Front-end Building Blocks for 100+ GS/s ADCs

by

Konstantinos Vasilakopoulos

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

© Copyright 2016 by Konstantinos Vasilakopoulos
Abstract

Front-end Building Blocks for 100+ GS/s ADCs

Konstantinos Vasilakopoulos
Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto
2016

In this thesis IC building blocks for future 100+Gbaud fiber-optic receivers, employing high-speed ADC converters above 100 GSps are investigated. A single-ended broadband, low-noise amplifier was designed in a 55nm SiGe BiCMOS technology. The DC-92GHz amplifier consumes 48 mW from a 2.3V supply and achieves 13 dB of gain and a noise figure of 6 dB up to 88 GHz. The circuit was measured to operate correctly up to 120 Gbps. To the best of the author’s knowledge, this is the lowest noise amplifier in any technology ever tested at this data rate.

A high-speed track-and-hold amplifier (THA) was designed in the same technology for time-interleaved 100+GS/s ADCs. It can support 90GS/s operation (was tested up to 108 GS/s), has 40GHz input bandwidth, and consumes 87 mW from 1.8V and 2.5V supplies, delivering record-breaking performance among previously published high-speed THAs.
Acknowledgements

I would like to thank my supervisor, Professor Sorin Voinigescu, for giving me the opportunity to study in Canada and providing expert guidance along the way. I would also like to thank my examination committee: Professor Joyce Poon, Professor Antonio Liscidini, and Professor Glenn Gulak.

I am grateful to all my colleagues in BA4182 who have made my graduate studies a pleasant experience. In particular, I would like to thank: James Hoffinan, Yingying Fu, Stefan Shopov, Hassan Farooq, James Bateman, Guy Alter, and Sadegh Dadash. Many thanks go to Jaro Pristupa for CAD support and Ioannis Sarkas for his excellent advice throughout my studies. I am also deeply thankful to Dan Case and The’ Linh Nguyen from Finisar Corporation for giving me the opportunity to work with them during my summer internship.

This thesis would not have been possible without the love and support of my family, and especially my girlfriend, who has stood by me since the first day I embarked on this journey.

The following organizations are acknowledged for their support: STMicroelectronics for providing design kits and fabrication, CMC for CAD support, Keysight, SHF Communication Technologies AG, and Anritsu for equipment, and Finisar Corporation for financial support.

This work is dedicated to the memory of my undergraduate supervisor, Kosta Efstathiou.
## Contents

**Abstract** ii  
**Acknowledgements** iii  
**List of Tables** vi  
**List of Figures** xi  
**List of Abbreviations and Notations** xii  

### 1 Introduction 1  
1.1 Motivation ........................................ 1  
1.1.1 State-of-the-Art ADCs .......................... 3  
1.2 Objectives ......................................... 7  
1.3 Thesis Overview .................................... 7  

### 2 Building Blocks for High-Speed A/D Converters 9  
2.1 Broadband Low-Noise Amplifier .................. 9  
2.1.1 Useful Results from the Theory of Noisy Linear Networks .......................... 9  
2.1.2 Low-Noise Broadband Topologies ................. 12  
2.2 Track-and-Hold Amplifier ......................... 27  
2.2.1 THA performance metrics ........................ 30  
2.2.2 THA topologies .................................. 31  

### 3 Low-Power, Low-Noise Broadband Amplifier Design 37  
3.1 TIA topology investigation ......................... 37  
3.2 80 GHz Low-Power TIA ............................ 40  
3.2.1 Design ......................................... 40  
3.3 92 GHz Linear TIA .................................. 45  
3.3.1 Design ......................................... 45  

### 4 Track & Hold Amplifier with New Quasi-CML MOS-HBT Switch 51  
4.1 Design Considerations .............................. 51  
4.1.1 MOS-HBT cascode .................................. 52  
4.1.2 Quasi-CML Topology ............................. 56  
4.2 Proposed Design .................................... 57
List of Tables

1.1 High-speed ADC requirements .............................................. 2
1.2 High-speed ADCs above 20 GS/s ........................................... 6
1.3 Number of blocks (N) in a flash and TI-SAR .......................... 6

2.1 State-of-the-Art Broadband, Low-Noise Amplifiers .................. 27
2.2 State-Of-The-Art Track & Hold Amplifiers ............................. 36

3.1 Power summary of broadband topologies in Fig. 3.1 ................... 40
3.2 Process Corners .............................................................. 44
3.3 $S_{21}$ standard deviation ($\sigma$) after 200 Monte Carlo simulations. 49

4.1 Summary of THA specifications .......................................... 51
4.2 Top 4 noise contributors .................................................. 74
4.3 Pedestal error for various differential inputs ............................ 76
4.4 Simulated DC parameters ................................................... 82
4.5 Simulated AC parameters .................................................. 82

5.1 ADC front-end I/O pin description ........................................ 91

6.1 Summary of low-power TIA performance at room temperature .... 101
6.2 Summary of linear TIA performance at room temperature .......... 109
6.3 Comparison with State-of-the-Art Broadband Amplifiers .......... 109
6.4 Breakdown of THA power consumption at room temperature ...... 118
6.5 Comparison With State-Of-The-Art Track & Hold Amplifiers ...... 118
# List of Figures

