</td>
</tr>
<tr>
<td>2.15</td>
<td>Spectrum of IIP3 measurement at 19 GHz.</td>
<td>24</td>
</tr>
</tbody>
</table>
2.16 Measured IIP3 versus input frequency. ........................................... 24

2.17 Measured THD at the 1dB compression point and SFDR versus input
frequency at 40 GSamples/s. ................................................................. 25

2.18 Beat frequency test: input signal at 40.002 GHz sampled at 40 GHz. ... 25

3.1 High-speed ADC architectures: (A) A time-interleaved architecture with
sub-rate THAs and multiphase clock generator (e.g., [2–4]). (A) Direct
driving the flash ADC comparator bank using a full rate THA (e.g., [5]).
(C) Using a transmission line to distribute the input and the clock signal
(e.g., [6]). (D) Implementation of active distribution networks to route
the input and clock signal to the flash ADC comparator bank, as in this
work [7]. .......................................................... 29

3.2 Various data distribution methods in high-speed ADCs: (A) Direct driving
of comparators or sub-ADC bank. (B) Passive data distribution in the
form of a broadband power splitter tree. (D) Passive data distribution
using a transmission line. (C) Active data distribution in the form of a
data tree. .............................................................. 32

3.3 This model is used to simulate the efficacy of a THA in the presence of
data tree bandwidth limitations. Without sufficient data tree bandwidth,
the THA is unable to combat the effects of the clock and data skew ($\tau_{d1-d15}$
and $\tau_{c1-c15}$, respectively). ............................................................... 37

3.4 A linear THA model used to analyze the impact of the analog front-end
bandwidth on the ADC performance. .................................................. 37

3.5 Simulated histogram of a 4-bit Flash ADC in the presence of clock and data
skew with standard deviations of $0.05T_{clk}$ and no THA. The performance
of the ADC degrades with increasing input frequency. ...................... 39
3.6 Impact of the data tree bandwidth on the THA waveform: (A) $\omega_{3dB} = 2\omega_{clk}$, (B) $\omega_{3dB} = 0.8\omega_{clk}$, (C) $\omega_{3dB} = 0.5\omega_{clk}$. The input signal frequency is $\omega_{3dB} = 0.14\omega_{clk}$. The hold mode behavior of the THA is mostly lost if the data tree bandwidth is limited.

3.7 Simulated ENOB of a 4-bit Flash ADC in the presence of clock and data skew with standard deviations of $0.05T_{clk}$. It can be observed that if the data tree bandwidth is limited to $0.5\omega_{clk}$, a THA offers little improvement in performance.

3.8 Block diagram of the fabricated breakout. This breakout allows for the characterization of the TIA, THA, and the BiCMOS data tree. The THA can be disabled for bandwidth measurement.

3.9 Schematic of the implemented fully bipolar TIA. The TIA has 12 dB of simulated differential gain and drives the THA.

3.10 Schematic of the MOS-HBT cascode differential pair used in the data tree. CMOS input transistors offer high linearity while the bipolar cascode transistors offer low output capacitance.

3.11 Chip micrograph of the ADC slice; the die area is $2.5 \times 1 \text{mm}^2$.

3.12 Small-signal characteristics of the fabricated breakout. Measured single-ended s-parameters and simulated single-ended gain of the TIA, THA, data tree and overall s-parameters of the breakout after extraction are shown. The combination of the TIA, THA and the data tree has a measured small-signal 3 dB bandwidth of 16 GHz.

3.13 Measured output of the fabricated breakout: (A) THA OFF; (B) THA ON. Due to the limited data tree bandwidth, the hold mode behavior of the THA amplifier is nearly eliminated.
3.14 Measured spectra of the breakout output showing the nonlinearity impact of the THA: (A) THA disabled; (B) THA enabled showing intermodulation products of the input and clock signals in the output spectrum.

3.15 Measured output of the single-bit thermometer output. The input signal is at 312.5 MHz with a comparator clock frequency of 40 GHz.

3.16 Measured single-bit thermometer output for two different comparator threshold settings: (A) zero-level threshold, (B) quarter-level threshold. The input signal is at 10 GHz with a comparator clock frequency of 40 GHz.

3.17 Block diagram of the implemented 35-GS/s, 4-bit, Flash ADC-DAC. The ADC uses data and clock distribution to drive the comparator bank.

3.18 Schematic of the merged MOS-HBT BiCMOS cascode and THA. The total capacitance at node (A) is reduced by connecting the collector of transistor $Q_9$ to the low impedance cascode node (B) instead.

3.19 Schematic of MOS-HBT BiCMOS cascode offset amplifier. The offset currents are drawn from the low impedance cascode nodes to mitigate their impact on bandwidth.

3.20 Schematic of the implemented 15-level thermometer code DAC. Thermometer code inputs (D0-D15) are summed together at different nodes of the double cascode DAC.

3.21 Chip micrograph of the ADC-DAC, die area is 2.5 x 3.2 mm².

3.22 Measured delay cell phase adjustment at 40 GHz.

3.23 Measured ADC INL and DNL with a 35-GHz clock frequency.

3.24 Measured DAC INL and DNL with a 35-GHz clock frequency.

3.25 Measured ENOB and SFDR versus input frequency for a 35-GHz clock frequency. The THA degrades the performance of the ADC for input frequencies above 8 GHz.
3.26 Output spectrum of a 7.98 GHz input signal sampled at 35 GHz after accounting for the sin(x)/x DAC roll-off and setup losses. The THA is disabled.

4.1 (Left) Example of a sampling diode bridge. (Right) Example of a switched emitter follower (SEF).

4.2 (Left) Example of a clocked series CMOS switch. (Right) Example of a switched source follower (SEF).

4.3 Block diagram of the implemented THA.

4.4 Schematic of the TIA and CS stages. All MOSFETs have a drawn length of 0.13 \( \mu \text{m} \).

4.5 Circuit diagram of the differential pair, SSF and output driver blocks. All MOSFETs have a drawn gate length of 0.13 \( \mu \text{m} \).

4.6 Circuit diagram of the differential pair, THA and output driver blocks. All MOSFETs have a drawn length of 0.13 \( \mu \text{m} \).

4.7 Die photo of the fabricated THA. The chip area is 1 mm\(^2\).

4.8 Simulated and measured single-ended THA input return loss (\( S_{11} \)), output return loss (\( S_{22} \)) and transmission (\( S_{21} \)).

4.9 Single ended outputs of a 7.5-GHz sinusoid sampled at 30 GHz.

4.10 Measured input compression point (\( P_{i1dB} \)) and output compression point (\( P_{o1dB} \)) versus input frequency with 30-GHz clock frequency.

4.11 Measured IIP3 and OIP3 versus input frequency with 30-GHz clock frequency.

4.12 Measured SFDR and THD versus frequency with 30-GHz clock frequency. The measured THD is shown for signals with -12 dBm input power.

4.13 Reported record operating speeds for multiplexers ([8–10]), T-FFs ([11–13]) and D-FFs ([14–16]).

4.14 Operational time chart for multiplexers, T-FFs and D-FFs.

4.15 The TIALA-Retimer block diagram.
4.33 Measured input and output eye diagrams at 78Gb/s. The input eye height/amplitude is $12\text{mV}_{\text{PP}}/60\text{mV}_{\text{PP}}$. .......................... 93

4.34 Measured input and output eye diagrams at 81Gb/s. The input eye height/amplitude is $63\text{mV}_{\text{PP}}/250\text{mV}_{\text{PP}}$. .......................... 93

4.35 Measured input and output eye diagrams at 81Gb/s. The input eye height/amplitude is $15\text{mV}_{\text{PP}}/80\text{mV}_{\text{PP}}$. .......................... 94

4.36 Correct measured input and output $2^7-1$ PRBS patterns at 81Gb/s corresponding to Figure 4.35] .......................... 94

4.37 Improved schematics of the tuned clock tree. .......................... 96
1

Introduction

1.1. A Brief History of Data Converters

Some of earliest implemented data converters were not electrical at all, but rather hydraulic. During the 18\textsuperscript{th} century in Turkey, sophisticated water metering systems were conceived using binary-weighted tubes and dams. These systems were undoubtedly not referred to as data converters, however their principle of operation were akin to a digital-to-analog converters (DAC) we know today \[1\].

During the 19\textsuperscript{th} century, the creation of communication systems led to the invention and wide spread use of electronic data converters. Telegraphs led to the invention of the telephone and the establishment of Bell Systems. The rapid growth of the telephone infrastructure demanded more capacity and the desire to push more data through existing copper wires. Thus, modulation schemes were born and in 1921 pulse code modulation (PCM) brought multi-bit digital data transmission to life. Later during that decade, Harry Nyquist published his classic papers where he calculated the minimum sampling frequency required to avoid loss of information in a sampled system. Furthermore, it took another two decades before the famous paper by Shannon, Bennett, and Oliver was published which solidified PCM and information theory for all time. \[2\].

During the second world war, Bell labs began their own research on PCM modulation with the intention of encrypted transmission of voice for military purposes. During
this time (although not published until several years after the war), the first successive approximation ADC, electron beam coding tube, “Shannon-Rack” decoder (DAC) and Germanium transistor were invented. During the 1950’s and 60’s, the first commercial ADCs became available: an Epsco, 11-bit, 50-kSPS vacuum-tube based ADC which consumed 500W and cost more than $8000. During this time the main driving force for developing faster ADCs with higher resolution was for radar applications. The United States military was in the process of implementing anti-ballistic missile (ABM) systems. These systems used a digitally controlled phased array radars for guiding short and long-range interceptor missiles. [1].

During the 1970’s ADCs had found their way into many applications which had previously used analog processing in the past. The ADC/DAC market was driven by high resolution digital voltmeters, industrial process control, digital video, military phased array radar, medical imaging, vector scan displays, and raster scan displays. Nearly two decades after the invention of the planer process by Jean Hoerni of Fairchild Semiconductor in 1959, commercial ADCs became monolithic, first using bipolars and finally CMOS. By the end of the 1980’s and early 1990’s, ADCs and DACs had become fully integrated, single-chip, complex CMOS circuits found in almost every application. Table 1.1 shows a time line of major developments in data converters [1].

1.2. **mm-Wave Data Converter Applications**

A robust and integrated wireline receiver solution is to employ digital signal processing (DSP) for dispersion and intersymbol interference (ISI) compensation and recovery of the clock. DSP-based equalization can be employed for optical links or to address the need to push higher data rates through existing low frequency copper wireline infrastructures. Electrical equalization (if possible) is an inexpensive alternative to replacing the existing wired infrastructure. However, a major bottleneck in realizing a DSP-based equalizer is the implementation of the preceding analog-to-digital converter (ADC). DSP-based
equalizers using baud-rate ADCs have been demonstrated in the past \[17,18\]. Furthermore, baud-rate ADCs enable soft-error correction within DSP-based receivers. For instance, \[19\] has demonstrated a powerful forward error correction for 10 Gb/s optical communication systems using a 3-bit soft decision IC.

Data converters aimed for high-end instrumentation have stringent bandwidth and resolution requirements but relaxed power, area and cost requirements. These type of expensive instruments are sold in low volumes and typically boast record breaking data converters at their core. Furthermore, depending on the application, the desired ENOB varies. Oscilloscopes and arbitrary waveform generators aim to maximize the effective number of bits (ENOB) while \[6\] recommends sampling at twice the data rate combined with an ENOB of over 4-bits in the frequency band of interest.

The aim of this thesis is to investigate the challenges associated with implementing mm-wave data converters. Chapter 2 explores the design methodology of mm-wave SiGe track and hold amplifiers (THA) while Chapter 3 investigates the impact of signal distribution on THA performance in mm-wave ADCs. Chapter 4 introduces an array of implemented CMOS building blocks for next generation ADCs and data communications systems. The concluding remarks and future work are presented in Chapter 5.
<table>
<thead>
<tr>
<th>Year</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1853</td>
<td>Time Division Multiplexing</td>
</tr>
<tr>
<td>1924</td>
<td>Nyquist Criterion</td>
</tr>
<tr>
<td>1937</td>
<td>Pulse Code Modulation</td>
</tr>
<tr>
<td>1939</td>
<td>Reeves Counting ADC</td>
</tr>
<tr>
<td>1946</td>
<td>Successive Approximation Register ADC</td>
</tr>
<tr>
<td>1948</td>
<td>Flash ADC With Electron Tube</td>
</tr>
<tr>
<td>1949</td>
<td>Grey Coding</td>
</tr>
<tr>
<td>1950</td>
<td>Delta Modulation, Differential PCM</td>
</tr>
<tr>
<td>1952</td>
<td>Voltage to Frequency Converter</td>
</tr>
<tr>
<td>1954</td>
<td>DATRAC: 11-bit, 50 Ksps Vacuum Tube ADC 500 W Power Dissipation</td>
</tr>
<tr>
<td>1956</td>
<td>Subranging ADC</td>
</tr>
<tr>
<td>1956</td>
<td>Subranging With Error Correction</td>
</tr>
<tr>
<td>1957</td>
<td>Dual Slope ADC</td>
</tr>
<tr>
<td>1962</td>
<td>Sigma Delta Modulation</td>
</tr>
<tr>
<td>1966</td>
<td>HS-810: 8-bit, 10 Msps, 150 W</td>
</tr>
<tr>
<td>1967</td>
<td>Triple Slope ADC</td>
</tr>
<tr>
<td>1969</td>
<td>Successive Approximation Register Realized With Fourteen 7400 Series IC</td>
</tr>
<tr>
<td>1973</td>
<td>Quad Slope ADC</td>
</tr>
<tr>
<td>1978</td>
<td>First Monolithic ADC: AD571 SAR, 10-bit, 25μs</td>
</tr>
<tr>
<td>1980</td>
<td>First Silicon Time-Interleaved Converter Array</td>
</tr>
<tr>
<td>1988</td>
<td>Bandpass Sigma Delta converter</td>
</tr>
<tr>
<td>2004</td>
<td>First mm-Wave (&gt; 30 GHz) Flash ADC: 3b, 40GS/s ADC-DAC in SiGe</td>
</tr>
</tbody>
</table>