1.1 Global data center IP traffic growth [3] (1 zettabyte = $2^{21}$ bytes) .......................... 2
1.2 Example of a flexible coherent optical transceiver architecture [11] ................................. 3
1.3 Simplifier block diagram of a modern fiber-optic receiver ............................................. 3
1.4 Conversion bandwidth vs. SNDR of recently published ADCs [19] ................................. 4
1.5 Schreier’s figure of merit $FOM_S$ vs. Nyquist sampling rate of recently published ADCs [19] 5
1.6 Power consumption plot of flash vs. time-interleaved SAR as a function of resolution (based on the power of a SR quasi-CML D-FF which will be discussed later). ............. 6
2.1 Parallel connection of two noisy two-ports with its equivalent input referred correlated voltage and noise sources [33] ................................................................. 10
2.2 Shunt-shunt feedback amplifier [33] ............................................................................... 11
2.3 Input referred noise (adapted from [33]). $C_T$ includes all capacitances at the input of the amplifier ................................................................................................................. 12
2.4 $f_T$, $f_{MAX}$ and $NF_{MIN}$ versus bias current density for a SiGe HBT device [36] .......... 12
2.5 Common-base amplifier ................................................................................................. 13
2.6 Noise sources in a bipolar transistor [37] ....................................................................... 14
2.7 SiGe HBT noise equivalent circuit [38] ......................................................................... 15
2.8 SiGe HBT noise equivalent circuit with input referred noise sources [38] ................... 16
2.9 Two-port network with input referred noise sources [39] ............................................. 17
2.10 Regulated cascode amplifier .......................................................................................... 18
2.11 HBT-source follower .................................................................................................... 20
2.12 Darlington cascode amplifier ....................................................................................... 22
2.13 HBT-TIA schematic .................................................................................................... 26
2.14 The simplest track-and-hold circuit [14] ..................................................................... 27
2.15 Output waveform of ideal THA [14] ........................................................................... 28
2.16 SHAs used in a time-interleaved ADC architecture [48] ............................................ 28
2.17 (a): Capacitive feedthrough in a flash ADC and (b): harmonic distortion due to nonlinear capacitance of the comparators .................................................. 29
2.18 Errors in ADCs due to timing and bandwidth limitations ............................................. 29
2.19 Errors in a THA [50] .................................................................................................... 31
2.20 Open-loop THA topology [50] ....................................................................................... 32
2.21 Closed-loop THA topology [50] ..................................................................................... 32
2.22 Open-loop THA with diode bridge switch [13] ............................................................ 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.23</td>
<td>Differential circuit implementation of switched emitter follower THA proposed in [52]</td>
</tr>
<tr>
<td>2.24</td>
<td>Feedthrough cancellation capacitor, $C_{ff}$, used in [52, 63]</td>
</tr>
<tr>
<td>2.25</td>
<td>Base-collector diode THA concept in [51]</td>
</tr>
<tr>
<td>3.2</td>
<td>Measured $f_T$ and $f_{MAX}$ of a high-speed HBT (100 nm x 4.5 µm) and a 55nm n-MOSFET</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulated $</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulated $S_{11}$ of the low-noise topologies depicted in Fig. 3.11</td>
</tr>
<tr>
<td>3.5</td>
<td>TIA schematic with component values</td>
</tr>
<tr>
<td>3.6</td>
<td>Simulated S-parameters and noise figure. Vertical dashed line indicates 3dB bandwidth</td>
</tr>
<tr>
<td>3.7</td>
<td>Simulated 1dB compression point (left) and THD (right) at different frequencies and input power levels</td>
</tr>
<tr>
<td>3.8</td>
<td>Simulated group delay</td>
</tr>
<tr>
<td>3.9</td>
<td>80 Gbps simulated eyes with 5 mV_{pp} input</td>
</tr>
<tr>
<td>3.10</td>
<td>Simulated $</td>
</tr>
<tr>
<td>3.11</td>
<td>Linear TIA schematic with component values</td>
</tr>
<tr>
<td>3.12</td>
<td>Layout detail of the feedback network</td>
</tr>
<tr>
<td>3.13</td>
<td>Simulated S-parameters and noise figure of linear TIA. Vertical dashed line indicates 3dB bandwidth</td>
</tr>
<tr>
<td>3.14</td>
<td>Simulated 1dB compression point (left) and THD (right) at different frequencies and input power levels for the linear TIA</td>
</tr>
<tr>
<td>3.15</td>
<td>Simulated group delay of linear TIA</td>
</tr>
<tr>
<td>3.16</td>
<td>120 Gbps $2^{31} - 1$ simulated eyes with 5 mV_{pp} input</td>
</tr>
<tr>
<td>3.17</td>
<td>120 Gbps $2^{31} - 1$ simulated eyes with 50 mV_{pp} input</td>
</tr>
<tr>
<td>4.1</td>
<td>Differential circuit implementation of switched emitter follower THA proposed in [52]</td>
</tr>
<tr>
<td>4.2</td>
<td>40GS/s SiGe BiCMOS THA [54]</td>
</tr>
<tr>
<td>4.3</td>
<td>BiCMOS cascode topologies: (a) MOS-MOS, (b) HBT-HBT, (c) HBT-MOS, and (d) MOS-HBT</td>
</tr>
<tr>
<td>4.4</td>
<td>Calculated time constants: (a) 130nm SiGe BiCMOS [60] and (b) 55nm SiGe BiCMOS node</td>
</tr>
<tr>
<td>4.5</td>
<td>(a) Measured $S_{21}$ for all cascode topologies fabricated in a 130nm SiGe BiCMOS technology [60] and (b) simulated AC gain of all cascode topologies designed in a 55nm SiGe BiCMOS technology [66]</td>
</tr>
<tr>
<td>4.6</td>
<td>Measured MAG of cascode topologies in: (a) a 130nm [60] and (b) a 55nmSiGe BiCMOS technology [59]</td>
</tr>
<tr>
<td>4.7</td>
<td>(a) Regular and (b) quasi-CML BiCMOS switch</td>
</tr>
<tr>
<td>4.8</td>
<td>Block diagram of the proposed track-and-hold amplifier</td>
</tr>
<tr>
<td>4.9</td>
<td>Linear input buffer topologies: (a) differential pair with emitter degeneration, (b) Karanicolias' folded diode loaded differential pair [72], (c) diode loaded differential pair [69], (d) Caprio's quad [74], (e) Quinn's cascomp [70], and (f) Miki differential pair [73]</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>4.10</td>
<td>Linear input buffer schematic.</td>
</tr>
<tr>
<td>4.11</td>
<td>THA in track-mode. Deactivated parts of the circuit are displayed in grey.</td>
</tr>
<tr>
<td>4.12</td>
<td>THA in hold-mode. Deactivated parts of the circuit are displayed in grey.</td>
</tr>
<tr>
<td>4.13</td>
<td>Effect of feedthrough capacitor, $C_{ff}$, on bandwidth.</td>
</tr>
<tr>
<td>4.14</td>
<td>Layout of the switched emitter follower core.</td>
</tr>
<tr>
<td>4.15</td>
<td>Clock amplifier schematic.</td>
</tr>
<tr>
<td>4.16</td>
<td>Layout of the clock amplifier.</td>
</tr>
<tr>
<td>4.17</td>
<td>AC simulation results of the clock amplifier after parasitic extraction.</td>
</tr>
<tr>
<td>4.18</td>
<td>Schematic of the 50Ω output driver.</td>
</tr>
<tr>
<td>4.19</td>
<td>Layout of the proposed THA (pads are not shown).</td>
</tr>
<tr>
<td>4.21</td>
<td>Bandwidth in track-mode across process corners.</td>
</tr>
<tr>
<td>4.22</td>
<td>Simulated noise transfer function after extraction of layout parasitics.</td>
</tr>
<tr>
<td>4.23</td>
<td>Simulated equivalent input referred noise for various sampling rates.</td>
</tr>
<tr>
<td>4.24</td>
<td>Differential output of a full-scale (300 mV_pp per side) 18GHz sinusoidal signal sampled at 90 GS/s. The dashed sinusoid corresponds to the sampling clock, the solid line to the differential output of the switched emitter follower, and the dashed line to the output of the output driver.</td>
</tr>
<tr>
<td>4.25</td>
<td>Single-ended outputs of a full-scale (300 mV_pp per side) 18GHz sinusoidal signal sampled at 90 GS/s. The dashed sinusoid corresponds to the sampling clock, the solid lines to the outputs of the switched emitter follower, and the dashed lines to the outputs of the output driver.</td>
</tr>
<tr>
<td>4.26</td>
<td>Time-domain waveform of a 15.95GHz sinusoidal input sampled at 64 GS/s.</td>
</tr>
<tr>
<td>4.27</td>
<td>4096-point FFT spectrum of a 15.95GHz sinusoidal input sampled at 64 GS/s.</td>
</tr>
<tr>
<td>4.28</td>
<td>Improvement of SNR due to oversampling: 4096-point FFT of a 20GHz sinusoid coherently sampled at (a) 64 GS/s (bin: 1283) and (b) 80 GS/s (bin: 1021).</td>
</tr>
<tr>
<td>4.29</td>
<td>Hanning window 1024-point FFT for different input frequencies sampled at 40 GS/s.</td>
</tr>
<tr>
<td>4.30</td>
<td>Hanning window 1024-point FFT for different input frequencies sampled at 90 GS/s.</td>
</tr>
<tr>
<td>4.31</td>
<td>Simulated 1dB input compression point ($P_{1dB}$) for a 10GHz input signal sampled at 90 GS/s (extracted from -50 dBm).</td>
</tr>
<tr>
<td>5.1</td>
<td>Block diagram of an N times time-interleaved ADC front-end.</td>
</tr>
<tr>
<td>5.2</td>
<td>Timing waveforms of an ideal time-interleaved front-end similar to Fig. 5.1. Crosses represent the samples captured within one period of the sampling clock $f_s$.</td>
</tr>
<tr>
<td>5.3</td>
<td>ADC front-end with quadrature sampling clocks for doubling the effective sampling rate.</td>
</tr>
<tr>
<td>5.4</td>
<td>Timing waveforms of quadrature sampling clock architecture in Fig. 5.3.</td>
</tr>
<tr>
<td>5.5</td>
<td>ADC front-end block diagram.</td>
</tr>
<tr>
<td>5.6</td>
<td>ADC front-end layout. Size is: 1070µm × 895µm.</td>
</tr>
<tr>
<td>5.7</td>
<td>64GHz quadrature hybrid schematic.</td>
</tr>
<tr>
<td>5.8</td>
<td>64GHz quadrature hybrid. Size is: 100µm × 110µm.</td>
</tr>
<tr>
<td>5.9</td>
<td>Simulated phase difference of quadrature hybrid.</td>
</tr>
<tr>
<td>5.10</td>
<td>Simulated 3dB bandwidth from S-parameters.</td>
</tr>
<tr>
<td>5.11</td>
<td>4096-point FFT simulated coherently sampled spectrum of a 10.016GHz sinusoidal input sampled at 64 GS/s (bin: 641) on the I-path.</td>
</tr>
</tbody>
</table>
6.33 ADC front-end S-parameter setup. .................................................. 120
6.34 Measured vs. simulated S-parameters of ADC front-end. .................. 120