Table 1.1: Time line of major development in data converters. [1]
2.1. Introduction and State of the Art

In the past, optical fiber was considered to be an infinite bandwidth medium. However, the rapid scaling of data rates over optical networks has exposed two important fiber impairments: polarization mode dispersion (PMD) in single-mode fibers and differential mode dispersion in multi-mode fibers. Thus, to ensure error-free communication at bit rates above 10 Gb/s, some method of data equalization is required. Electronic equalization integrated in the receiver offers lower cost and faster adaptation than all-optical dispersion compensation. Decision feedback and analog equalizers have been demonstrated up to 40 Gb/s [20,21]. It can be argued that analog equalization is more practical at high data rates than digital equalization. However, assuming high-speed analog-to-digital converters (ADCs) can be realized, digital equalization is more robust, scalable, and offers more flexibility. Figure 2.1 illustrates a system block diagram of a DSP-based fiber optic equalizer. At the heart of this system is an ADC clocked at the full bit rate. The design of a 40-GSamples/s track and hold amplifier (THA) is a prerequisite to implementing 40-Gb/s ADCs. Applications for mm-wave THAs and ADCs are not limited to DSP-based equalizers. THAs can also be used for satellite and wireless communication [19,22], high-speed soft-decision-based forward error correction systems [9], military radar systems [5] and instrumentation (i.e., wide bandwidth sampling oscilloscopes) [23].
The fastest THA reported to date is a distributed switched emitter-follower, 50-GSamples/s THA implemented in SiGe HBT technology [24]. A 3-bit, 40-GSamples/s ADC-DAC, employing a switched diode bridge sampler as its input stage, is presented in [5]. The sample and hold (S/H) block is followed by a rather high-noise emitter-follower/differential pair (EF/DIFF) stage with resistive degeneration, which acts as a preamplifier before the quantizer. Since the switched diode bridge provides no gain, its input-referred noise is further degraded by the high noise of the second stage (EF/DIFF). The reported spurious-free dynamic range (SFDR) at frequencies beyond 15 GHz is less than 13 dB (1.5 bits) and the overall bandwidth is limited by the S/H block to 13 GHz [5].

In this work, the design and characterization of a 40-GSamples/s switched emitter-follower THA is described. The goal is to extend the bandwidth of this architecture to make it suitable for the next generation mm-wave ADCs while achieving at least 4-bits of linearity. It is based on a topology first proposed in [25] and also used in [23, 26, 27]. Unlike [5], the order of the sampling block and of the preamplifier is reversed, and unlike [5, 23, 25–27], a low-noise preamplifier is used as an input stage. This combination is one of the underlying reasons for achieving good performance. Sampling frequency and bandwidth of 40 GHz are achieved by applying a systematic design methodology which combines a low-noise broadband front-end, signal feedthrough cancelation, and a droop-rate-minimizing architecture. The bandwidth of every circuit block is optimized with attention to proper biasing, layout, and component sizing. The design methodology of the THA and the measurement results are presented next.
2.2. Track and Hold Amplifier Design

Figure 2.2 shows a simplified block diagram of the track and hold amplifier. At the front end is a differential broadband low-noise amplifier (LNA) which is noise and impedance matched to 50 Ω per single-ended input. Its linearity, bandwidth, and noise figure determine the overall system performance. The output of the LNA goes to an emitterfollower/differential pair (EF/DIFF) stage which, in turn, drives the track and hold block at the core of this system. The output of the track and hold stage is buffered by a differential pair and fed to a linear output driver with 50-Ω output impedance. A high-bandwidth clock distribution network delivers a 40-GHz differential clock signal to the THA. The following sections cover the design methodology for each block.

2.2.1. Broadband Low-Noise Amplifier (LNA)

Traditionally, designers have used emitter-followers and heavily degenerated differential pairs as input stages [23, 25–27] (shown in Figure 2.3). While this approach can provide adequate bandwidth and linearity, using 50-Ω resistive loads for matching or heavy resistive degeneration ($R_E$) results in poor noise figure. In contrast, a transimpedance amplifier (TIA), Figure 2.4, can be designed to provide simultaneous noise and impedance matching without the need for 50-Ω matching resistors [28]. It employs a shunt-shunt feedback scheme which can be used to adjust the input impedance by varying the feedback resistor $R_F$.

The linear output voltage swing of the TIA is set to 600 mV. Input impedance matching is achieved as described in [28], which results in the use of 200-Ω feedback resistors, $R_F$. To improve the linearity at the input of the TIA, two 15-Ω degeneration resistors are also added. It is important to note that the linear output swing must be chosen with consideration to the desired bandwidth as well. A large value for $R_L$ would result in a low pole frequency at that node, limiting the overall bandwidth. In this design, simulations show that the bandwidth of the TIA exceeds 40 GHz. Noise matching is achieved by siz-
ing the input transistors ($Q_3, Q_4$) to have 50 Ω optimum source impedance. Given noise parameters, $G, G_C, R$ and $B$ (technology constants which depend on the bias current density), and $\omega_o = \omega L_o/R_F$, the following equation can be used to determine the optimal emitter length of the input devices at angular frequency [28]:

$$l_{E_{OPT}} = \frac{1}{\omega} \left[ \frac{1}{R_F (1 + \omega_o^2)} \right] \sqrt{\frac{R}{(G/R) + G_C^2 + B^2}}. \quad (2.1)$$

The transistors are biased at minimum noise figure current density $J_{OPT}$ and $\omega_o$ is set for $20/(2\pi)$ Grad/s. The bandwidth of this stage is further enhanced by using inductive peaking both in the feedback path and at the output. The feedback inductor $L_o$ also filters high-frequency noise. The measured noise figure of SiGe HBT TIAs and emitter-follower/differential pairs for 40-Gb/s applications were compared in [29], demonstrating the superiority of the TIA stage.
Figure 2.3: Schematic of an emitter-follower/differential pair (EF/DIFF).
Figure 2.4: Schematic of a transimpedance amplifier (TIA).
2.2.2. Emitter Follower/Differential Pair and Track and Hold Stages (EF/DIFF and T/H)

The output of the TIA is fed to an emitter-follower/differential pair stage which in turn drives the track and hold (T/H) block, as shown in Figure 2.5. The T/H block consists of a pair of switched emitter-followers and the hold capacitor, \( C_H \). The behavior of the circuit in Track and in Hold modes must be optimized as each mode of operation contributes to the overall system performance.

2.2.2.1. Track Mode Operation

Figure 2.6 illustrates the equivalent half-circuit schematic of the T/H block in track mode. During track mode, the signal \( Clk_A \) is at a high current mode logic (CML) level while the signal \( Clk_B \) is low. Thus, currents \( I_{T1} \) and \( I_{T2} \) flow through the transistors \( Q_{EF1} \) and \( Q_{EF1} \) respectively, which act as emitter-followers. The signal \( In_A \) (for instance) is amplified by the EF/DIFF stage and buffered by the two consecutive emitter-followers before arriving at the output terminal \( Out_A \). However, there are a few major sources of nonidealities in track mode; for instance, modulation of the base-emitter voltage of the emitter-follower as a function of the input voltage and the nonlinearities of the emitter-follower/differential pair.

During track mode, the current of \( Q_{EF1} \) is modulated by the current \( i_{CH} \) required to charge and discharge the hold capacitance. This modulation leads to variations in the base-emitter voltage of \( Q_{EF1} \) as a function of the input voltage and is undesired. If \( v_{in} \) and \( v_{out} \) denote the small-signal input and output voltages and \( V_t \) is the thermal voltage \( kT/q \), the output voltage of the switched emitter-follower is given by

\[
v_{out} = v_{in} - V_t \ln \left( \frac{I_{T1} + 2i_{CH}}{I_{T1}} \right).
\]  

(2.2)

For a sinusoidal input with frequency \( f_{in} \) and amplitude \( A \), the maximum charge
Figure 2.5: Circuit diagram of the emitter-follower/differential pair followed by the track and hold block. Signals $I_{inA}$ and $I_{inB}$ are provided by the TIA output.

Figure 2.6: Equivalent T/H half-circuit in track mode.
current is

\[ i_{CH} = C_H \frac{dv_{in}}{dt} = 2\pi f_{in} AC_H. \quad (2.3) \]

Given that the circuit is implemented differentially, only the third-order harmonic is of interest. It can be shown that the third-order harmonic \((HD_3)\) is

\[ HD_3 = \frac{V_{t}^3 i_{CH}^3}{12I_{T1}^2 \left( \frac{AI_{T1}}{2} + V_{t} i_{CH} \right)} = \frac{V_{t}^3 (2\pi f_{in} AC_H)^3}{12I_{T1}^2 \left[ \frac{AI_{T1}}{2} + V_{t} (2\pi f_{in} AC_H) \right]} \quad (2.4) \]

It is immediately apparent that the total harmonic distortion can be improved by increasing the tail current \(I_{T1}\) or by reducing the hold capacitance \(C_H\). Increasing the tail current beyond 5 mA makes transistor \(Q_{EF1}\) too large, resulting in inadequate bandwidth. On the other hand, reducing the hold capacitance excessively would result in significant hold mode distortion.

Based on a maximum swing of 240 mVpp at the input of the TIA, the gain of the two blocks preceding the T/H, a sampling rate of 40-GSamples/s, and a tail current of 5 mA, a hold capacitance of 175 fF was chosen to attain 28 dB of linearity. Simulations indicate a total of 25 fF of parasitic capacitance at the hold node. The remaining required capacitance is provided via a 150 fF MIM capacitor. The size of the hold capacitance also impacts the hold mode performance of the THA. As a result, a compromise between the track mode and hold mode performance is made to achieve acceptable bandwidth and linearity.

The EF/DIFF stage preceding the T/H block is also optimized for linearity. Referring to Figure 2.5, the maximum linear input swing of the EF/DIFF stage can be calculated based on the circuit component parameters and biasing. Given the small-signal gain of the differential pair \(A_v\), the collector-emitter voltage \(V_{CE-Q3,4}\) and tail current \(I_{DIFF}\), the
maximum linear input voltage swing is

\[ MaxSwing = \min \left( \frac{I_{\text{DIFF}} R_L}{A_V}, \frac{2(V_{\text{CE-Q3,4}} - V_{\text{CE-SAT}})}{A_V}, (1 + g_{m-Q3,4} R_E) 2V_t \right) \]  

(2.5)

Knowing the desired signal amplitude at the input of the T/H block, one can ensure sufficient linear voltage swing range, both at the input and output. Transistors \( Q_{1-4} \) are biased at peak current density \( J_{PfT} = 1.2 \text{ mA/\mu m} \) to maximize speed. Inductive peaking is used to extend the 3-dB bandwidth of this stage.

Due to large capacitive loads at their emitters, which can lead to negative resistance and oscillation, transistors \( Q_{EF1}, Q_{EF2} \) and are biased at 1/4 and 1/3 peak current density (as a rule of thumb), respectively. Biasing these transistors at higher current densities would result in excessive peaking in the AC response of the circuit. The switching transistors in this block are biased for maximum switching speed at 1.5 times peak \( f_T \) current density, when fully switched.

2.2.2.2. Hold Mode Operation

During hold mode, the signal \( Clk_N \) is at a CML level low while the signal \( Clk_P \) is high. Thus, \( I_{T1} \) and \( I_{T2} \) flow through \( Q_5 \) and \( Q_6 \), respectively (Figure 2.7). The hold capacitance is now isolated from the input by \( Q_1 \) and separated from the next stage by \( Q_2 \). Three major hold mode distortions degrade the performance of the circuit: 1) inadequate isolation from the input to the hold node due to \( Q_1 \) remaining on; 2) input signal coupling to the hold capacitance \( C_H \) through \( C_{BE-Q1} \); and 3) signal droop during hold mode. During hold mode \( C_H \), must be isolated from the input signal, which arrives from the EF/DIFF stage. \( I_{T1} \) is routed away from \( Q_1 \), through the resistive load \( R_L \). The resulting voltage drop at the base of \( Q_1 \) reduces \( V_{BE-Q1} \) by nearly a half of its DC value to ensure that \( Q_1 \) is fully turned off. The voltage drop at the base of \( Q_1 \) can be
approximated by

$$\Delta V_{b-Q1} \approx I_T R_L + V_T$$  \hspace{1cm} (2.6)$$

and is set to approximately $V_{BE-Q1}/2$. It is important to note that the value of $R_L$ is chosen to concurrently satisfy linearity and hold-mode isolation criteria.

Hold mode signal feedthrough occurs as the input signal couples through the parasitic capacitance $C_{BE-Q1}$ to the hold capacitance $C_H$. The hold mode feedthrough is

$$V_{fth} = \frac{C_{BE-Q1}}{C_{BE-Q1} + C_H} V_{in}.$$  \hspace{1cm} (2.7)$$

In a high-speed THA, the hold capacitance must be small to improve bandwidth and total harmonic distortion (THD). Furthermore, transistor $Q_1$ is large and therefore $C_{BE-Q1}$ is also large. Hence, signal feedthrough poses limitations on the effective number of bits the THA can achieve. With the addition of two feedthrough cancellation capacitors $C_F$, feedthrough can be reduced. Due to the differential nature of the circuit, the feedthrough caused by capacitors $C_F$ is of opposite polarity compared to $C_{BE-Q1}$.
Therefore, signal feedthrough can be expressed as

\[ V_{fth} = \frac{C_{BE-Q1}}{C_{BE-Q1} + C_H} \left( \frac{C_{BE-Q1} - C_F}{C_{BE-Q1}} \right) V_{in}. \]  

(2.8)

If \( C_{BE-Q1} \), hold mode signal feedthrough should be completely eliminated. However, the addition of capacitors \( C_F \) significantly degrades the bandwidth of the system. Note that the impact of \( C_F \) is doubled due to the Miller effect. As a result, in this design only partial feedthrough cancellation \( (C_F = 0.66C_{BE-Q1}) \) is applied in order to achieve more bandwidth.

Signal droop during hold mode is caused by the gradual discharge of the hold capacitance and must be minimized. In this design, \( Q_2 \) isolates the hold capacitor from the next stage. The current required by the input of the next stage (roughly \( I_{TAIL}/\beta \)) is provided by the emitter current of \( Q_2 \), and not by the discharge current of \( C_H \). This approach improves the droop rate by a factor of \( \beta \). The disadvantage is the need to provide a high-speed clock to these switching pairs. This droop rate improvement technique has been employed in [26] and [25], but some designs, such as [23] and [27] omit the second switching emitter-followers.

The input-referred noise power spectral density of the first two stages of the THA (TIA + EF/DIFF) obtained from simulation reaches a value of \( 4 \text{nV}/\sqrt{\text{Hz}} \) at 20 GHz and a maximum of \( 5 \text{nV}/\sqrt{\text{Hz}} \) at 40 GHz. The equivalent input-referred noise integrated from 10 MHz to 40 GHz is \( 757 \mu \text{V}_{\text{RMS}}, (-50 \text{dBm}) \). For a receiver with signal-to-noise ratio (SNR) of 17 dB, corresponding to an eye \( Q \) of 7 and a bit-error rate (BER) of \( 10^{-12} \), this results in an input sensitivity of 5.3 mV.

2.2.2.3. Clock Distribution Network (Clock Tree)

The clock path converts a single-ended 150 mV clock input to a differential signal with 300 mV per side swing and drives four switched emitter-follower pairs in-phase. Its bandwidth must exceed 40 GHz. The clock distribution consists of a tree of emitter-follower
inverters (EF/INV), each with a fan-out of two. The four differential outputs of the clock tree drive the T/H circuit through double emitter-follower stages (EF/EF). Figure 2.8 illustrates the schematic of the final two stages, EF/INVEF/EF.

2.2.2.4. Output Driver

The output driver is a differential pair with 50-Ω loads and a tail current of 24 mA. Resistive degeneration (20 Ω) is used to accommodate an input swing of 500 $mV_{P-P}$ per side (determined by the desired input swing of the TIA and the overall system gain). The output return loss is improved by inductive peaking. A differential pair stage precedes the output driver and acts as a level shifter and buffer. The transistors in both stages are biased at peak $f_T$ current density for maximum speed.

Figure 2.9 presents the simulation results for the entire chip. Transistor level extraction tools were unavailable at the time of the fabrication, however inductors and interconnects were modeled using the ASITIC tool [30]. Using this tool, the conductive
Figure 2.9: Simulated (Top) single-ended and (Bottom) differential outputs of a 10 GHz sinusoid sampled at 40 GHz.
substrate, proximity effects (capacitive and inductive coupling) and high frequency skin effects of monolithic inductors and transformers can be simulated. A broadband circuit model of the inductor can then be realized and used in the circuit simulations [31].

Figure 2.9 (Top) shows the single-ended outputs with a 10 GHz sinusoid signal sampled at 40 GHz. The 40 GHz clock signal is superimposed to illustrate the location of the sampled values. Figure 2.9 (Bottom) shows the differential output of a 10 GHz signal sampled at 40 GHz. It can be observed that the common mode clock feedthrough is significantly reduced.

The chip was fabricated in Jazz Semiconductors SBC18HX 0.18-μm SiGe BiCMOS technology with $f_T$ a of 160 GHz. The chip area is $1.1 \text{mm}^2$ and the die photo is shown in Figure 2.10. The circuit blocks of the THA are indicated in the die photo. Careful attention was paid to layout symmetry. Each block consists of identical half-circuits to ensure matching and low skew propagation, both in the signal and in the clock paths. The inductors were designed using ASITIC [30] and are realized in the thick top metal layer.

The circuit operates from a 3.6-V supply and draws 150 mA (540 mW). Due to the large number of stacked transistors in the track and hold block, a 3.3-V supply was not sufficient. The T/H block and the clock distribution network consume 24 mA and 75 mA, respectively, while the remaining current is drawn by the input and output stages. All measurements were conducted on-wafer.

The simulated and measured S-parameters are shown in Figure 2.11. The measured single-ended input and output return loss are better than -15 dB up to 26 GHz and 42 GHz, respectively. $S_{21}$ shows a bandwidth of 43 GHz when the circuit is configured in track mode. The ripple in the measured $S_{21}$ of the circuit (which is not present in the simulated $S_{21}$) is due to larger than anticipated inductors in the differential pairs and the peaking due to the emitter-followers in the track and hold block.

Time domain measurements were conducted using an Agilent 86100C DCA-J oscillograph.
Figure 2.10: Chip micrograph of the 40-GSamples/s THA.

Figure 2.11: Simulated and measured single-ended THA input return loss ($S_{11}$), output return loss ($S_{22}$), and transmission ($S_{21}$).
Figure 2.12: Measured (A) single-ended and (B) differential outputs of a 10 GHz sinusoid sampled at 40 GHz.
scope and an Agilent E8257D signal source. The two outputs of the THA were displayed on channels 3 and 4, respectively, and the differential output was calculated using the built-in functions of the oscilloscope. A constant delay was added to one of the channels to compensate for the difference in electrical delays between cables. The input power used is -12 dBm with -16 dBm of clock input power. Figure 2.12 illustrates the measured single-ended and differential outputs of a 10-GHz signal sampled at 40 GSamples/s. In Figure 2.12(A), the measured 40-GHz clock signal is superimposed on the single-ended output signal to more clearly indicate the location of each sample. As in the simulations, clock feedthrough (which appears in common mode) is present in the single-ended output and is significantly reduced in the differential measurement Figure 2.12(B). The observed improvement between the measured differential output (Figure 2.12(B)) and the simulated differential output (Figure 2.9(Bottom)) is due to variation in clock signal amplitude delivered to the SEF in each case. For consistency, the clock input power to the chip is kept constant in both simulation and measurement, however due to uncaptured bandwidth limitations of the clock tree in simulation, the SEF is overdriven, which results in excessive clock feedthrough and distortion. During measurements however, the SEF receives optimal clock signal amplitude which results in a better differential output.

The spectral content of the output signal was captured using an Agilent E4448A PSA spectrum analyzer. The input compression point of the circuit was measured for frequencies ranging from 2 to 20 GHz and is plotted in Figure 2.13. Figures 2.14 and 2.15 show the spectra for two tone tests at 11 and 19 GHz, respectively, for input signal powers of -14 dBm and -24 dBm. The input tones were separated by 100 MHz. The IIP3 is measured using two input tones separated by 100 MHz and is illustrated as a function of frequency in Figure 2.16 and has a 3-dB bandwidth of 10 GHz.

Figure 2.17 shows the measured SFDR and THD. The SFDR was calculated as the difference between the measured fundamental signal power and that of the third-order harmonic at the input power where the harmonic power exceeds the noise floor. THD
Figure 2.13: Measured P1dB versus input frequency sampled at 40 GSamples/s.

Figure 2.14: Spectrum of IIP3 measurement at 11 GHz.
Figure 2.15: Spectrum of IIP3 measurement at 19 GHz.

Figure 2.16: Measured IIP3 versus input frequency.
Figure 2.17: Measured THD at the 1dB compression point and SFDR versus input frequency at 40 GSamples/s.

Figure 2.18: Beat frequency test: input signal at 40.002 GHz sampled at 40 GHz.
values were measured as the difference between the fundamental signal power and the third-order harmonic at the P1dB input power. For a 19-GHz input, the SFDR and THD were 35 dB and 27 dB, respectively. The single-ended spectral characteristics of a beat frequency test with \( f_{in} = f_s + \Delta f \), \( f_s = 40 \, GHz \), and \( \Delta f = 2 \, MHz \) are shown in Figure 2.18. The difference between the power of the fundamental and of the second- and third-order harmonics is 24.7 dB and 42 dB, respectively, for a 16 dBm input signal power. It is expected that the second harmonic content should be much lower when the circuit is measured and operated in differential mode. Table 2.1 compares this work with other state-of-the-art track and hold amplifiers presented to date. At the time of the publication of this circuit, it represented the fastest reported THA.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Technology</th>
<th>( f_{sample} ) (GS/s)</th>
<th>Track BW (GHz)</th>
<th>THD dB @ ( f_{in} )</th>
<th>Supply (V)</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[32]</td>
<td>0.18µm SiGe BiCMOS</td>
<td>40</td>
<td>43</td>
<td>-27 @ 20 GHz</td>
<td>3.6</td>
<td>540</td>
</tr>
<tr>
<td></td>
<td>( f_T = 160 GHz )</td>
<td></td>
<td></td>
<td>-29 @ 10 GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[33]</td>
<td>0.18µm CMOS</td>
<td>10</td>
<td>N/A</td>
<td>-24.7 @ 5 GHz</td>
<td>1.8</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>( f_T = 45 GHz )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[26]</td>
<td>InP</td>
<td>12</td>
<td>14</td>
<td>-23.3 @ 12 GHz</td>
<td>-5.2</td>
<td>390</td>
</tr>
<tr>
<td></td>
<td>( f_T = 120 GHz )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[22]</td>
<td>0.25µm SiGe BiCMOS</td>
<td>12</td>
<td>5.5</td>
<td>-52.4 @ 1.5 GHz</td>
<td>3.7</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>( f_T = 200 GHz )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[24]</td>
<td>0.18µm SiGe BiCMOS</td>
<td>50</td>
<td>42</td>
<td>N/A</td>
<td>4, 3.3</td>
<td>640</td>
</tr>
<tr>
<td></td>
<td>( f_T = 200 GHz )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[34]</td>
<td>0.13µm CMOS</td>
<td>30</td>
<td>7</td>
<td>-30 @ 1 GHz</td>
<td>1.8</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td>( f_T = 85 GHz )</td>
<td></td>
<td></td>
<td>-29 @ 7 GHz</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Table of comparison of state of the art published THAs.
2.3. Conclusion

A 40-GSamples/s THA was designed and fabricated. Following a systematic design procedure, and combining a low-noise front-end, signal feedthrough cancelation, and a second pair of switched emitter-followers for droop rate improvement has yielded the highest sampling frequency for a switched emitter-follower THA at the time of publication. For 19-GHz input signals, a THD of 27 dB at the 1dB compression point has been measured and an SFDR of 35 dB has been achieved.
3.1. State of the Art mm-Wave ADCs and Architectures

At the time of writing, the fastest commercial data converter is an 8-bit, 50-GS/s ADC, albeit only available as part of a DPO/DSA72004B Tektronix oscilloscope [35]. This ADC uses a full-rate THA front-end and produces an ENOB of 5.6-bits for a 13-GHz input signal.

High speed ADCs reported in the literature use a variety of technologies and architectures. The highest sampling rates have been realized with SiGe and InP technologies [5, 6, 36]. To overcome the capacitive load associated with the comparator bank, in mm-wave flash or time-interleaved architectures, different techniques have been employed.

To achieve a combination of high sampling rate and high resolution, time-interleaved architectures have been demonstrated in [2], [3] and [4]. These designs rely on parallelism and require periodic calibration for offset, gain, skew and mismatch correction. A simplified block diagram of this approach is shown in Figure 3.1(A). The reported CMOS implementations use time-interleaved THAs to drive banks of time-interleaved sub-ADCs. In [2], the front-end consists of eighty parallel THAs and a SiGe amplifier is used to drive the 4-pF of total input capacitance. In [3] and [4], a power splitter is used to break up the input capacitance of the THAs. The latter approach requires a full scale
Figure 3.1: High-speed ADC architectures: (A) A time-interleaved architecture with sub-rate THAs and multiphase clock generator (e.g., [2–4]). (A) Direct driving the flash ADC comparator bank using a full rate THA (e.g., [5]). (C) Using a transmission line to distribute the input and the clock signal (e.g., [6]). (D) Implementation of active distribution networks to route the input and clock signal to the flash ADC comparator bank, as in this work [7].
Amongst the highest reported conversion speeds, 40 GHz sampling rate was achieved using a SiGe technology ($f_T/f_{MAX} = 210/310$ GHz) for a 3-bit ADC. In this approach (shown in Figure 3.1(B)) the comparator bank is treated as a lumped load to be driven by a THA.

A 5-bit, 22 GS/s ADC is presented in [6] where the input capacitance of the comparators is absorbed along a transmission line. This approach, which was originally implemented in [36], is shown in Figure 3.1(C). No track-and-hold amplifier (THA) or other active input stage was integrated. The implemented ADC in [6] requires a differential input signal amplitude of $1.28 V_{PP}$ and precise matching between the delay of the input data path and the clock distribution network to avoid clock-to-data skew.