7.1 ×4 time-interleaved ADC system with 25% duty cycle sampling clocks. .... 123
7.2 Behavioral simulation results of the first two SAR ADC lanes. ................. 123
7.3 (a): Conceptual block diagram of a SAR ADC [12] and (b): successive approximation register [93]. ................................................................. 124
7.4 Schematic of the proposed SR D latch. .............................................. 124
# List of Abbreviations and Notations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>100GbE</td>
<td>100Gbps Ethernet</td>
</tr>
<tr>
<td>BEOL</td>
<td>Back-End Of Line</td>
</tr>
<tr>
<td>BiCMOS</td>
<td>Bipolar complementary metal oxide semiconductor</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>$B_{SOPT}$</td>
<td>Optimum noise susceptance</td>
</tr>
<tr>
<td>BW</td>
<td>3dB bandwidth</td>
</tr>
<tr>
<td>CML</td>
<td>Current Mode Logic</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary MetalOxide-Semiconductor</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital to Analog Converter</td>
</tr>
<tr>
<td>D-FF</td>
<td>Data Flip Flop</td>
</tr>
<tr>
<td>DMT</td>
<td>Discrete Multitone</td>
</tr>
<tr>
<td>DP-QPSK</td>
<td>Dual Polarization Quadrature Phase Shift Keying</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processor</td>
</tr>
<tr>
<td>EF</td>
<td>Emitter Follower</td>
</tr>
<tr>
<td>ENOB</td>
<td>Effective Number Of Bits</td>
</tr>
<tr>
<td>FDSOI</td>
<td>Fully Depleted Silicon On Insulator</td>
</tr>
<tr>
<td>$f_{MAX}$</td>
<td>Maximum oscillation frequency</td>
</tr>
<tr>
<td>$F_{MIN}$</td>
<td>Minimum noise factor</td>
</tr>
<tr>
<td>FoM</td>
<td>Figure of Merit</td>
</tr>
<tr>
<td>$f_s$</td>
<td>Sampling frequency</td>
</tr>
<tr>
<td>$f_T$</td>
<td>Unity gain frequency</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>GS/s</td>
<td>Giga-Samples per second (also: GSps)</td>
</tr>
<tr>
<td>HBT</td>
<td>Heterojunction Bipolar Transistor</td>
</tr>
<tr>
<td>HD</td>
<td>High Definition</td>
</tr>
<tr>
<td>I/Q</td>
<td>In phase / Quadrature phase</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>IIP&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Input 3rd order intercept point</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>ISSCC</td>
<td>International Solid-State Circuits Conference</td>
</tr>
<tr>
<td>J&lt;sub&gt;OPT&lt;/sub&gt;</td>
<td>Optimal noise current density</td>
</tr>
<tr>
<td>J&lt;sub&gt;pFt&lt;/sub&gt;</td>
<td>peak fT current density</td>
</tr>
<tr>
<td>l&lt;sub&gt;Eopt&lt;/sub&gt;</td>
<td>Optimal emitter length (HBT)</td>
</tr>
<tr>
<td>LNA</td>
<td>Low Noise Amplifier</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple Input Multiple Output</td>
</tr>
<tr>
<td>MOSFET</td>
<td>Metal-Oxide-Semiconductor Field-Effect Transistor</td>
</tr>
<tr>
<td>NF</td>
<td>Noise Figure</td>
</tr>
<tr>
<td>NRZ</td>
<td>Non-Return to Zero</td>
</tr>
<tr>
<td>P&lt;sub&gt;1dB&lt;/sub&gt;</td>
<td>1dB compression point of power gain</td>
</tr>
<tr>
<td>PAM</td>
<td>Pulse Amplitude Modulation</td>
</tr>
<tr>
<td>PRBS</td>
<td>Pseudo-random binary sequence</td>
</tr>
<tr>
<td>PSA</td>
<td>Power Spectrum Analyzer</td>
</tr>
<tr>
<td>Q</td>
<td>Quality factor</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>Rx</td>
<td>Receiver</td>
</tr>
<tr>
<td>SAR</td>
<td>Successive Approximation Register</td>
</tr>
<tr>
<td>SFDR</td>
<td>Spurious-Free Dynamic Range</td>
</tr>
<tr>
<td>SHA</td>
<td>Sample and Hold Amplifier</td>
</tr>
<tr>
<td>SiGe</td>
<td>Silicon Germanium</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SNDR</td>
<td>Signal to Noise and Distortion Ratio</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SoC</td>
<td>System On Chip</td>
</tr>
<tr>
<td>SOI</td>
<td>Silicon On Insulator</td>
</tr>
<tr>
<td>SRF</td>
<td>Self Resonant Frequency</td>
</tr>
<tr>
<td>THA</td>
<td>Track and Hold Amplifier</td>
</tr>
<tr>
<td>THD</td>
<td>Total Harmonic Distortion</td>
</tr>
<tr>
<td>TI</td>
<td>Time Interleaving</td>
</tr>
<tr>
<td>TIA</td>
<td>Transimpedance Amplifier</td>
</tr>
<tr>
<td>VT</td>
<td>thermal voltage of HBT</td>
</tr>
<tr>
<td>Vi</td>
<td>Mosfet threshold voltage</td>
</tr>
<tr>
<td>VLSI</td>
<td>Very Large Scale Integration</td>
</tr>
<tr>
<td>VNA</td>
<td>Vector Network Analyzer</td>
</tr>
<tr>
<td>Y_{COR}</td>
<td>Correlation admittance</td>
</tr>
<tr>
<td>Y_{SOPT}</td>
<td>Optimum noise admittance</td>
</tr>
<tr>
<td>Z_{COR}</td>
<td>Correlation impedance</td>
</tr>
<tr>
<td>Z_{SOPT}</td>
<td>Optimum noise impedance</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

It is an indisputable fact that information technology has become an essential commodity of our data-driven society. Recent advances in wireless communication and integrated circuit (IC) technology have made information easily accessible to everyone worldwide. Thanks to the unprecedented level of integration offered by advanced 28nm or even 14nm CMOS nodes, complex functions that were previously available only as discrete solutions are now realized in a single system-on-chip (SOC) with lower manufacturing costs, lower-power, and improved performance. This has led to the proliferation of millions of products such as laptop computers, smartphones, tablets, and other handheld gadgets allowing people to exchange information, improve their productivity and stay in touch using social networks (Facebook, Twitter) and other online services (Skype).

As noted in [1], the number of devices connected to the Internet in 2010 exceeded the Earth’s population. The amount of data produced by smartphones, tablets, smart sensors, radio-frequency indention (RFID) tags, and other devices is expected to surpass 500 exabytes by 2020 [2]. This explosive increase in global data traffic will be primarily driven by current as well as future applications, each with its own requirements in terms of bandwidth, latency, data capacity, etc. Examples include: 4k Ultra HD video, industrial automation, telemedicine [3], self-driving cars, and big scientific research projects like the Large Hadron Collider [4], to name only a few. According to predictions by Cisco [3], most of the data traffic will continue to take place inside data centers and it is projected that it will grow at a compound annual growth rate (CAGR) of 25% from 2014 to 2019, as shown in Fig. I.1. Hence, Big Data is on its way to establish itself as one of the most profitable markets of the future, projected to generate an annual revenue of over 122$ billion by 2025 [5].

In order to meet the tough future requirements, a radical reformation of the entire data center infrastructure is called for. The large bandwidth of optical fiber makes it the perfect candidate, as it can support the massive amount of data and aggregate data rates of future applications. The current trend is to move from the well-established 100Gbps Ethernet (100GbE) standard to 400GbE solutions (anticipated in 2017) and in the future to even 1Tb/s per carrier [6]. Due to increased power consumption traditional ON-OFF light intensity modulation techniques have become obsolete at higher bit rates and have been replaced with more sophisticated modulation formats like dual-polarization quadrature phase-shift keying (DP-QPSK) [7]. The use of multiple carriers and higher-order modulation leads to a net
increase of the overall data rate [8]. In the case of dual polarization quadrature phase shift keying (DP-QPSK), for example, a coherent digital receiver is necessary with 4 analog-to-digital converter (ADC) channels, in order to digitize the I/Q information of the two optical polarizations [7]. Even though the design of a coherent receiver tends to be significantly more complex compared to a direct detection receiver, the former offers the advantage of compensating for various optical nonidealities, such as chromatic dispersion, in the digital domain [8] and has thus become the norm in most fiber-optic links today.

An example of a digital, single-carrier, optical transceiver was designed by Ciena [9] and is shown in Fig. 1.2. The main bottleneck on the Rx side is located in the ADC, where good resolution (6 bits or more), high sampling rate and low-power consumption are required. As data traffic keeps increasing, more complex modulation schemes are called for, which in turn put more pressure on high-speed ADC design as confirmed by the numbers of Table 1.1. To address future 1Tb/s per carrier communication links, dual-polarization 16QAM solutions at 125 Gbaud are needed, which translates to 8 data lanes running at 125 Gb/s [10]. This means that ADCs for systems like these must be able to support bandwidths in excess of 100 GHz with at least 6 bit of resolution at 200 GSps.

<table>
<thead>
<tr>
<th>Modulation</th>
<th>Resolution [bits]</th>
<th>Sampling Rate [GSp]</th>
<th>Bandwidth [GHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>40G DP QPSK</td>
<td>6</td>
<td>23</td>
<td>5-10</td>
</tr>
<tr>
<td>100G DP QPSK Dual Carrier</td>
<td>6</td>
<td>29</td>
<td>5-10</td>
</tr>
<tr>
<td>100G DP QPSK</td>
<td>6</td>
<td>56-65</td>
<td>15-20</td>
</tr>
<tr>
<td>200G DP QAM16</td>
<td>6-8</td>
<td>56-65</td>
<td>15-20</td>
</tr>
<tr>
<td>Future Systems</td>
<td>6-8</td>
<td>&gt;100</td>
<td>&gt;40</td>
</tr>
</tbody>
</table>

Table 1.1: High-speed ADC requirements [12]
III. DIGITAL-TO-ANALOG CONVERTERS

Digital-to-analog converters at the transmitter allow DSP for equalization and the capability of a wealth of software definable modulation formats from a single transceiver. In combination with electro-optics, near-arbitrary 4-D (magnitude and phase on each of two polarizations) optical fields can be constructed at the transmitter output. Fig. 2(a) shows a configuration to generate such fields using 5-bit DACs ($2^5 = 32$ levels) for a linear electro-optics (E/O) transduction. Each dot of the transmitter complex output field is addressable. Fig. 2(b) shows an example of a 200 Gb/s payload 16-QAM field values pre-compensated with –40 ns/nm of pre-dispersion (one polarization is shown). In operation, the transmitter implements a field closest to the desired field at each instant.

In practice, the E/O transduction may not be linear and addressable field locations are, consequently, not regularly spaced as in Fig. 2. In such cases, pre-compensation for a memory-less nonlinearity might be used. This can take the form of a lookup table which conditions the DAC input.

High-speed DAC, and ADC, design attends: bit resolution, sample rate, signal-to-noise-plus-distortion ratio (SNDR), clock speed, jitter, and power dissipation. As consequence of bit resolution at high speed, it is necessary to co-integrate converters with the DSP. For example, four 6-bit 40 GSa/s converters would require an overall transfer rate of 960 Gb/s between DACs and DSP. This requirement informs the choice of technology, such as CMOS or BiCMOS, as well as designs for low power dissipation and small footprint.

A. DAC Topologies

High-speed DAC cores are usually based on a current-steering architecture. Examples include: i) thermometer [5], ii) R-2R ladder network [6], iii) binary weighted [6], and iv) segmented [7]. These are shown in the schematics of Fig. 3. In the thermometer architecture, $2^N - 1$ switchable identical current sources are used, where $N$ is the converter’s number of bits [see Fig. 3(a)]. Only one current source is switched by means of decoding logic for any single least significant bit (LSB) change in the digital input. The circuit is comparatively large but has guaranteed monotonicity.

1.1.1 State-of-the-Art ADCs

A simplified block diagram of a modern receiver used for high-speed optical link applications is illustrated in Fig. 1.3. A photodiode produces a weak current signal proportional to the intensity of light, which gets converted to voltage by a transimpedance amplifier (TIA). The output of the TIA is then amplified by a broadband low-noise amplifier (LNA), in order to maintain a good signal-to-noise ratio (SNR) before being sampled by the track-and-hold amplifier (THA). Finally, the analog samples are digitized by the ADC and processed by a digital signal processor (DSP) in the digital domain. The best candidates for ADCs in high performance fiber-optic systems are the time-interleaved successive approximation register (SAR) and flash architecture [12]. The interested reader is referred to [13, 14] for more information on these architectures. Note that in the case of a flash ADC, the preceding THA is often omitted [15].

In the quest for higher sampling rates and low-power operation, which are critical parameters of

![Fig. 1.3: Simplifier block diagram of a modern fiber-optic receiver.](image-url)
ADCs targeting data-demanding environments, such as data centers, several figures of merit (FoMs) have been derived. It has been observed that the overall quality of the A/D conversion in high-speed designs as represented by the signal-to-noise and distortion ratio (SNDR), is mainly limited by the jitter of the clock being used [16]. Therefore, it makes sense to quantify the spectral purity of a high-speed converter design by taking into account the effect of clock jitter as given by [17]:

$$SNDR_{jitter} \approx \frac{1}{(2\pi f_{in})^2 \sigma_j^2}$$

(1.1)

where $\sigma_j^2$ is the aperture jitter in rms. Although the above metric leads to rather pessimistic results as it counts all ADC nonidealities (quantization noise, distortion, thermal noise, etc.) as jitter [17], it provides a fair comparison for designs operating with GHz sampling clocks. Fig. 1.4 displays the conversion speed of recently published ADCs from VLSI Symposium and ISSCC in the past 20 years. On the y-axis $f_{in,hf}$ represents the highest input frequency for which the SNDR of the x-axis was reported (also known as conversion bandwidth). It is obvious that in 2015 most designs have surpassed the 0.1ps rms boundary with [18] reaching the lowest reported jitter of 127 fs rms.

Relation (1.1), however, does not include power consumption, which can quickly reach high values at mm-wave sampling clocks. For recently published ADC designs with moderate to high resolutions that approach the fundamental thermal limits of analog circuits, it has been observed that power quadruples for each effective bit of resolution [20]. A suitable figure of merit that captures this trend has been
Fig. 1.5: Schreier’s figure of merit $FOM_{S}$ vs. Nyquist sampling rate of recently published ADCs [19], proposed by Schreier in [21]:

$$FOM_{S} = SNDR(dB) + 10\log\left(\frac{f_{s}}{P}\right)$$  \hspace{1cm} (1.2)$$

where $f_{s}$ is the sampling rate and $P$ the power consumption. Looking at Fig. 1.5, the fastest ADC reported to date achieves 141 dB SNDR at 90 GS/s and is a CMOS time-interleaved ADC in IBM’s 32nm SOI CMOS technology [22]. Table 1.2 summarizes the performance of recently published ADCs with sampling rates above 20 GS/s. The SNDR values on the table correspond to the highest measured input frequency reported in the BW column. Interestingly, the majority of the designs with 6 or more bits of resolution follow a time-interleave successive approximation (TI-SAR) architecture, whereas lower resolution ADCs make use of a flash architecture. This is because the number of comparators in a flash ADC grows exponentially with the resolution of the converter. For example an 8-bit flash ADC would require $2^8 - 1$ comparators resulting in excessive power dissipation and a slow input time constant due to increased capacitance at the input.

Although the simplicity of the flash architecture looks attractive at low resolution values, as the number of bits grows, the exponentially increasing number of comparators incurs a significant bandwidth and power penalty. To quantify the power requirements of the two architectures, the number of components (N) needed in each case is listed in Table 1.3.

Assuming a very optimistic scenario, where every block consumes as much as a settable and resettable D flip-flop, SR D-FF, (a high-speed version of which will be described in Chapter 7), the border where the power benefits of the TI-SAR approach shine, is approximately at 4.5 bits of resolution as plotted in
Table 1.2: High-speed ADCs above 20 GS/s