This chapter presents an alternative architecture which is shown in Figure 3.1(D). The fabricated ADC ([7]) uses a tree of linear buffers to drive the capacitive load of the comparator bank. A TIA with 12 dB of differential gain is used as a relatively low-noise front-end amplifier. This combination allows the ADC to process small input signals ($0.24 V_{PP}$ differential), and drive the comparator bank through a symmetric data tree for minimized skew. A fully symmetric bipolar clock distribution network is employed to minimize clock skew. This technique relies only on the matching between identical blocks in each distribution tree. A THA may be inserted between the front-end amplifier and the data tree, but it will be shown that unless sufficient bandwidth is provided in the data tree such a track and hold amplifier can actually degrade the performance of the ADC.

The following sections discuss the impact of the data tree (passive or active) on the ADC performance and the efficacy of the THA for high frequency input signals. Measurements of a fabricated breakout comprising a THA followed by an active data tree are also presented. The implementation and measurement results of the fabricated 35-GS/s flash ADC are then discussed followed by concluding remarks.
3.2. Impact of the Data Distribution On ADC Performance

From the perspective of data distribution, high-speed flash ADCs and time-interleaved ADCs pose a similar problem: the problem of delivering the input signal to a bank of comparators or sub-ADCs with sufficient bandwidth and linearity while minimizing noise, skew and power consumption. Figure 3.2 demonstrates various methods of data distribution in mm-wave flash ADCs. In the case of an \( n \)-bit flash converter, \( 2^n - 1 \) comparators must be clocked at \( \omega_{\text{clk}} \). Each comparator presents an input capacitance of \( C_C \) with a quantization level of \( \Delta V \). This results in a full-scale input range of

\[
A = \Delta V (2^n - 1) .
\]  
(3.1)

Doubling the number of comparators results in one bit of resolution improvement. Given the above assumptions, simple analysis can yield a better understanding of the advantages and disadvantages of various data distribution methods; this is covered in the next few sections.

3.2.1. Direct Driving Method

This method (shown in Figure 3.2(A)) is the simplest solution to data distribution where all the comparators are driven simultaneously via an on chip or off-chip amplifier/THA block. Assuming a system impedance of \( Z_o \), the ADC input pole can be approximated by

\[
\omega_{3dB} = \left( \frac{1}{(2^n - 1) Z_o C_C} \right).
\]  
(3.2)

The front-end amplifier block must provide the full scale input \( A \) to the ADC with a bandwidth of \( \omega_{3dB} \geq \omega_{\text{clk}}/2 \). While the linearity (swing) condition may be met relatively easily (possibly using an on-chip silicon amplifier/THA), the large fan-out poses serious bandwidth limitations and scalability challenges. However, this type data distribution
Figure 3.2: Various data distribution methods in high-speed ADCs: (A) Direct driving of comparators or sub-ADC bank. (B) Passive data distribution in the form of a broadband power splitter tree. (D) Passive data distribution using a transmission line. (C) Active data distribution in the form of a data tree.
method does not consume power. An example of this type of data distribution is presented in [5] using an on-chip amplifier and [2] requiring the use of an external amplifier in order to meet the bandwidth requirements.

3.2.2. Passive Data Distribution Using Power Splitters

This method of passive data distribution divides the large input capacitance amongst multiple broadband power splitters (shown in Figure 3.2(B)). In order to minimize the depth of the distribution tree, the largest fan-out of $k$ must be chosen to meet the bandwidth requirements. For a fan out of $k$, the overall bandwidth can be approximated as

$$\omega_{3dB} = \left( \frac{1}{kZ_oC_C} \right).$$

(3.3)

This method can provide a significant bandwidth improvement over 3.2 without any power consumption overhead. However, due to the losses associated with broadband power splitters, the full-scale ADC input is increased to

$$A_{PS} = \Delta V (2^n - 1) 2^{[\log_k(2^n - 1)]}. \quad (3.4)$$

Equation 3.4 demonstrates the scalability challenges associated with this method. The required large input swing may impose the use of an external amplifier/THA whose power consumption should be included in the overall power budget. Two instances where this method has been used are presented in [3] and [4] where a single power divider is used at the input of the ADC.

3.2.3. Passive Data Distribution Using Transmission Lines

By using a transmission line along the comparator chain, the input capacitance can be absorbed by the transmission line while maintaining the characteristic impedance $Z_o = \sqrt{L_o/(C_o + C_C)}$ (shown in Figure 3.2(C)). This method has been demonstrated
in [6]. Ideally, the full-scale ADC input should be unaffected. Furthermore, in a flash converter, a similar transmission line needs be used on the clock path to align the clock and data signals.

While, as with other types of passive data distribution, this method does not consume power, it relies on precise matching between the delay of the input data path and the clock distribution network to avoid clock-to-data skew which becomes more challenging with increased conversion rates. Furthermore, due to high frequency losses of transmission lines, comparator-dependant input signal attenuation can further degrade the ADC performance. These limitations also hinder the scalability of this architecture as the length of the transmission line is proportional to the number of comparators.

### 3.2.4. Active Data Distribution

By using a tree of linear buffers to drive the comparators of the ADC, the large capacitance associated with the comparator bank is divided amongst buffers with reduced fan-out (shown in Figure 3.2(D)). Since the input capacitance of the data tree is significantly smaller than that of the entire comparator bank, a high gain front-end amplifier (in this design, a TIA) can then be used to drive the data tree with sufficient bandwidth. It is important to note however, that in contrast to purely passive data distribution, this active implementation consumes power.

A simple approach is to design the data tree using identical amplifiers with a fan-out of $k$, in which case the required depth of the tree is $d = \lceil \log_k(2^n - 1) \rceil$ for an $n$-bit flash converter. In order to minimize the number of cascaded stages, the highest fan-out which satisfies the overall required data tree bandwidth must be chosen. In practice, layout considerations and symmetry requirements of the data tree must also be considered. To achieve a data tree bandwidth of $\omega_{clk}/2$, each buffer with a fan-out of $k$ must have a 3 dB
The choice of each amplifier’s output impedance is not limited to $Z_o$, the designer may choose any output impedance as long as the bandwidth and linearity requirements are met.

If the amplifiers used in the data distribution have a gain of 1, the front-end amplifier/THA must provide the full-scale swing of $A$ to the distribution network. Similar to the direct driving method, the input swing requirements can be met using an on-chip silicon amplifier/THA. This work uses this type of data distribution [7].

### 3.2.5. Data Distribution Selection

Depending on the mm-wave ADC resolution, full scale input range and the need for a single-chip or multi-chip solution, a specific data distribution method may be most suitable. Equations 3.2, 3.3, 3.4 and 3.5 highlight the advantages and disadvantages of each method.

In the case of low resolution ($\leq 3$-bits) mm-wave ADCs, the direct driving method may be most feasible. Due to the low number of comparators and low full-scale input range, a single chip solution may be possible. In fact, the only reported single-chip mm-wave ADC (besides the work presented in this thesis) uses this method [5]. However, if the input capacitance is increased, for instance in the case of higher resolution ADCs, or a time-interleaved ADC with many sub-ADCs, external amplifiers may be required. For example, the time-interleaved ADC reported in [2] requires the use of an external BiCMOS amplifier.

To build a single-chip mm-wave ADC with $\geq 4$-bits of resolution, the problem of large input capacitance and large full-scale input signal must be solved simultaneously. The use of power-splitters can mitigate the bandwidth problem, but will introduce the
problem of large full-scale inputs. In fact, the only existing mm-wave ADCs which use power-splitter method require the use of external amplifiers \[3, 4\]. However, if a single-chip solution is desirable, the active data distribution presented in this thesis becomes an attractive solution despite its power consumption overhead. At the time of writing, the 35-GS/s ADC presented here remains the only single chip mm-wave ADC with 4-bits of resolution \[7\].

### 3.2.6. Impact of Data Distribution on THA

In an \(n\)-bit converter and under otherwise ideal conditions, a front-end amplifier and data tree bandwidth of \(\omega_{\text{clk}}/2\) can achieve an ENOB of better than \((n - 0.5)\)-bits up to the Nyquist frequency. However, sources of sampling uncertainty can further degrade the performance of the ADC. The maximum achievable SNR in the presence of sampling jitter is \[38\]

\[
SNR_{\text{MAX}} = 20 \log \left( \frac{1}{\omega_{\text{in}}\tau_j} \right) .
\]  

(3.6)

In flash ADCs, three major sources of timing uncertainty are clock skew, data skew and random jitter of the clock signal, but clock and data skew dominate if the ADC is driven by a low phase noise clock \[6\]. To combat clock and data skew in mm-wave ADCs, a THA may be employed. A full rate THA can be inserted after the front-end amplifier. For half the clock period the THA holds the input signal constant, ideally reducing the rate of change of the input signal to zero. During this time, the clocked comparators make decisions based on the held value. For the second half of the clock period, the THA tracks the input signal. In the presence of an ideal THA, the sampling uncertainty due to clock and data skew is eliminated.

The THA produces the desired zero slope regions in the time domain by introducing high frequency content to the spectrum of the input signal. A model of an ideal THA is shown in Figure 3.4. For half of each clock period, the THA output is identical to
Figure 3.3: This model is used to simulate the efficacy of a THA in the presence of data tree bandwidth limitations. Without sufficient data tree bandwidth, the THA is unable to combat the effects of the clock and data skew ($\tau_{d1-d15}$ and $\tau_{c1-c15}$, respectively).

Figure 3.4: A linear THA model used to analyze the impact of the analog front-end bandwidth on the ADC performance.
the input, $x(t)$. For the other half of each clock period, the output is a sample-and-held
version of the input, $u(t)$. The two are multiplexed by the clock to form the THA output,
$z(t)$. The bandwidth limitation of the ADC’s data tree is represented by the low pass
filter block with a 3dB bandwidth of $\omega_{3dB}$. Using the time-frequency domain duality
principal, the spectrum of the THA output, $Z(\omega)$, is related to that of the input signal,
$X(\omega)$ as follows:

$$Z(\omega) = V(\omega) + U(\omega) = \frac{1}{2} \sum_{k=-\infty}^{\infty} \left[ X(\omega + k\omega_{clk}) + Y(\omega + k\omega_{clk}) \right] \text{Sinc} \left( \frac{k\pi}{2} \right). \quad (3.7)$$

Furthermore, by substituting for $Y(\omega)$, equation (3.7) simplifies to

$$Z(\omega) = \frac{1}{2} \sum_{k=-\infty}^{\infty} \left[ X(\omega + k\omega_{clk}) + \text{Sinc} \left( \frac{(\omega + k\omega_{clk}) \pi}{\omega_{clk}} \right) \sum_{p=-\infty}^{\infty} X(\omega + (k+p)\omega_{clk}) \right] \times \text{Sinc} \left( \frac{k\pi}{2} \right). \quad (3.8)$$

The frequency content of $Z(\omega)$ extends beyond $\omega_{clk}$, however, since its spectrum is
shaped by a $\text{Sinc}$ function, more than 90% of its energy is captured within $\omega_{clk}$ and
thus the infinite summations can be approximated by taking only three terms in each:
$(p,k) = \{-1, 0, 1\}$. The expression for $Z(\omega)$ indicates that high frequency signal content
is present near the clock frequency of the ADC. This high frequency content must be
preserved otherwise the hold mode behavior of the THA is lost.

In practice, poles introduced at the output of the THA and comparator pre-amplifiers
attenuate this content, thus reducing the efficacy of the THA. Figure 3.3 shows the block
diagram of a 4-bit flash ADC used to model this phenomenon. In our behavioral simu-
lations, an ideal THA is followed by an $N$th order low-pass filter

$$T_{LPF}(s) = \left( \frac{1}{1 + \frac{s}{\omega_0}} \right)^N \quad (3.9)$$
Figure 3.5: Simulated histogram of a 4-bit Flash ADC in the presence of clock and data skew with standard deviations of $0.05T_{clk}$ and no THA. The performance of the ADC degrades with increasing input frequency.

... and with a 3 dB bandwidth of $\omega_{3dB}$. The low-pass filter represents the frequency response of the data tree. For a constant 3 dB bandwidth, simulation results are similar for $N = \{2, 3, 4\}$. The results shown here are for $N = 3$. Clock and data skew are modeled by delaying the input and clock of each comparator by independent Gaussian distributed random variables ($\tau_{d1-d15}$ and $\tau_{c1-c15}$ in Figure 3.3) with zero mean and standard deviations of $0.05T_{clk}$ where $T_{clk}$ is the clock period of the ADC. Behavioral simulations of one hundred 4-bit ADCs with a data tree bandwidth of $\omega_{clk}/2$ and no THA show that with increasing input frequency, the average ENOB decreases due to skew (Figure 3.5). In this simulation, the input frequencies are chosen to demonstrate the limitation on achieving 3.5-bits, 3-bits and better than 2-bits of ENOB. For instance, while 95% of all simulated ADCs achieve an ENOB of better than 3.5 bits for an input frequency of $0.08\omega_{clk}$, only 3% of the simulated ADCs achieve the same ENOB for an input frequency of $0.22\omega_{clk}$.

An ideal THA followed by a data tree with infinite bandwidth can fully eliminate the
effect of the skew in more than 99.7% of all simulated ADCs resulting in an ENOB of 4-bits for any input frequency. However, an ideal THA followed by a finite-bandwidth data tree can be expected to have performance somewhere between these two extremes. Therefore we can investigate the data tree bandwidth requirements such that a THA can successfully combat skew.

Figure 3.6 shows the input to the comparator bank for a sinusoidal input signal \( x(t) = \sin(\omega_{in} t) \) with \( \omega_{in} = 0.14 \omega_{clk} \) where \( \omega_{clk} \) is the clock frequency of ADC in rad/s. With a very high bandwidth data tree, the slope during the hold time is zero, providing immunity against skew. However, as the bandwidth is reduced, the input to the comparator bank looks increasingly like a sinusoid at the ADC input. Hence, the THA provides little improvement in performance unless the bandwidth of the following stages is high enough.