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>TI-SAR</td>
<td>32nm CMOS</td>
<td>90</td>
<td>19.9</td>
<td>8</td>
<td>33</td>
<td>0.667</td>
<td>141</td>
</tr>
<tr>
<td>23</td>
<td>TI-Flash</td>
<td>0.13µm SiGe BiCMOS</td>
<td>50</td>
<td>22</td>
<td>5</td>
<td>22.1</td>
<td>5</td>
<td>118</td>
</tr>
<tr>
<td>21</td>
<td>TI-SAR</td>
<td>28nm FD-SOI CMOS</td>
<td>46</td>
<td>23</td>
<td>6</td>
<td>25.2</td>
<td>0.381</td>
<td>133</td>
</tr>
<tr>
<td>23</td>
<td>Flash</td>
<td>0.13µm SiGe HBT</td>
<td>40</td>
<td>13</td>
<td>3</td>
<td>20</td>
<td>3.8</td>
<td>117</td>
</tr>
<tr>
<td>25</td>
<td>TI-SAR</td>
<td>65nm CMOS</td>
<td>40</td>
<td>18</td>
<td>6</td>
<td>25.2</td>
<td>1.5</td>
<td>126.5</td>
</tr>
<tr>
<td>24</td>
<td>Flash</td>
<td>0.13µm SiGe BiCMOS</td>
<td>35</td>
<td>11</td>
<td>4</td>
<td>19.8</td>
<td>4.5</td>
<td>116</td>
</tr>
<tr>
<td>26</td>
<td>TI-Flash</td>
<td>40nm CMOS</td>
<td>25</td>
<td>10</td>
<td>6</td>
<td>25.8</td>
<td>0.5</td>
<td>129.8</td>
</tr>
<tr>
<td>29</td>
<td>TI-two step</td>
<td>65nm CMOS</td>
<td>25</td>
<td>12.5</td>
<td>6</td>
<td>29.7</td>
<td>0.088</td>
<td>141</td>
</tr>
<tr>
<td>30</td>
<td>TI-SAR</td>
<td>90nm CMOS</td>
<td>24</td>
<td>12</td>
<td>6</td>
<td>22.8</td>
<td>1.2</td>
<td>122.8</td>
</tr>
<tr>
<td>31</td>
<td>Flash</td>
<td>0.13µm SiGe BiCMOS</td>
<td>22</td>
<td>10</td>
<td>5</td>
<td>20</td>
<td>3</td>
<td>115.6</td>
</tr>
</tbody>
</table>

Table 1.3: Number of blocks ($N$) in a flash and TI-SAR [10].

<table>
<thead>
<tr>
<th>Flash</th>
<th>TI-SAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>comparators: $2^N - 1$</td>
<td>comparators: $N + 2$</td>
</tr>
<tr>
<td>flipflops: $2^N - 1$</td>
<td>flipflops: $N(N + 2)$</td>
</tr>
<tr>
<td>DACs: $N + 2$</td>
<td></td>
</tr>
<tr>
<td>samplers: $N + 3$</td>
<td></td>
</tr>
<tr>
<td>clock tree fanout: $2^N - 1$</td>
<td>clock tree fanout: $N^2 + 3N + 3$</td>
</tr>
<tr>
<td>analog tree fanout: $2^N - 1$</td>
<td>analog tree fanout: $N + 2$</td>
</tr>
</tbody>
</table>

Fig. 1.6: Power consumption plot of flash vs. time-interleaved SAR as a function of resolution (based on the power of a SR quasi-CML D-FF which will be discussed later).
Chapter 1. Introduction

Fig. 1.6 based on the relations in Table 1.3. Therefore, the TI-SAR topology proves more power-efficient after 5-6 bits of resolution.

As far as technology is concerned, CMOS is preferred at high resolutions as it consumes no static power and, thus, results in low overall power dissipation. Furthermore, as long as a fast track-and-hold amplifier (THA) exists [32], multiple slower SAR CMOS ADC lanes can be employed and their mismatches and nonidealities can be corrected via powerful digital signal processors (DSP), which can seamlessly interface with the ADC in advance CMOS nodes. On the other hand, the input bandwidth still remains an issue as can be seen from the 90GS/s design in [22] that exhibits only about 20GHz bandwidth. The latter is limited by the bandwidth and fanout of the cascaded CMOS switches in the 32nm SOI CMOS node. This is where SiGe HBT technology shows a definite advantage. The higher $g_m$ of the HBT device leads to bigger gains at lower bias currents, allowing for the integration of faster THAs and other blocks, such as broadband low-noise amplifiers (LNAs) or transimpedance amplifiers (TIAs) on the same die. Comparing [23] with [24] we can observe that the sampling rate and bandwidth are approximately the same at similar resolution, even though [24] is designed using a much newer 28nm FDSOI CMOS technology. SiGe solutions, however, tend to result in higher power dissipation due to the higher voltage supplies needed for their operation.

1.2 Objectives

The objective of this thesis is to investigate IC building blocks for 100+Gbaud fiber-optic receivers employing high-speed ADC converters above 100 GS/s, able to support the speeds of future fiber-optic systems. As illustrated in Fig. 1.2 the first block of such a receiver is the linear gain amplifier. This amplifier must simultaneously meet different criteria. It must provide enough gain (>10 dB) and linearity to support PAM-4 and higher-order modulation formats. It must also be broadband (from DC to above 80 GHz for 100 Gb/s operation) and low-noise (less than 7 dB noise figure), while dissipating as little power as possible (less than 100 mW).

An important component of every high-speed ADC is the THA. Its performance determines the linearity, sampling rate, and input bandwidth of the ADC system. To comply with rich modulation formats such as DP QAM16, a target linearity of 7 bits with over 65GS/s sampling rate and over 30GHz is assumed. Using a master and 9 slave THAs a sampling front-end suitable for time-interleaved 7bit SAR ADCs can be designed. By clocking two masters in quadrature, the overall sampling rate can be doubled reaching 128 GS/s (2×64 GS/s) or even higher values. Hence, low-power operation (less than 100 mW) of the THA is important to ensure low power dissipation of the ADC system.

Since SiGe BiCMOS offers the best of both worlds (high $f_T$ and $g_m$ of SiGe HBT and fast MOS devices) with superior BEOL for passive elements such as inductors, it can potentially lead to faster and lower power circuits compared to CMOS or SiGe-only nodes, and will be the technology of choice for the design of the linear amplifier and the THA circuits.

1.3 Thesis Overview

Chapter 2 provides some background knowledge on broadband low-noise amplifiers and track-and-hold amplifiers. Chapter 3 discusses the design of a 92GHz SiGe HBT linear amplifier with 6dB noise figure. Chapter 4 deals with the design of a SiGe BiCMOS THA with quasi-CML sampling switch and
Chapter 5 describes its use in a 128GS/s ADC sampling front-end prototype. This prototype was created in such a way, so that it can be easily extended to a fully working 128GS/s time-interleaved SAR ADC, which had been the initial purpose of this M.A.Sc. thesis. However, due to the complexity of such an endeavor, the successful completion of this goal required far more time than the duration of a M.A.Sc. degree, and so it was decided that for the purposes of this thesis it was sufficient to focus on the sampling front-end part of the system. The experimental results of the low-noise broadband amplifier, the THA and the ADC front-end are discussed in Chapter 6. Finally, Chapter 7 draws the main conclusions of the thesis and shows how the THA could be used in a future high-speed ADC.
Chapter 2

Building Blocks for High-Speed A/D Converters

In this section, the basic building blocks that constitute the backbone of modern fiber optic receivers will be discussed. Among the blocks of the receiver, the broadband low-noise amplifier and the track-and-hold amplifier play a critical role in the sensitivity and sampling rate of the overall system and will be the focus of this chapter.

2.1 Broadband Low-Noise Amplifier

An essential part of any electro-optical communication system is the broadband amplifier. Its main purpose is to amplify the weak signal at the receiver’s input with minimal distortion and low noise. Therefore, it is important to provide sufficient gain and bandwidth, while maintaining low noise figure and constant group delay. Therefore, the main specifications of such amplifiers can be summarized in the following:

- Wide bandwidth starting from DC as quantified by $S_{21}$
- Adequate linearity as quantified by the 1dB compression point ($P_{1dB}$), in order to support complex modulation schemes such as 16-QAM,
- Low-noise operation as quantified by NF or equivalent input referred voltage ($V_{2n, in}^2$) or current ($I_{n, in}^2$), in order to detect and amplify weak signals.

If the amplifier is to be used in a 50Ω environment, it is also important to provide an input impedance matched to 50Ω (quantified by $S_{11}$).

2.1.1 Useful Results from the Theory of Noisy Linear Networks

Fig. 2.1 shows two noisy two-ports connected in parallel. According to [33, chapter 3] the parallel connection of two linear networks represented by their Y-parameter matrices $Y_a$ and $Y_f [34]$ under the
assumption that the main amplifier, $Y_a$, is unilateral yields the following results:

$$Y_{SOPT} = \sqrt{G_{sopta}^2 + \frac{G_{uf}}{R_{na}} + 2G_{cora}\Re(Y_{11f}) + \Re^2(Y_{11f}) + \frac{|Y_{corf} - Y_{11f}|^2R_{nf}}{R_{na}} + j[B_{sopta} - \Im(Y_{11f})]}$$

(2.1)

$$F_{MIN} = 1 + 2R_{na}[G_{cora} + G_{sopt} + \Re(Y_{11f})] \geq F_{MINa}$$

(2.2)

where $Y_{sopt}$ is the optimum noise admittance of the parallel combination of the two two-ports, $G_{sopta}$ the noise conductance of $Y_a$, $G_{cora}$ the correlated noise conductance of $Y_a$, $R_{na}, R_{nf}$ the noise resistances of $Y_a$ and $Y_f$ respectively, $B_{sopta}$ the optimum noise susceptance of $Y_a$ and $F_{MINa}, F_{MIN}$ the minimum noise factor of $Y_a$ and of combined two-port, respectively.

From equations 2.1 and 2.2 the following observations can be made [33]:

- Employing a lossless feedback network, $Y_f$, does not degrade the noise factor of the parallel combination.

- By using lossy feedback, $Y_{sopt}$ increases, meaning that the optimum noise impedance, $Z_{sopt}$, decreases. This proves to be very useful, when the signal source impedance is lower than the noise impedance of the original two-port. At the same time, the input impedance is also lowered due to the shunt negative feedback and so the two-port can be noise and impedance matched with the lowest possible current consumption.

- Lossy feedback lowers the imaginary and raises the real part of $Y_{sopt}$, resulting in decreased noise quality factor of the noise admittance, $Q = \frac{B_{sopt}}{G_{sopt}}$ and eventually in broadband noise matching.

![Fig. 2.1: Parallel connection of two noisy two-ports with its equivalent input referred correlated voltage and noise sources.](image)

A very important conclusion from the above listed points is that resistive (lossy) negative feedback is an effective way to achieve broadband input and noise matching. This comes in contrast with tuned LNAs that employ lossless components such as inductors to achieve low-noise operation in a very narrow band of interest. Hence, a suitable broadband, lumped topology is a simple amplifier with shunt (transimpedance) feedback as the one shown in Fig. 2.2, where $C_{in}$ represents all capacitance associated with the input node.
For this configuration it can be shown [33, chapter 3] that:

\[ Y_{SOPT} = \sqrt{G_{sopta}^2 + 2 \frac{G_{cora}}{R_F} + \frac{1}{R_{na}R_F} + \frac{1}{R_F} + jB_{sopta}} \]  \hspace{1cm} (2.3)  

\[ F_{MIN} = 1 + 2R_{na} \left( G_{cora} + G_{sopt} + \frac{1}{R_F} \right) > F_{MINa} \]  \hspace{1cm} (2.4)  

Equation (2.4) shows that the minimum noise figure is increased because of the resistor in the feedback path, \( R_F \), and its contribution can be greatly minimized by using large values. Note, however, that a large resistor value may reduce the bandwidth of the amplifier, because the input time constant \( R_FC_{in} \) may become the dominant pole slowing down the circuit. In general, the noise performance of a shunt-shunt amplifier is dominated by the feedback resistor, \( R_F \), at low frequencies and by the noise of the device used to implement the forward amplifier at high frequencies, as illustrated in Fig. 2.3.

To come up with an optimal design, the feedback resistor value must be chosen sufficiently large to minimize the noise figure at low frequencies (without sacrificing the bandwidth of course) and the sizing and biasing of the device must be optimized to keep the noise low close to the desired 3dB frequency. With the help of equations (2.1) and (2.3), it can be shown that there is an optimal device size at which the noise impedance becomes 50\( \Omega \) [33]. As far as biasing is concerned, it has been verified experimentally that the there is an optimal bias current density, \( J_{OPT} \), that minimizes the noise figure [35] as shown from the measured data in Fig. 2.4. These important observations lead to an optimal design methodology for broadband, low-noise amplifiers, which will be applied in the case of two TIAs demonstrated in Chapter 3.
Fig. 2.3: Input referred noise (adapted from [33]). $C_T$ includes all capacitances at the input of the amplifier.

Fig. 2.4: $f_T$, $f_{MAX}$ and $NF_{MIN}$ versus bias current density for a SiGe HBT device [36].

2.1.2 Low-Noise Broadband Topologies

For fiber-optic applications the amplifier must provide a wide bandwidth starting from DC, so tuned low-noise topologies such as inductive degeneration [37] that are widely used in narrow-band applications, are excluded. In addition, the amplifier must be low-noise and provide enough gain to minimize the noise contribution of subsequent stages in the receiver chain. In the remainder of this section some popular low-noise topologies for broadband amplification will be discussed. Even though the topologies shown consist primarily of bipolar devices, they could equally be made out of MOS devices. The choice of bipolar serves only as an example in this case to analyze the various topologies and does not affect the validity of the conclusions drawn. Specific information on the technology of choice is provided in Chapter 3.

Common Base

The schematic of the common-base amplifier is plotted in Fig. 2.5. In order to provide 50Ω input matching, the resistance seen looking at the emitter must be equal to 50Ω (neglecting the parasitic
emitter resistance):

\[ R_s = \frac{1}{g_m} = 50 \, \Omega \]  \hspace{1cm} (2.5)

meaning that a \( g_m \) of 20 mS is necessary for input matching. Given the fact that for a bipolar device \( g_m = I_C/V_T \), where \( V_T = kT/q \) is the thermal voltage equal approximately to 25 mV, a bias current of 0.5 mA is required. Due to the dependence of the input matching on \( g_m \), the input linear swing of this topology is limited to 0.5 mA \( \times 50 \, \Omega = 25 \, \text{mV}_{pp} \). Note that the emitter resistor, \( R_E \), must be much greater than \( 1/g_m \) so as not to affect the input matching. Since the input is matched to 50 \( \Omega \) the overall voltage gain is [37]:

\[ A_v = \frac{1}{2} g_m R_L = \frac{R_L}{2R_s} \]  \hspace{1cm} (2.6)

Using the open circuit time constant technique (OCTC) [13], the input and output time constants can be identified:

\[ \tau_1 = R_{in} C_{in} = (R_s + r_b) C_{be} \]  \hspace{1cm} (2.7)
\[ \tau_2 = R_{out} C_{out} = R_L (C_{cs} + C_{bc}) \]  \hspace{1cm} (2.8)

note that \( \tau_1 \) can be rewritten as: \( \tau_1 = \frac{g_m}{C_{be}} > 2\pi f_T \). Because both time constants are small (\( \tau_1 \) exceeds \( f_T \) and \( C_{cs} \) is small), this topology can achieve large bandwidths. The 3-dB bandwidth can be approximated by:

\[ f_{3dB} = \frac{1}{2\pi(\tau_1 + \tau_2)} \]  \hspace{1cm} (2.9)

At this point, the input impedance and gain of this stage can be found as a function of frequency.
(s = jω):

\[ Z_{in}(s) = \frac{1}{g_m / sC_{be}} = \frac{1}{gm + sC_{be}} \]  \hspace{1cm} (2.10)

\[ A_v(s) = \frac{R_L / R_s}{s(C_{bc} + C_{cs})} = \frac{g_m R_L}{1 + sR_L(C_{bc} + C_{cs})} \]  \hspace{1cm} (2.11)

Before deriving the noise figure of this amplifier, let us remind ourselves of the noise sources in a bipolar device. The bipolar transistor has thermal noise due to its terminal resistances and shot noise associated with the base and collector current, as shown in Fig. 2.6. Generally, in low-noise circuits at low frequencies the collector shot noise, base shot noise, and base resistance thermal noise dominate [37]:

\[ I_{2n,c} = 2qI_C = 4kTg_m \]  \hspace{1cm} (2.12)

\[ I_{2n,b} = 2qI_B \]  \hspace{1cm} (2.13)

\[ V_{2n,b} = 4kT\tau_b \]  \hspace{1cm} (2.14)

The collector shot noise from eq. 2.12 can be also referred to the input as an equivalent noise voltage source by dividing by \( g_m^2 \):

\[ V_{2n,c}^2 = \frac{4kT}{2g_m} \]  \hspace{1cm} (2.15)

In order to calculate the noise figure, the output noise of the circuit, \( V_{n,o}^2 \), is divided by the gain from \( V_{in} \) to \( V_{o} \), normalized to the source resistance \( 4kTR_s \) and a 1 is added to the result [37].

\[ V_{n,o}^2 = 4kT(r_b + \frac{1}{2g_m})(\frac{R_L}{R_s})^2 + 4kTR_L \]  \hspace{1cm} (2.16)

\[ NF = 1 + \frac{V_{n,o}^2}{4kTR_s(\frac{R_L}{2R_s})^2} = 1 + \frac{4(r_b + \frac{1}{2g_m})}{R_s} + \frac{4R_s}{R_L} \]  \hspace{1cm} (2.17)

Assuming a gain of 5 (14 dB) and remembering that \( R_s = 1/g_m \) the noise figure can be calculated from 2.17 equal to 5.3 dB (neglecting the effect of the base resistance).
At high frequencies, however, the correlation between the base and collector shot noise cannot be ignored [33]. A detailed noise equivalent circuit of an HBT device that captures correlation as well as access resistance thermal noise is shown in Fig. 2.7.