The simulated ADC ENOB profile is shown in Figure 3.7. It shows the minimum ENOB achieved by the best 95% of all simulated ADCs. With a data tree bandwidth of \( \omega_{clk}/2 \), no THA, and in the presence of skew, an ENOB better than 3-bits can only be expected for input frequencies up to \( 0.18 \omega_{clk} \). With the THA enabled and the data tree bandwidth extended to beyond \( \omega_{clk} \), nearly all simulated ADCs achieved 4-bits of ENOB. However, it can be observed that if the bandwidth of the data tree is close to the Nyquist frequency, \( \omega_{3dB} = \omega_{clk}/2 \), implementing a THA gives little improvement in performance. Therefore, if a THA is to be employed effectively, the bandwidth of the data tree must exceed the clock frequency; typically a very difficult design specification for mm-wave conversion rates.

### 3.2.7. THA and Data Tree Breakout

To investigate the impact of the data tree on the ADC performance, a breakout of the TIA, THA and data tree has been fabricated. Figure 3.8 shows the block diagram of the implemented breakout. This breakout allows for the characterization of the TIA, THA, and the BiCMOS data tree as well as an adjustable threshold comparator. The front-end
Figure 3.6: Impact of the data tree bandwidth on the THA waveform: (A) $\omega_{3dB} = 2\omega_{clk}$, (B) $\omega_{3dB} = 0.8\omega_{clk}$, (C) $\omega_{3dB} = 0.5\omega_{clk}$. The input signal frequency is $\omega_{3dB} = 0.14\omega_{clk}$. The hold mode behavior of the THA is mostly lost if the data tree bandwidth is limited.
Figure 3.7: Simulated ENOB of a 4-bit Flash ADC in the presence of clock and data skew with standard deviations of 0.05\(T_{clk}\). It can be observed that if the data tree bandwidth is limited to 0.5\(\omega_{clk}\), a THA offers little improvement in performance.
Figure 3.8: Block diagram of the fabricated breakout. This breakout allows for the characterization of the TIA, THA, and the BiCMOS data tree. The THA can be disabled for bandwidth measurement.
amplifier is a low noise broadband TIA, followed by a buffering stage which drives the THA. The buffering stage provides biasing current to the THA and improves the single-ended to differential conversion when the TIA is driven single-ended [32]. Unlike [2,3,5], the implemented THA needs to only drive a small output capacitance. A data tree of BiCMOS cascode buffers is employed. The tree is intended for use in a 4-bit flash converter. Each stage has a fan-out of two and, to save area in this breakout, only one branch of the tree is implemented with the other branches terminated on chip. A BiCMOS 50Ω driver allows the analog output of the tree to be taken off-chip. A single bit thermometer code output generated by the adjustable comparator is also taken off-chip.

3.2.7.1. Transimpedance Amplifier (TIA)

A transimpedance amplifier can be designed to provide simultaneous noise and impedance matching without the need for 50Ω matching resistors [28]. Figure 3.9 shows the schematic of the TIA included in this ADC. The TIA is employed as the front-end amplifier and is driven as a voltage amplifier. In order to achieve the desired 12 dB of differential gain, the TIA is implemented using SiGe HBTs only. In this design, we have added 12Ω degeneration resistors to improve the linearity of the TIA without significantly degrading the noise figure. Diode connected transistors $Q_3$ and $Q_4$ prevent the breakdown of $Q_1$ and $Q_2$. A switched emitter-follower THA is also implemented following the design methodology described in [32]. The TIA has a simulated bandwidth of better than 20 GHz. This simulation is also supported by the measurement results presented in [32].

3.2.7.2. BiCMOS Data Tree

Each block in the BiCMOS data tree consists of a MOS-HBT cascode amplifier shown in Figure 3.10. The BiCMOS cascode amplifiers have a tail current of 8 mA with the MOS-FETs biased near the peak $f_T$ current density ($J_{f_T} = 0.3 mA/\mu m$). Although smaller nMOS transistors are more susceptible to mismatches, biasing at $J_{f_T}$ was necessary to achieve a combination of high bandwidth and high linearity. In this design, the layout of
Figure 3.9: Schematic of the implemented fully bipolar TIA. The TIA has 12 dB of simulated differential gain and drives the THA.
Figure 3.10: Schematic of the MOS-HBT cascode differential pair used in the data tree. CMOS input transistors offer high linearity while the bipolar cascode transistors offer low output capacitance.
each data tree block is optimized by using dummy resistors, transistors and interdigitating and merging the nMOS transistors with common sources a single well to minimize mismatches. The 0.18µm nMOS pair offers a simulated differential input compression point of 1.5V_{PP} at low frequencies. The HBT common-base transistors create low impedance cascode nodes with high output slew rate and bandwidth. The data tree is implemented completely symmetrically to minimize systematic sources of skew, comparator dependent signal attenuation and offsets. Series-shunt inductive peaking is also used throughout the data tree. Series peaking is provided by the interconnect between the data tree blocks. The simulated bandwidth of a single block with a fan-out of two is 45 GHz and the simulated gain of the entire data tree is 0 dB.

3.2.8. THA and Data Tree Measurement Results

The die photo of the fabricated breakout is shown in Figure 3.11. All measurements on the breakout are done single-ended and using wafer probing. The measured and post-extraction simulated small-signal characteristics of the TIA, THA and the data tree are shown in Figure 3.12. The BiCMOS output driver used to drive off-chip 50Ω loads has a simulated loss of 6 dB. The combination of the TIA, THA and the data tree has a 3 dB bandwidth of 16 GHz. The TIA input return loss is better than -10 dB up to 20 GHz. Although a similar THA in the same technology demonstrated a bandwidth exceeding 40 GHz [32], the bandwidth of this design is limited by the many buffer stages required to drive the capacitive load of the comparator bank. While each buffer stage has a bandwidth of 45 GHz on its own, cascading four of them along with the THA and TIA greatly reduces the overall bandwidth down to 16 GHz. Figure 3.13 shows the measured time domain outputs of the breakout for input frequency \( \omega_{in} = 5 GHz \) and clock frequency \( \omega_{clk} = 35 GHz \) without [Figure 3.13(A)] and with [Figure 3.13(B)] the THA enabled. It can be observed that the hold mode behavior of the THA has nearly disappeared in Figure 3.13(B) due to the insufficient bandwidth of the data tree.
A significant disadvantage of enabling the switched emitter-follower (SEF) THA is that it introduces nonlinearities which can actually degrade the SNDR of the ADC. There are two major sources for these non-linearities. First source is the nonlinear modulation of the base-emitter junction voltage of the SEF during the charging and discharging of the hold capacitance [32]. The second source is the mixing of the harmonics of the THA input signal (2nd, 3rd and 4th depending on the input frequency) with the clock signal which folds the harmonics back into the Nyquist bandwidth of the converter. Since ADCs are typically tested using a single tone input, the folded harmonics are only present while the THA is active. Figure 3.14 shows the measured output spectrum of a fabricated breakout, with and without the THA activated. To disable the THA, it is switched to track-mode only. In the case where the THA is disabled [Figure 3.14(A)], the second and third harmonics at the output are mostly due to the nonlinearities of the TIA. However, once the THA is activated [Figure 3.14(B)], mixed products of the input signal harmonics, as well as the clock feed-through signal, are present at the output. The third order harmonic also increases in magnitude. Depending on the resolution of the ADC and the severity of the clock jitter and skew, the benefit of a THA can outweigh the nonlinearities introduced
Figure 3.12: Small-signal characteristics of the fabricated breakout. Measured single-ended s-parameters and simulated single-ended gain of the TIA, THA, data tree and overall s-parameters of the breakout after extraction are shown. The combination of the TIA, THA and the data tree has a measured small-signal 3 dB bandwidth of 16 GHz.
Figure 3.13: Measured output of the fabricated breakout: (A) THA OFF; (B) THA ON. Due to the limited data tree bandwidth, the hold mode behavior of the THA amplifier is nearly eliminated.
by it. However, in the case where the data distribution bandwidth is insufficient for proper operation of the THA, it is better to omit the implementation of the THA altogether. In this work, since the data tree bandwidth is insufficient for proper operation of the THA at the targeted clock frequencies, minimizing clock and data skew is of great importance. This criterion was strictly observed by designing symmetric and well-matched clock and data distribution trees.

Measurements of the single thermometer code output show the operation of the adjustable threshold comparator. For instance, Figure 3.15 shows the measured output when a 312.5-MHz sinusoid input signal and a 40-GHz clock signal are used. By adjusting the comparator threshold, a different ratio of ones and zeros present themselves at the output. By increasing the input frequency to 10 GHz, while maintaining a 40-GHz clock signal, a 4-bit periodic pattern can be measured. Figure 3.16 shows the 4-bit pattern for two different threshold settings. When the threshold is set to zero-level, a ”1100” pattern is observed (Figure 3.16(A)). Similarly, by adjusting the threshold to quarter-level, a ”1110” pattern is measured (Figure 3.16(B)). A full ADC can be constructed by using an array of such comparators each with a different threshold.

3.3. The 4-bit FLASH ADC-DAC CHAIN

The fabricated ADC (presented in [7]) uses the same TIA and BiCMOS data tree as the breakout and merges the buffer and THA. The block diagram of the implemented ADC-DAC chain is shown in Figure 3.17. The data tree of BiCMOS cascode buffers distributes the THA output to the entire comparator bank. Each of the 15 comparators, required for 4-bit operation, consists of an offset amplifier, a high gain preamplifier, a flip-flop and a differential pair. A 16th dummy comparator is also employed to maintain the symmetry of the data tree and provide a single thermometer code output off-chip for testing. The comparator bank produces a 15-level thermometer code. For testing, a 15-level thermometer code DAC has been implemented on chip. No decoder logic or
Figure 3.14: Measured spectra of the breakout output showing the nonlinearity impact of the THA: (A) THA disabled; (B) THA enabled showing intermodulation products of the input and clock signals in the output spectrum.
bubble correction circuitry has been implemented. The clock tree comprises 17 series-shunt inductively peaked bipolar differential pairs with a fan-out of two per stage; the final stages drive four latches each. A separate clock path drives the THA. A two-stage tuneable delay cell has been implemented on chip to align the THA clock with the comparator clock. The active delay cells consist of phase interpolating blocks between fast and slow paths [39]. The THA can be disabled by forcing it into track mode only via external controls.

3.3.1. ADC-DAC Circuit Description

3.3.1.1. Merged BiCMOS Cascode and THA

In the design of the ADC, a MOS-HBT BiCMOS cascode buffer is merged with the switched emitter-follower THA as shown in the schematics on Figure 3.18. The tail current of the THA is 6 mA per side. Compared with [32], the total capacitance at node (A) is reduced by connecting the collector of transistor \( Q_9 \) to the low impedance cascode node (B) instead. This approach eliminates the need for level shifting diodes, and further
Figure 3.16: Measured single-bit thermometer output for two different comparator threshold settings; (A) zero-level threshold, (B) quarter-level threshold. The input signal is at 10 GHz with a comparator clock frequency of 40 GHz.
Figure 3.17: Block diagram of the implemented 35-GS/s, 4-bit, Flash ADC-DAC. The ADC uses data and clock distribution to drive the comparator bank.
Figure 3.18: Schematic of the merged MOS-HBT BiCMOS cascode and THA. The total capacitance at node (A) is reduced by connecting the collector of transistor $Q_9$ to the low impedance cascode node (B) instead.
isolates the hold capacitor, $C_H$, from the clock signal.

### 3.3.1.2. Comparator Design

The comparator block diagram is shown in Figure 3.17. Offset currents were used in this design to establish the quantization levels without a resistor ladder [5]. Each of the comparators is preceded by a differential BiCMOS cascode offset amplifier illustrated in Figure 3.19. The unit quantization current in this design is $120\mu A$, which in combination with the $170\Omega$ load resistors, produces a differential quantization level of $40.8$ mV at the inputs of the comparators. Furthermore, since the simulated gain of the data tree is $0$ dB, one LSB referred back to the input of the data tree is also $40.8$ mV. The offset currents are drawn from the low impedance cascode nodes, (A) and (B), to mitigate their impact on bandwidth. In order to maintain equal delay, the total capacitance at nodes (A) and (B) is kept constant for all comparators. This is accomplished by adding dummy off current sources where needed. Offset mismatches introduced by the data tree also present themselves at the output of this block. These types of mismatches may be corrected by calibration of the offset currents. Similarly, data tree gain variations may also be corrected through this type of calibration. However, in this design, due to the large quantization levels and the low data tree gain, no calibration technique has been deemed necessary. Fully bipolar cascode pre-amplifiers with $15$ dB of differential gain and $25$ GHz bandwidth follow the BiCMOS offset amplifiers. The comparator’s metastability window, which refers to the smallest input voltage required to switch a comparators state, is reduced by the addition of these pre-amplifiers. The preamplifiers also isolate the offset stages from flip-flop kickback. Each flip-flop has two latches with $4$ mA tail current and emitter-followers both on the clock and data paths. Inverting stages with $4$ mA tail current follow every flip-flop to eliminate the latch induced clock feedthrough and drive the DAC. The combination of the pre-amplifier and flip-flop has a simulated metastability window of better than $20$ mV at $35$ Gbps.
Figure 3.19: Schematic of MOS-HBT BiCMOS cascode offset amplifier. The offset currents are drawn from the low impedance cascode nodes to mitigate their impact on bandwidth.
3.3.1.3. Thermometer Code DAC

The comparator bank produces a 15-level thermometer code. In implementations where thermometer- to-binary conversion is required, bubble error removal can be applied in the form of three-input NAND gates at the cost of increased power consumption [40]. Another method to remove bubble errors that does not increase the power consumption is by adding extra voting transistors to the latches as proposed by [41]. However, in this approach the added capacitive loading in the latches may reduce the maximum conversion frequency.