The noise equations now include the correlation between the base and collector noise currents [38]:

\[
\langle v^2_{n,bx} \rangle = 4kTR_{bx} \tag{2.18}
\]

\[
\langle v^2_{n,bi} \rangle = 4kTR_{bi} \tag{2.19}
\]

\[
\langle v^2_{n,b} \rangle = 4kTR_c \tag{2.20}
\]

\[
\langle v^2_{n,e} \rangle = 4kTR_e \tag{2.21}
\]

\[
\langle i^2_{n,c} \rangle = 2qI_C \tag{2.22}
\]

\[
\langle i^2_{n,b,} \rangle = 2q(I_B + |1 - e^{-j\omega\tau_n}|^2 I_C) \tag{2.23}
\]

where \(\tau_n\) is a modelling parameter used to express the statistical correlation between the base and collector noise currents given by [38]: \(\langle i_{n,b,}i_{n,c}^* \rangle = 2q(e^{-j\omega\tau_n} - 1)I_C\), where the asterisk denotes the complex conjugate quantity.

The device in Fig. 2.7 can be studied as a two-port and its noise properties can be completely characterized by an equivalent voltage and current noise source connected at its input as illustrated in Fig. 2.8. In order to bring the noisy 2-port to the form shown in Fig. 2.8, the following algorithm is used [33]: the inputs and outputs of the original circuit are shorted and the output short-circuit current is set equal to that of the final circuit. This way \(v_n\) is obtained. Keeping the inputs and outputs of the original circuit open, the open-circuit output voltage is assumed equal to that of the final circuit yielding the expression for \(i_n\).

Using the noise admittance formalism [33], the correlation of the input-referred noise sources of the
two-port is expressed through the admittance correlation parameter: $Y_{\text{COR}} = G_{\text{COR}} + jB_{\text{COR}}$:

\[ i_n = i_u + Y_{\text{COR}} v_n \]  
\[ Y_{\text{COR}} = \frac{\langle i_n v_n^* \rangle}{\langle v_n^2 \rangle} \]

where $i_u$ represents the uncorrelated and $i_c$ the correlated part of the current, respectively. The noise resistance, $R_n$, and noise conductance, $G_n$, are defined as follows [39]:

\[ R_n = \frac{\langle v_n^2 \rangle}{4kT\Delta f} \]
\[ G_n = \frac{\langle i_u^2 \rangle}{4kT\Delta f} \]

where $\Delta f$ is the bandwidth of interest. In the case of HBT devices the above parameters become [40]:

\[ R_n \approx \frac{V_T}{2J_C w_E l_E} + r_B + r_E \]
\[ G_n \approx \frac{J_C w_E l_E}{2V_T^2} \left( \frac{f}{f_T} \right)^2 \]
\[ Y_{\text{COR}} \approx \frac{f}{f_T R_n} \]

where $V_T$ is the thermal voltage, $J_C$ is the collector current density, $r_B$ is the base resistance, $r_E$ the emitter resistance, $w_E$ and $l_E$ the emitter width and length, respectively. Using $G_n$, $R_n$, $G_{\text{COR}}$, and $B_{\text{COR}}$ the noise of the 2-port can be fully described. When the 2-port is driven by a source with an
admittance $Y_S = G_S + jB_S$ as illustrated in Fig. 2.9, its noise factor is given by 39:

$$F = 1 + \frac{G_n}{G_S} + \frac{R_n}{G_S}[(G_{COR} + G_S)^2 + (B_{COR} + B_S)^2]$$

$$= F_{MIN} + \frac{R_n}{G_S}|Y_S - Y_{SOPT}|^2$$

(2.31)

$$Y_{SOPT} = G_{SOPT} + jB_{SOPT} = \sqrt{\frac{G^2_{COR}}{R_n} + \frac{G_n}{R_n}} - jB_{COR}$$

(2.32)

$$F_{MIN} = 1 + 2R_nG_{COR} + 2R_n\sqrt{\frac{G^2_{COR}}{R_n} + \frac{G_n}{R_n}}$$

(2.33)

Equation (2.31) reveals that the circuit achieves its minimum noise factor ($F_{MIN}$), when its input is matched to $Y_{SOPT}$. In the case of the common-base amplifier depicted in Fig. 2.5, the input referred noise current is:

$$i^2_n = i^2_{n_b} + i^2_{n_c} + \frac{4kT}{r_E} + \frac{4kT}{R_L}$$

(2.34)

finding the short-circuit output current as described previously, the input referred noise voltage, $v^2_n$, can be calculated and via (2.25) the admittance correlation, $Y_{COR}$. Finally, the noise factor can be calculated by (2.26)-(2.31).

### Regulated Cascode

The regulated cascode is displayed in Fig. 2.10. It is also known as noise cancelling topology because the noise of the transconductor device, Q1, can be theoretically canceled when the output is taken differentially. The analysis is discussed in [37] and its main points are summarized here. If the noise of Q1 is referred to its base as an equivalent noise voltage, then this noise sees a common-emitter path to node A (inverted polarity) and an emitter-follower path to node X (no inversion). The noise at X arrives at the common emitter of Q2 and appears inverted at node B. Thus, the noise of Q1 appears as a common-mode signal (same polarity) at both outputs A and B and gets canceled in differential operation. The signal on the other hand, sees a common-base through Q1 and a common-emitter through Q2 and gets amplified at the output when taken differentially.

Similarly to the common-base amplifier the input condition for 50Ω matching is:

$$R_s = \frac{1}{g_{m1}} = 50 \Omega$$

(2.35)
assuming that $R_E >> 1/g_{m1}$. Since the input is matched, the gain from the base of Q1 to node X is approximately equal to $g_{m1}(R_S//1/g_{m1})$. Thus, only half of the noise voltage of Q1 appears at node X and this noise gets amplified producing an output noise voltage at B equal to $-g_{m2}R_2V_{n1}$. The noise of Q1 also sees a degenerated common-emitter path (with $R_s//R_E$ as the degeneration resistance) to A and so the output noise at that node is: $-g_{m1}R_1V_{n1}/2$. For perfect cancellation the following relation must hold [37]:

$$g_{m1}R_1 = g_{m2}R_2$$

(2.36)

For differential operation the swing at both output nodes must be symmetrical requiring that $R_1 = R_2$ and as a result of the noise cancellation condition (2.36) $g_{m1} = g_{m2}$. Now that the noise of Q1 has been cancelled, the noise figure depends on Q2, Q3, Q4, $R_1$, and $R_2$:

$$\bar{V}_{n,A}^2 = 4kTR_1$$

(2.37)

$$\bar{V}_{n,B}^2 = 4kTR_2 + 4kT\frac{g_{m1}}{2}R_1^2 + 2 \times 4kT\frac{g_{m2}}{2}R_2^2$$

(2.38)

where the last two terms of (2.38) account for the shot noise of Q2, Q3, and Q4, respectively. The single-ended to differential gain of the circuit is given by [37]:

$$A_v = \frac{g_{m1}R_1 + g_{m2}R_2}{2} = g_{m1}R_1 = \frac{R_1}{R_s}$$

(2.39)

Dividing the total output noise by the gain from (2.39) and adding 1 to the result, the low-frequency noise figure can be computed as:
\[ NF = 1 + \frac{4kTR_1 + 4kTR_2 + 4kT \frac{g_{m1}}{2} R_2^2 + 2 \times 4kT \frac{g_{m2}}{2} R_2^2}{4kTR_s} \left( \frac{R_1}{R_s} \right)^2 \] (2.40)

\[ = 1 + 2 \frac{R_s}{R_1} + 1.5 \] (2.41)

For a gain of 14 dB equation (2.41) gives a NF of 4.62 dB which is somewhat lower from that of the common-base amplifier derived in the previous section (again the effect of \( r_b \) has been omitted). Note that due to the input matching condition (2.35), the maximum input linear swing is the same for both topologies.