For testing purposes, a 15-level thermometer code DAC has been implemented on this chip. Its schematic is reproduced in Figure 3.20. It consists of a differential, double bipolar cascode. Each thermometer code bit drives one CML inverter with a tail of current of 4 mA. The output currents of these CML inverters are summed in the first cascode nodes in groups of four. The remaining four output currents are summed in the second cascode nodes in pairs. Finally the remaining two output currents are summed at the on-chip 50Ω resistors. Simulations have shown that, in this technology and with this DAC topology, buffers smaller than 4 mA do not have sufficient bandwidth to operate at 35 Gbps. Each thermometer code produces a differential swing of 200 $mV_{PP}$ for a total differential output swing of 3 $V_{PP}$. The emitter-followers driving the inverting stages of the DAC operate from 3.3 V while the double cascodes require a 5 V power supply (Figure 3.20).

3.3.1.4. Clock Distribution Network

The challenging task of distributing the 35 GHz clock to 32 latches (16 masterslave comparators) is accomplished by a fully bipolar clock distribution tree. The tree consists of series-shunt inductively peaked stages with a fan-out of two per stage; the final stages drive four latches each. Each stage has an 8 mA tail current with 500 $mV_{PP}$ of differential output swing. A separate clock path drives the THA. A two-stage tuneable delay cell is
Figure 3.20: Schematic of the implemented 15-level thermometer code DAC. Thermometer code inputs (D0-D15) are summed together at different nodes of the double cascode DAC.
also implemented on chip [39]. It is used to align the clock and data signals and consists of phase interpolating blocks between fast and slow paths.

### 3.3.2. ADC-DAC Measurement Results

The ADC-DAC chain has been implemented in Jazz Semiconductors SBC18HX, 0.18 \( \mu m \) SiGe BiCMOS HBT technology. The process offers an \( f_T \) of 160 GHz/50 GHz for bipolar and nMOS devices respectively. The ADC operates from a 3.3 V supply while the DAC requires a 5 V supply. The chip occupies an area of 2.5 mm x 3.2 mm and the ADC-DAC chain consumes a total of 5 W (4.5 W for the ADC). The data and clock tree require 0.65 W and 1 W respectively. The comparator bank consumes 2.4 W while the TIA and THA amplifier use a total of 0.23 W. The adjustable delay cells and biasing circuitry need a total of 0.22 W. Figure 3.21 shows the chip micrograph. All measurements are performed on wafer under nominal power supply voltages and at room temperature.

The clock output provides a means to measure the delay cell’s performance. Figure 3.22 shows that the delay cell achieves a 13.5 ps of delay at 40 GHz. The INL and DNL were measured differentially by applying a full scale DC sweep at the input and capturing the DAC outputs while clocking the ADC at 35 GHz. Figure 3.23 shows that the INL and DNL of the ADC are below 0.5 LSB for all analog input levels. The measured INL and DNL of the DAC are better than 0.1 LSB for all digital codes (Figure 3.24).

All dynamic measurements are performed singled-ended by wafer probing only one of the differential pair outputs. The ADC was tested with the same input signal power for all input frequencies. The spectral content of the DAC output is displayed on an Agilent spectrum analyzer. After accounting for the \( \sin(x)/x \) DAC roll-off and high frequency losses in the wafer probes and cables, the SFDR and ENOB of the ADC are plotted in Figure 3.25 for both THA operation modes. At low frequencies the ENOB is 3.7 bits (24 dB SNDR). Figure 3.26 shows a sample output spectrum of a 7.98 GHz input signal with the THA disabled. The effective resolution bandwidth is about 8 GHz.
Figure 3.21: Chip micrograph of the ADC-DAC, die area is $2.5 \times 3.2 \text{mm}^2$.

Figure 3.22: Measured delay cell phase adjustment at 40 GHz.
Figure 3.23: Measured ADC INL and DNL with a 35-GHz clock frequency.

Figure 3.24: Measured DAC INL and DNL with a 35-GHz clock frequency.
irrespective whether the THA is enabled or not. However, beyond 8 GHz the SFDR and ENOB degrade more rapidly with the THA activated. This is the result of the input signal harmonics mixing with the clock signal and showing up at low frequencies, within the bandwidth of the data tree. The third-order harmonics of the input signal is also degraded due to the SEF THA nonlinearities. Therefore, the THA actually degrades the performance of this ADC. In this design, with 4 bits of precision, careful layout is sufficient to prevent skew from becoming a performance limitation.

### 3.4. Conclusions

A 4-bit, 35-GS/s flash ADC-DAC chain was demonstrated in this chapter. The ADC uses active clock and data trees to distribute the input and the clock to the comparators. Because both clock and data are symmetrically distributed, this implementation minimizes timing skew achieving 3.7 ENOB with 8 GHz effective resolution bandwidth and 3 ENOB up to 11 GHz without a full-rate THA. Table 3.1 shows a comparison of the state-of-the-art ADCs.
Figure 3.25: Measured ENOB and SFDR versus input frequency for a 35-GHz clock frequency. The THA degrades the performance of the ADC for input frequencies above 8 GHz.

Figure 3.26: Output spectrum of a 7.98 GHz input signal sampled at 35 GHz after accounting for the sin(x)/x DAC roll-off and setup losses. The THA is disabled.
### Table 3.1: Table of comparison of state of the art published ADCs.

<table>
<thead>
<tr>
<th>Year</th>
<th>Resolution (bits)</th>
<th>Sample Rate (GS/s)</th>
<th>SFDR ($f_m \leq 1, GHz$)</th>
<th>SNDR ($f_m \leq 1, GHz$)</th>
<th>SFDR ($f_m = 11, GHz$)</th>
<th>SNDR ($f_m = 11, GHz$)</th>
<th>ENOB (bits)</th>
<th>ERBW (GHz)</th>
<th>Full Scale (V&lt;sub&gt;pp&lt;/sub&gt; Diff.)</th>
<th>Power (W)</th>
<th>Architecture</th>
<th>Technology ($f_T$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>4</td>
<td>35</td>
<td>28.5</td>
<td>24.1</td>
<td>27.3</td>
<td>19.8</td>
<td>3.7</td>
<td>8</td>
<td>0.24</td>
<td>4.5</td>
<td>Flash</td>
<td>SiGe (160 GHz)</td>
</tr>
<tr>
<td>2004</td>
<td>3</td>
<td>40</td>
<td>28</td>
<td>18.6</td>
<td>25</td>
<td>N/A</td>
<td>2.8</td>
<td>N/A</td>
<td>N/A</td>
<td>3.8</td>
<td>Flash</td>
<td>SiGe (210 GHz)</td>
</tr>
<tr>
<td>2003</td>
<td>8</td>
<td>20</td>
<td>N/A</td>
<td>40.8</td>
<td>N/A</td>
<td>N/A</td>
<td>6.5</td>
<td>2</td>
<td>0.5</td>
<td>9.0</td>
<td>Time Intrvl.</td>
<td>SiGe BiCMOS (210 GHz)</td>
</tr>
<tr>
<td>2006</td>
<td>5</td>
<td>22</td>
<td>35</td>
<td>30</td>
<td>&lt; 20</td>
<td>&lt; 15</td>
<td>4.7</td>
<td>5</td>
<td>1.2</td>
<td>3.0</td>
<td>Flash</td>
<td>CMOS (150 GHz)</td>
</tr>
<tr>
<td>2004</td>
<td>3</td>
<td>24</td>
<td>24.6</td>
<td>19.5</td>
<td>23.3</td>
<td>15.6</td>
<td>4.7</td>
<td>6</td>
<td>1.2</td>
<td>3.84</td>
<td>Flash</td>
<td>BiCMOS (150 GHz)</td>
</tr>
<tr>
<td>2008</td>
<td>6</td>
<td>24</td>
<td>44</td>
<td>33.3</td>
<td>35.3</td>
<td>23.1</td>
<td>5.25</td>
<td>8</td>
<td>N/A</td>
<td>1.2</td>
<td>Time Intrvl.</td>
<td>CMOS (150 GHz)</td>
</tr>
<tr>
<td>2009</td>
<td>5</td>
<td>20</td>
<td>N/A</td>
<td>28.8</td>
<td>N/A</td>
<td>22.2</td>
<td>4.5</td>
<td>8</td>
<td>N/A</td>
<td>4.8</td>
<td>Flash</td>
<td>BiCMOS (200 GHz)</td>
</tr>
<tr>
<td>2010</td>
<td>6</td>
<td>40</td>
<td>&lt; -50</td>
<td>34.8</td>
<td>&lt; -35</td>
<td>31.3</td>
<td>5.5</td>
<td>8</td>
<td>1.2</td>
<td>1.5</td>
<td>Time Intrvl.</td>
<td>CMOS (200 GHz)</td>
</tr>
</tbody>
</table>

Sample Rate: $f_m$ is the input frequency.

SFDR: Signal-to-Noise and Distortion Ratio.

SNDR: Signal-to-Noise Ratio.

ENOB: Effective Number of Bits.

ERBW: Effective Resolution Bandwidth.

Full Scale: $V_{pp}$ is the peak-to-peak voltage range.

Power: $W$ is the power consumption.
Toward CMOS
Millimeter-Wave Data Converters

4.1. Introduction

Baud-rate wireline transceivers and mm-wave flash ADCs require full-rate flip-flops, clock distribution and various low-noise analog front-ends. Furthermore, the operational frequency of CMOS THAs has lagged those of SiGe HBTs despite the fact that nano-scale CMOS $f_T$ and $f_{MAX}$ is on par with production SiGe technologies. The goal of this section is to explore the potential of CMOS technology for implementing fundamental building blocks for mm-wave ADCs and wireline transceivers. This chapter describes the design of two record-breaking circuits: A 30-GS/s THA in 0.13-$\mu$m CMOS technology followed by an 81-Gb/s TIALA-Retimer in 65nm CMOS technology.

4.2. A 30-GS/s THA in 0.13-$\mu$m CMOS Technology

4.2.1. Introduction and State of the Art

Traditionally, high speed THAs implemented in III-V and SiGe technologies either use high speed sampling diode bridge or switched emitter follower topologies (Figure 4.1). The sampling diode bridge requires high-speed diodes and large power supplies and thus isn’t a viable solution for CMOS technologies [19]. For CMOS implementations, a clocked series CMOS switch is most common (Figure 4.2). However, the series resistance associated with the CMOS switch limits the bandwidth of this topology [43]. This section
Chapter 4. Toward CMOS Millimeter-Wave Data Converters

presents the design methodology and measurement results of a 30-GS/sec track and hold amplifier in CMOS. This work makes use of a switched source follower (SSF) to achieve a 30-GS/s clock rate similar to the bipolar switched emitter follower (SEF) (Figures 4.1 and 4.2).

All previously-reported THAs operating above 10 GHz have employed bipolar devices in the switched emitter follower topology, first proposed in [25]. The fastest THA reported to date is a distributed switched-emitter follower, 50-GS/s THA implemented in SiGe HBT technology and consuming 640 mW from a 3.3-V supply [24]. The fastest CMOS THA uses a switched series transistor topology in a 0.18-µm technology for a maximum clock frequency of 10 GHz from a 1.8-V supply [33]. In this design, by systematically biasing transistors in the signal and clock paths for maximum speed, record breaking sampling rate and bandwidth are achieved. Unlike all other CMOS THAs, this circuit employs a transimpedance amplifier (TIA) input to maximize bandwidth and minimize noise. The speed and performance of this CMOS THA are comparable to those of the fastest reported SiGe HBT and InP THAs [22, 24, 26, 32] but are obtained with half the supply voltage and thus lower power due to the lower CMOS power supply voltage requirements.

4.2.2. Track and Hold Amplifier Design

Figure 4.3 shows the block diagram of the track and hold amplifier. At the front-end is a TIA followed by a pair of common source (CS) amplifiers. The CS blocks drive two consecutive differential pair stages which in turn drive the track and hold (T/H) block differentially. The output drivers are single-ended common source amplifiers with 50Ω resistors. The data path of the THA is designed for maximum dynamic range. This is realized by choosing a low noise broadband input stage and biasing the transistors in the data path at 0.25 mA/µm for a good compromise between bandwidth, low-noise and high linearity. A clock distribution network provides the differential clock signal to the
Figure 4.1: (Left) Example of a sampling diode bridge. (Right) Example of a switched emitter follower (SEF).

Figure 4.2: (Left) Example of a clocked series CMOS switch. (Right) Example of a switched source follower (SEF).
T/H block. It consists of the same TIA-CS front-end as the data path, followed by three CML inverters. These stages operate in digital mode and are biased at 0.15 mA/µm for maximum switching speed. The circuit is powered from a 1.8-V supply and consumes 270 mW. The T/H block and the clock distribution network consume 12 mA and 70 mA respectively, while the remaining 68 mA of current is drawn by the input and output blocks.

4.2.3. Transimpedance Amplifier (TIA)

Instead of using 50Ω-terminated differential pairs, this design employs a low noise TIA input stage (Figure 4.4). A TIA provides simultaneous noise and signal matching without the need for 50Ω matching resistors. Noise matching is achieved by sizing the input transistors ($Q_{1-2}$) to produce an optimum source impedance of 50Ω [29]. The input impedance is set by the feedback resistor $R_F$. Active PMOS loads ($Q_{3-4}$) are used to increase the open loop gain and maximize the linearity of the TIA. At DC, transistors $Q_1$ and $Q_2$ are diode connected and thus the output of the TIA biases the transistors $Q_{5-6}$ [44]. The input-referred noise power spectral density of the cascade of the TIA and the differential pair stages obtained from simulation and integrated up to 30 GHz is 0.5 mV$_{\text{rms}}$.

4.2.4. Track and Hold Stage

Fig. 4.3 shows the schematic diagram of the THA with a differential pair input and output driver. In track mode, $Clk_P$ is high, $Q_{SF}$ acts as a source follower and the output follows the input signal. In hold mode, $Clk_N$ is high and the tail current $I_T$ flows through the loads of the differential pair, $R_L$. The value of $R_L$ is chosen such that the voltage drop $I_TR_L$ turns off transistor $Q_{SF}$ and provides good isolation between the input and the output. The power supply voltage required by this design is dictated by $I_{Diff}R_L + V_{GS-QSF} + V_{GS-QDRV}$. Taking advantage of the triple well option in this process, the bulk and source of $Q_{SF}$ are shorted together to minimize $V_{GS-QSF}$ and allow
Figure 4.3: Block diagram of the implemented THA.

Figure 4.4: Schematic of the TIA and CS stages. All MOSFETs have a drawn length of 0.13 μm.
for a 600-mV \( P-P \) signal swing at the output of the source follower. Even though a 1.8-V supply is required, the voltage drop across individual transistors does not exceed 1.2 V. The hold capacitance, \( C_H = 250 \, fF \), includes the parasitic capacitances at that node and a 50fF MIM capacitor. The value of \( C_{fth} \) is chosen to match \( C_{GS-QSF} \) and to cancel hold mode signal feedthrough.

### 4.2.5. Clock Distribution Network

The clock path converts a single-ended 30-GHz clock input to a differential signal with 750mV \( P-P \) swing per side at the two switching pairs (\( Q_T \) and \( Q_H \)) in the T/H block. The blocks in the clock path operate at 30 GHz, the highest operating frequency of this circuit. It consists of a TIA stage (identical to Figure 4.4) followed by three CML inverters. The schematic of the final three CML inverting stages is shown in Figure 4.6. In order to ensure that every stage is fully switched, the gain of the inverters in the clock path is designed to exceed \( \sqrt{2} \).

### 4.2.6. Fabrication and Testing

The chip was fabricated in a 0.13\( \mu \)m CMOS technology. The chip area is 1 mm\(^2\) and the die photo is shown in Figure 4.7. All measurements were conducted on-wafer with single-ended input and clock signal. The simulated and measured S-parameters are shown in Figure 4.8. The input and output return loss are better than -10 dB up to 35 GHz. \( S_{21} \) has a bandwidth of 7 GHz when the circuit is operated in track mode. Since the front end of the data and clock paths are identical, the input return loss of the data path is also representative of the clock input return loss which is less than -15 dB from 22 GHz to 32 GHz. Figure 4.9 illustrates the two single ended outputs for a -12 dBm, 7.5-GHz input signal sampled at 30 GHz. Signal droop rate was measured at low clock frequencies to be less than 10 mV/ns. For a 30-GHz sampling frequency, this translates to a droop of only 0.2 mV per held value.

The input and output compression points of the circuit were measured from 1 GHz to
Figure 4.5: Circuit diagram of the differential pair, SSF and output driver blocks. All MOSFETs have a drawn gate length of 0.13µm.

Figure 4.6: Circuit diagram of the differential pair, THA and output driver blocks. All MOSFETs have a drawn length of 0.13µm.
Figure 4.7: Die photo of the fabricated THA. The chip area is 1mm$^2$.

Figure 4.8: Simulated and measured single-ended THA input return loss ($S_{11}$), output return loss ($S_{22}$) and transmission ($S_{21}$).
14 GHz in 1-GHz steps and are illustrated in Figure 4.10. Figure 4.11 shows the measured IIP3 and OIP3 as a function of frequency. The OIP3 plot indicates a circuit bandwidth of 7 GHz, in close agreement with the $S_{21}$ measurement. Simulations show that the hold node is the bandwidth-limiting node of this circuit. The measured SFDR and THD are illustrated in Fig. 4.12. SFDR is calculated from the measured spectra and simulated noise floor integrated over a 30-GHz bandwidth, (evaluated to be -53 dBm). The THA has an SFDR of 40 dB at 1 GHz with 7 GHz of bandwidth.

### 4.2.7. Conclusions

A 30-GS/sec CMOS THA was designed and fabricated in a 0.13-µm CMOS process. Following a systematic design procedure, and combining a low noise front-end with signal feedthrough cancelation and power supply voltage reduction techniques, has yielded the highest sampling frequency THA in CMOS. With a THD of better than -29 dB and 7 GHz of bandwidth, this THA is a potential contender for the front end of a 30 GS/sec flash ADC. Such an ADC can be used in over-sampling DSP based receivers for 10 Gb/s applications. Table 4.1 compares the state of the art published THAs.
Figure 4.10: Measured input compression point ($P_{i1dB}$) and output compression point ($P_{o1dB}$) versus input frequency with 30-GHz clock frequency.

Figure 4.11: Measured IIP3 and OIP3 versus input frequency with 30-GHz clock frequency.
Figure 4.12: Measured SFDR and THD versus frequency with 30-GHz clock frequency. The measured THD is shown for signals with -12 dBm input power.

4.3. An 81Gb/s, 1.2V TIALA-Retimer in Standard 65nm CMOS

4.3.1. Introduction

As an essential clocked-system component, high-speed CML latches have been extensively studied and implemented in a variety of technologies. High-speed CML latches in a D-type flip-flop (D-FF) configuration can be used in flash ADCs, wireline transceivers and equalizers. T-type flip-flops (T-FFs) can be used as frequency dividers in PLLs and high-speed counters. Latches may also serve as a benchmark circuit to demonstrate the high-speed and low-power capabilities of a technology [11–16,45–47].

CMOS, with its $f_T$ and $f_{MAX}$ values exceeding those of all other commercial semiconductor technologies, is of particular interest for emerging wireline applications with up to 110Gb/s serial data rates and the next generation low-power ADCs with sampling rates exceeding 40GHz. The purpose of this section is to demonstrate the mm-wave
capabilities of standard 65nm CMOS process by implementing an 81Gb/s, 1.2V TIALA-Retimer. The text describes the design methodology and challenges associated with achieving record performance.

4.3.1.1. State of the Art

Figure 4.13 shows the reported record operating speeds for multiplexers (MUXs), T-FFs and D-FFs implemented in various technologies. Although III-V implementations of multiplexers, T-FFs, and D-FFs have achieved operating speeds of 165Gb/s, 150GHz and 80Gb/s respectively, the CMOS equivalent circuits continue to lag behind. Furthermore, compared to the fasted multiplexer, the fasted D-FF operates at less than half the bit-rate. This is due to the differences in the timing requirement and the spectral content of the processed data in each case. Figure 4.14 illustrates operational time charts for each circuit.

A D-FF processes broadband random data at every clock period. It therefore requires a full-rate clock at the frequency equal to the input data bit-rate. Since a random data-pattern has most of its spectral content up to half the bit-rate frequency, date buffers with wide bandwidth as well as fast switching clock transistors are required to realize D-FFs. In contrast, a MUX requires only a half-rate clock and data at its input and thus has more relaxed timing constraints. A T-FF does require fast switching clock transistors, however it does not process random broadband data patterns; the input signal is fed back from the output in a positive feedback configuration. Therefore a D-FF, compared to MUXs and T-FFs, poses more design challenges and this is the primary reason why D-FF speeds always lag those of MUXs and T-FF in every technology node.

Static frequency dividers up to 100GHz have been reported in SOI CMOS, however retiming flip-flop speeds continues to lag those reported in SiGe BiCMOS and III-V technologies. For example, full-rate retimers have been demonstrated up to 80Gb/s from -5.7V supply in InP HEMT technology, at 48Gb/s from 2.5V in SiGe BiCMOS and at 40Gb/s from a 1.2V supply in 90nm CMOS. The proceeding sections show
Figure 4.13: Reported record operating speeds for multiplexers ([8][10]), T-FFs ([11][13]) and D-FFs ([14][16]).

Figure 4.14: Operational time chart for multiplexers, T-FFs and D-FFs.
for the first time that by adapting the circuit topologies to low-voltage nanoscale CMOS technology and by appropriately sizing, biasing and laying out the circuit cells, a record 81Gb/s transimpedance-limiting-amplifier (TIALA) and retimer circuit can be fabricated in a 65nm general purpose (GP) and low power (GPLP) CMOS process with multiple threshold voltage devices.

4.3.2. Circuit Design

The circuit block diagram of the TIALA-Retimer is shown in Figure 4.15 and Figure 4.16 illustrates the functional behavior of the system. The data path comprises a broadband low-noise TIA, four pseudo-differential gain stages and four differential CML inverters whose role is to provide single-ended to differential conversion and to boost the signal before it reaches the retiming flip-flop without degrading the SNR. The clock signal is applied through a transformer-input and broadband distribution network to the flip-flop whose outputs are taken off-chip via a chain of output drivers.

4.3.2.1. Data Path (TIALA and Differential Amplifiers)

The TIA takes advantage of the quintessential digital circuit, the CMOS inverter, in a transimpedance feedback configuration. The gate width ratio of the p-MOS and n-MOS transistors is set approximately equal to the inverse of the $f_T$, $I_{ON}$, and characteristic current density ratios of the two devices [16]. The p-MOSFET and n-MOSFET are biased at 0.1mA/µm and 0.2mA/µm, respectively, which loosely correspond to the optimum noise figure current density and maximum power gain for each device, respectively. The total gate periphery and gate capacitance of the p-MOS and n-MOS devices ensures that the noise impedance of the TIA is also matched to 50Ω over a broadband. The layout of the CMOS inverter takes advantage of the equal number of gate fingers of the p-MOS and n-MOS transistors to minimize interconnect footprint and layout parasitic capacitance (Figure 4.18).

A major disadvantage of this CMOS TIA topology is the dependency of the bias
Figure 4.15: The TIALA-Retimer block diagram.

Figure 4.16: The TIALA-Retimer functional diagram.
current on the supply voltage. Due to the TIA’s single-ended topology and absence of any current source transistors, the supply voltage is divided between the n-MOS and p-MOS transistors \((V_{GS-nMOS} + V_{SG-pMOS} = V_{DD})\). Furthermore, the CMOS TIA also biases the following limiting amplifier stages. Therefore, the entire TIALA suffers from poor power supply rejection. However, in order to achieve the desired bandwidth and gain from the TIA using a 1.2V supply, this topology had to be adopted.

The role of the pseudo-differential limiting amplifier stages, whose schematic is shown in Figure 4.17, is to provide at least 20dB of gain to overcome the metastability of the flip-flop for TIA input signal levels as low as 10mV\(_{PP}\), even when applied in a single-ended fashion. The limiting-amplifier chain uses HVT and standard-\(V_T\) (SVT) transistors to maximize the \(V_{DS}\) of the input transistors. Since the TIALA has no common-mode rejection, a chain of differential n-MOS inverters (Figure 4.19) with inductive series-shunt peaking and active tail current sources is inserted between the TIALA and the flip-flop. This combination of differential TIA, a pseudo-differential and differential CML inverter chain ensures that sensitivity and bandwidth are maximized even for single-ended inputs. All signal-path n-MOSFETs are biased at 0.2mA/\(\mu\)m. Deep sub-micron CMOS transistors typically have a drawn and physical gate length that is smaller than the technology node name. The process feature size typically represents the smallest metal pitch of the back-end. In this 65nm process, the transistors have a drawn gate length

![Figure 4.17: Transimpedance amplifier (TIA) and limiting-amplifier (LA) schematics. All MOSFETs have a drawn gate length of 60nm.](image-url)
of 60nm and a physical gate length of 45nm. The simulated small-signal bandwidth of the TIALA after extraction of RC layout parasitics is 45GHz, limited by the limiting-amplifier chain. The simulated small-signal gain of the TIALA is shown in Figure 4.20. The TIA achieves a single-ended gain of 7dB over a 3-dB bandwidth of 80GHz.

4.3.2.2. Clock Distribution

The clock distribution network (Figure 4.21) features a broadband planar transformer for single-ended to differential conversion. A series-shunt network \((C_M, L_M)\) is used to match the transformer input to 50Ω. The same CMOS TIA as in the data path is implemented in the clock path. By AC-coupling the TIA outputs to a chain of tuned differential pairs with active current sources, 6dB of differential gain is achieved over a broad frequency band extending from 50GHz to 85GHz (Figure 4.22). The tuned differential pairs provide sufficient common-mode rejection at 80GHz to eliminate the impact of transformer asymmetries on the clock signal. The first differential pair is tuned at approximately 70GHz and the second stage is tuned at 80GHz. The final stage of the clock distribution uses a 0.8V supply and its outputs directly bias the clock path transistors of the flip-flop. The clock tree uses SVT and HVT transistors.
Figure 4.19: Schematic of the differential CML inverters for single-end to differential conversion. All MOSFETs have a drawn gate length of 60nm.

Figure 4.20: Simulated small-signal gain of the TIALA.
Figure 4.21: Tuned 81GHz clock distribution schematics. All MOSFETs have a drawn gate length of 60nm.

Figure 4.22: Simulated small-signal gain of the tuned clock distribution.
4.3.2.3. Flip-Flop

The schematic of the flip-flop, shown in Figure 4.23, is similar to that in [15]. The use of high-$V_T$ (HVT) devices on the clock path (which increases the $V_{DS}$ of the clock distribution transistors) and low-$V_T$ (LVT) devices on the data path allows the latch to operate from a 1.2V supply and at 81Gb/s. Each latch consumes 8mA. In this design, the layout of each block is optimized by interdigitating and merging transistors with common sources or drains in a single well. For instance, the layout of the latch uses interdigitated clock-pair and data-path transistors with gate fingers of 0.8µm and 1.6µm, respectively. This technique, shown in Figure 4.24, results in a compact latch layout which minimizes interconnect capacitance and transistor mismatch. Such layout techniques are particularly important in nanoscale CMOS where interconnect parasitic capacitances can approach the value of the intrinsic capacitance of the device.

4.3.2.4. Output Driver

The output driver consists of a chain of four scaled CML inverters with inductive peaking and a maximum fan-out of 1.5 (Figure 4.25). The final stage produces a swing of 0.4Vpp per side in 50Ω loads. The output driver uses HVT and SVT transistors.

4.3.3. Fabrication

The circuit was fabricated in a standard digital 65nm GPLP CMOS process with a 7 layer metal back-end. The measured $f_T$ and $f_{MAX}$ for an 80x60nmx1µm n-MOSFET are 170GHz and 250GHz at $V_{DS} = 0.6V$. The technology provides HVT, SVT and LVT transistors with the same $f_T$ and $f_{MAX}$. Figure 4.26 shows the die photo with an area of 0.57mmx1.2mm, including pads.

4.3.4. Measurement Results

The power consumption of the entire circuit is 200mW from a 1.2V supply, with 9.6mW, 48mW and 30mW being consumed by each latch, the TIALA, and the clock distribution
Figure 4.23: D-FF schematics. The flip-flop uses LVT and HVT transistors to achieve 81Gb/s operation from a 1.2V supply. All MOSFETs have a drawn gate length of 60nm.

Figure 4.24: Optimized layout of the CML latch core transistors.
Figure 4.25: Schematics of the output driver. All MOSFETs have a drawn gate length of 60nm.

Figure 4.26: Chip micrograph.
network, respectively. The measured and simulated input return loss of the TIA and output driver is shown in Figure 4.27. Similarly, the input return loss of the tuned clock tree is shown in Figure 4.28.

Figure 4.29 shows the post layout simulated eye diagram of the TIALA-retimer at various nodes within the circuit. The retimer was tested on-wafer with a single-ended 2\(^7\)-1 PRBS signal generated by the transmitter described in [48]. The block diagram of the test setup for 81Gb/s operation is shown in Figure 4.30. A 75-100GHz Millitech (x6) multiplier is used to generate the 81GHz clock signal. The multiplier is driven by an RF source at 1/6th of the clock frequency which also triggers the precision time-base module of the Agilent DCA-J oscilloscope. The 90Gb/s transmitter requires a half-rate clock signal which is provided by a second RF source. The two RF sources are synchronized via a 10MHz reference signal. A mechanical phase shifter is employed to align the retimer clock signal with the PRBS transmitter data. The half rate clock is also passed through a by-four divider to generate a (1/8)th-rate clock which is in turn used trigger the pattern-lock module of the DCA-J oscilloscope.

The PRBS generator was mounted on a separate probe station. One side of its differential output was connected to the TIALA-retimer under test via a 24-inch, 50GHz cable. The other side of the PRBS generator output was connected to the oscilloscope via an identical cable allowing the oscilloscope to capture the input and output signals of the TIALA-retimer simultaneously and illustrate the effect of the cable losses on the input signal. All measurements are performed single-endedly and are reproduced in Figures 4.31, 4.32, 4.33, 4.34, 4.35, 4.36. The minimum input eye amplitude is 40mVpp at 75Gb/s, 60mVpp at 78Gb/s and 80mVpp at 81Gb/s. The improvement in rise-time, eye opening and jitter (e.g. from 2.5ps RMS to 384fs RMS in Figure 4.31) are illustrated in all cases. In the absence of an 80Gb/s BERT, the pattern-lock feature of the DCA-J was employed to verify the correctness of the input and output bit streams, Figures 4.32, 4.36.
Figure 4.27: Simulated and measured input return loss of the TIA and output driver.

Figure 4.28: Simulated and measured input return loss of the tuned clock tree.
Figure 4.29: Simulated eye diagram at the input, after the TIALA, retimer and output driver at 80Gb/s.

Figure 4.30: Block diagram of the test setup.
Figure 4.31: Measured input and output eye diagrams at 75Gb/s. The input eye height/amplitude is $5mV_{PP}/40mV_{PP}$.

Figure 4.32: Correct measured input and output $2^7-1$ PRBS patterns at 75Gb/s corresponding to Figure 4.31.
Figure 4.33: Measured input and output eye diagrams at 78Gb/s. The input eye height/amplitude is $12\text{mV}_{PP}/60\text{mV}_{PP}$.

Figure 4.34: Measured input and output eye diagrams at 81Gb/s. The input eye height/amplitude is $63\text{mV}_{PP}/250\text{mV}_{PP}$.
Figure 4.35: Measured input and output eye diagrams at 81Gb/s. The input eye height/amplitude is $15\text{mV}_{PP}/80\text{mV}_{PP}$.

Figure 4.36: Correct measured input and output $2^7$-1 PRBS patterns at 81Gb/s corresponding to Figure 4.35.
4.3.5. Conclusions and Future Work

A record full-rate TIALA-retimer, capable of operating with input data rates up to 81Gb/s is reported. The input TIA stage exhibits a figure of merit of 118.5\mu W/Gb/s. The output eye diagrams show rise/fall times better than 6ps and RMS jitter below 380fs. These results demonstrate that 65nm and 45nm CMOS technologies can compete with SiGe and InP technologies for the next-generation serial 80-110Gb/s wireline transceivers. Table 4.2 shows a comparison of state of the art published retimers.

For the next phase of this project, an 81Gb/s clock and data recovery (CDR) unit may be designed. Some of the basic building blocks for a phase-lock loop (PLL) based CDR already exists in this technology such as a 90GHz static divider [15]. Furthermore, a feed-forward equalizer may be implemented in the place of the limiting-amplifier to combat some of the channel losses at these bit-rates. Offset cancelation circuits are also required for high-gain TIALAs and may be implemented for this circuit [28].

Some of the existing circuit blocks can also be improved. For instance, the final stage of the clock tree uses a 0.8V power supply with an input common-mode of 1.2V. This results in a low input transistor $V_{DS}$ of 400mV. This sub-optimal biasing was overlooked at the time of the original design. Figure 4.37 shows an improved schematic of the tuned clock tree. A common-mode feedback network adjusts the common-mode output of the clock tree which biases the following latch stage. Furthermore, the final stage is capacitively coupled. This allows the input common-mode voltage to be lowered to 0.9V.
Figure 4.37: Improved schematics of the tuned clock tree.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Technology</th>
<th>( f_{\text{sample}} ) (GS/s)</th>
<th>Track BW (GHz)</th>
<th>THD ( \text{dB} @ f_{\text{in}} )</th>
<th>Supply (V)</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>0.13(\mu)m CMOS</td>
<td>30</td>
<td>7</td>
<td>(-30 @ 1) GHz (-29 @ 7) GHz</td>
<td>1.8</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td>( f_T = 85) GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>0.18(\mu)m CMOS</td>
<td>10</td>
<td>N/A</td>
<td>(-24.7 @ 5) GHz</td>
<td>1.8</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>( f_T = 45) GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>InP</td>
<td>12</td>
<td>14</td>
<td>(-23.3 @ 12) GHz</td>
<td>-5.2</td>
<td>390</td>
</tr>
<tr>
<td></td>
<td>( f_T = 120) GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.25(\mu)m SiGe BiCMOS</td>
<td>12</td>
<td>5.5</td>
<td>(-52.4 @ 1.5) GHz</td>
<td>3.7</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>( f_T = 200) GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.18(\mu)m SiGe BiCMOS</td>
<td>50</td>
<td>42</td>
<td>N/A</td>
<td>4, 3.3</td>
<td>640</td>
</tr>
<tr>
<td></td>
<td>( f_T = 200) GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.18(\mu)m SiGe BiCMOS</td>
<td>40</td>
<td>43</td>
<td>(-27 @ 20) GHz (-29 @ 10) GHz</td>
<td>3.6</td>
<td>540</td>
</tr>
<tr>
<td></td>
<td>( f_T = 160) GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Table of comparison of state of the art published THAs.
Table 4.2: Table of comparison of state of the art published retimers.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Technology</th>
<th>Bit-Rate (Gb/s)</th>
<th>Supply (V)</th>
<th>$P_{\text{Latch}}$ (mW)</th>
<th>$P_{\text{Total}}$ (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>InP HEMT, $f_T = 245\text{GHz}$</td>
<td>80</td>
<td>-5.7</td>
<td>N/A</td>
<td>1200</td>
</tr>
<tr>
<td>15</td>
<td>SiGe BiCMOS, $f_T = 150\text{GHz}$</td>
<td>48</td>
<td>2.5</td>
<td>23</td>
<td>288</td>
</tr>
<tr>
<td>16</td>
<td>90nm CMOS, $f_T = 120\text{GHz}$</td>
<td>40</td>
<td>1.2</td>
<td>10.8</td>
<td>130</td>
</tr>
<tr>
<td>49</td>
<td>65nm CMOS, $f_T = 170\text{GHz}$</td>
<td><strong>81</strong></td>
<td><strong>1.2</strong></td>
<td><strong>9.6</strong></td>
<td><strong>200</strong></td>
</tr>
<tr>
<td>(This Work)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions and The Future of mm-Wave ADCs

While data converters have been around for nearly ninety years, mm-wave data converters are still in their infancy. Only recently the 40-GHz sampling barrier was broken with the introduction of the next generation Tektronix oscilloscopes. Meanwhile, data communication is the main driving force behind mm-wave data converter development. As with all mm-wave circuits, designers must go beyond simply relying on technology advancement to achieve acceptable performance. Careful device and passive modeling is critical and systematic design methodology may offer repeatable and scalable mm-wave designs.

5.1. Contributions

In Chapter 2, the design methodology to enable mm-wave operation of track and hold amplifiers in a SiGe process is presented. Based on this design methodology, a SEF THA was demonstrated to operate at 40 GS/s for the first time [32]. Furthermore, in Chapter 4 it was shown that a similar design procedure can be applied to CMOS processes in order to bring CMOS THAs into the mm-wave era. Based on this work, a record breaking 30-GS/s SSF THA was published in [34]. This was the first instance of a SSF THA operating at mm-wave frequencies.

However, as explored in Chapter 3, it is not trivial to effectively incorporate high-speed THAs in mm-wave ADCs due to very high data distribution bandwidth requirements.
Chapter 5. Conclusions and The Future of mm-Wave ADCs

This was experimentally verified and published in [7] through the implementation of a 4-bit, 35 GS/s ADC. In this chapter, various data distribution networks are compared and contrasted to illustrate the difficulties associated with building a single-chip mm-wave ADC. Furthermore, an active data distribution network was presented for the first time to demonstrate the single-chip 35-GS/s ADC. The presented ADC is currently the only mm-wave ADC which does not require an external power amplifier.

Finally, the potential of CMOS for implementing mm-wave broadband receivers and data converters is explored in Chapter 4 and a record breaking 81-Gb/s TIALA-retimer has been presented based on that work [49]. Critical layout techniques are also presented in this chapter maximize the performance of mm-wave CMOS designs.

5.2. Future of mm-Wave ADCs

The current industry trend for implementing mm-wave data converters is toward time-interleaved architectures. This may seem counter-intuitive at first, since historically flash converters have been the architecture of choice for breaking sampling rate barriers. Several factors seem to contribute to this paradigm shift. First, mm-wave ADCs are built increasingly using multi-chip solutions. The front-end amplifier and even the THA may be implemented as a separate chip. This allows the front-end amplifier/THA to be implemented using high breakdown/faster technologies. Therefore, the large capacitive load associated with massively parallel time-interleaved architectures can be driven by the off-chip front-end amplifier. To the best of my knowledge, every reported mm-wave time-interleaved ADC to date uses this approach [2, 4, 35]. Second, with the recent wide availability of efficient, multi-gigahertz CMOS SARs and pipeline ADCs, designers need to interleave fewer sub-ADCs to reach mm-wave sampling rates. This, combined with the advances in calibration techniques and high levels of integration CMOS offers, makes time-interleaved CMOS mm-wave ADCs an attractive option [2, 4]. However, for low resolution (≤ 4-bits), a flash ADC (in any technology) may still offer the simplest
solution, certainly the simplest solution to break sampling records. The flash mm-wave ADC presented in this thesis has been implemented using 0.18$\mu$m SiGe BiCMOS process. By switching to a 0.13$\mu$m SiGe BiCMOS process, the data distribution tree can be greatly improved for better bandwidth. Furthermore, it is worthwhile to investigate the feasibility of time-interleaving two of such ADCs to achieve $\geq$ 70 GHz sampling rates.

Currently, a 65nm SiGe BiCMOS process is not available from any foundry. However, such a process can offer several advantages for realizing mm-wave ADCs where either a 65nm process or an older generation SiGe BiCMOS process would fall short. For instance, the high voltage breakdown of the SiGe HBT transistors combined with their high $f_T$ and $f_{MAX}$ makes them ideal for realizing high dynamic range front-end amplifiers (i.e. TIAs), high linearity track and hold amplifiers (i.e. SEF) and high linearity active data distribution networks. Meanwhile, the high-speed 65nm CMOS process, which operates from a lower power supply, can be utilized for low-power retimers (i.e. flip-flops), highly integrated on-chip deep-memory (i.e. SRAM), complex digital calibration techniques and de-multiplexing digital circuits. This combination can provide a single-chip mm-wave ADC solution with a built in front-end amplifier, memory and CMOS compatible digital interfaces.
Bibliography