For high-frequency noise calculations the equivalent input referred noise current source is calculated as follows:

\[ i_{n}^2 = i_{b1}^2 + i_{b2}^2 + i_{b3}^2 + i_{c1}^2 + i_{c3}^2 + \frac{4kT}{r_E} + \frac{4kT}{R_1} + \left( 1 + \frac{\omega^2 C_m^2}{g_{m1}} \right) (i_{c2}^2 + i_{c4}^2 + i_{c4}^2 + \frac{4kT}{R_2}) \] (2.42)

The last term accounts for the noise current of \( R_2 \), \( Q2 \), and \( Q4 \) when referred to the input through the gain of the common-emitter transistor \( Q2 \). Using the algorithm described earlier, the input referred noise voltage can be determined and ultimately the noise factor via (2.31).

As far as the bandwidth is concerned the following time constants can be identified:

\[ \tau_1 = R_{in}C_{in} = (R_s + r_b)(C_{bc1} + C_{bc2} + 2C_{bc2}) \] (2.43)

\[ \tau_2 = R_{out}C_{out} = R_1(C_{cs1} + C_{bc1}) \] (2.44)

It is obvious that the bandwidth is somewhat lower in this case because \( \tau_1 \) is larger this time, however the use of the regulated cascode topology proves advantageous in multichannel applications where crosstalk is a serious concern due to improved power supply rejection [41].

The input impedance and gain as a function of frequency are given by:

\[ Z_{in}(s) = \frac{1}{R_s} + \frac{1}{s(C_{bc1} + C_{bc2} + 2C_{bc2})} = \frac{1}{g_{m1} + s(C_{bc1} + C_{bc2} + 2C_{bc2})} \] (2.45)

\[ A_v(s) = \frac{R_1}{R_s[1 + sR_1(C_{bc1} + C_{cs1})]} = \frac{g_{m1}R_1}{1 + sR_1(C_{bc1} + C_{cs1})} \] (2.46)

**HBT-source follower**

The schematic of the HBT-source follower is given in Fig. 2.11. Note that an HBT emitter follower cannot be used instead of the low-V\(_t\) source follower in this case, as it would push the \( V_{CE} \) of the main HBT beyond breakdown (in modern BiCMOS processes the maximum \( V_{CE} \) is less than 2 \times \( V_{BE} \) [42]).

Thanks to the shunt-shunt negative feedback the amplifier can be independently optimized to achieve low noise figure and input matching in contrast to the topologies discussed previously. Since the input match does not solely depend on the \( g_m \), this gives an extra degree of freedom to configure the input linear swing of this circuit by choosing the voltage drop on the emitter degeneration resistor, \( R_E \). A good starting point for choosing \( R_E \) is given by [43]:

19
where \( r_E \) is the parasitic emitter resistance and \( V_{IIP3} \) is the input swing that meets the \( IIP_3 \) linearity requirements. To a first-order analysis let’s assume that the gain of the source follower is ideally 1. If the amplifier is matched at the input, the gain from \( V_{in} \) to \( V_o \) can be approximated as:

\[
A_v \approx -\frac{R_L}{2R_E}
\] (2.48)

In reality, the gain of the source follower will be less than unity. To be more precise, the gain of the source follower can be found from [13]:

\[
A_{v, SF} \approx \frac{g_{mMOS}}{g_{mMOS} + g_s + g_{ds}}
\] (2.49)

where \( g_s \) and \( g_{ds} \) account for the body effect and output admittance \( 1/g_{ds} = r_{ds} \). In modern CMOS processes \( g_{mMOS} \) is of the order of 1.5 mS/\( \mu \)m and \( g_{ds} \) of the order of 0.18 mS/\( \mu \)m [3], while \( g_s \) can be roughly 10% of \( g_{mMOS} \) [13]. Therefore, a more realistic value for the gain of the source follower according to (2.49) is 0.8. Nonetheless, in the derivations to come, we will assume for simplicity a gain of 1.

Negative feedback reduces the input resistance to:

\[
R_{in} = R_s = \frac{R_F}{1 - A_v}
\] (2.50)

For a gain of 5, a feedback resistor value of 300 \( \Omega \) guarantees that the input is matched to 50 \( \Omega \).

The high-frequency performance of this topology is governed mainly by three time constants: one at the input (\( \tau_1 \)), one at the intermediate node between the HBT and MOS stages (\( \tau_2 \)), and one at the output (\( \tau_3 \)). \( \tau_3 \) is very small since the output resistance is approximately \( 1/g_{mMOS} \) and can thus be considered negligible.
\[
\tau_1 = R_{in}C_{in} = (R_s + r_b)[C_{bc} + (1 - A_v)C_{bc}]
\] (2.51)

\[
\tau_2 = R_L(C_{cs} + C_{bc} + C_{gs} + C_{gd})
\] (2.52)

Since Q1 should be biased at low current density for optimal noise performance, \(C_{bc}\) can be quite large. The input time constant is further exacerbated by the Miller effect of \(C_{bc}\). What is more, in order to provide adequate gain and linearity the value of \(R_L\) tends to be big and sees an equivalent capacitance from the collector of the HBT and the gate of the MOS device. Since the source follower must be able to drive a fairly large resistive load consisting of \(R_F\) and \(R_s\) \[43\], it needs sufficient bias current to ensure that \(1/g_m\) is kept small. This leads to a wide source follower device with considerable capacitance at its gate limiting the 3-dB bandwidth of this topology.

In the frequency domain the gain can be expressed as:

\[
A_v(s) = \frac{R_L}{1 + sC_{in}}A_{v, SF}(s) = -\frac{R_L}{R_E[1 + sR_L(C_{cs} + C_{bc} + C_{gs} + C_{gd})]}A_{v, SF}(s)
\] (2.53)

where \(A_{v, SF}(s)\) is the gain of the source follower given by \[13\]:

\[
A_{v, SF}(s) = \frac{g_mC_{gs}s}{s[C_{gs} + C_{sb}] + g_m[1/\left(\frac{R_s + R_F}{(1/g_s)}\right)]}
\] (2.54)

Substituting (2.54) into (2.53) results in a second-order transfer function with a high frequency zero at \(\omega_z = \frac{g_mC_{gs}}{C_{gs}}\). The input impedance is calculated by the parallel combination of the impedance due to \(C_{in}\) in (2.51) and the input resistance given by (2.50):

\[
Z_{in}(s) = \frac{R_F}{1 - A_v(s)} = \frac{R_F}{sC_{in} + (1 - A_v)C_{bc}R_F + 1 - A_v(s)}
\] (2.55)

resulting again in a second-order transfer function. The total output noise assuming unity gain for the source follower is given by:

\[
V_{n,o}^2 = 4kTR_F + \frac{4kT\gamma}{g_m^MOS} (1)^2 + 4kTR_L(1)^2 + \frac{4kT}{R_E} R_s^2(1)^2 + 4kT \frac{1}{2g_mHBT} \left(\frac{R_L}{R_E}\right)^2(1)^2
\] (2.56)

where \(\gamma\) is "excess noise coefficient", which is equal to 2/3 for long-channel devices but can rise up to 2 for nano-scale MOSFETs \[37\]. The low-frequency noise figure is given by:

\[
NF = 1 + \frac{4kT(R_F + \frac{\gamma}{g_m^MOS} + R_L + \frac{R_s^2}{R_E} + \frac{g_mHBT}{2} R_s^2)}{4kTR_s\left(\frac{R_L}{2R_E}\right)^2}
\] (2.57)

Assuming a degeneration resistor of 15 \(\Omega\) for an input linear swing of 150 mV (HBT biased at 10 mA) and a gain of 5, a load resistor of 150 \(\Omega\) is required. Given a relatively high \(g_m\) of 60 mS, the noise figure of this circuit is roughly 3.9 dB.

Based on the theory of negative feedback in linear noisy networks and accounting for the statistical correlation between the collector and base current noise at high frequencies, it can be shown that the
50Ω noise factor can be expressed as [33]:

\[
F = 1 + R_{n,Q1} R_s \left| Y_{COR,Q1} + \frac{1}{R_s} + \frac{1 - j\omega_f}{R_F(1 + \omega_f^2)} \right|^2 + G_{n,Q1} R_s + \frac{R_s}{R_F(1 + \omega_f^2)}
\] (2.58)

where \( R_{n,Q1}, Y_{COR,Q1}, G_{n,Q1} \) refer to the noise resistance, the correlation admittance and noise conductance of \( Q1 \), respectively, and \( \omega_f = \omega L_F / R_F \).

**Darlington Cascode**

The Darlington cascode topology has been demonstrated to achieve operation beyond 100 GHz [44] and is illustrated in Fig. 2.12. The transconductance of the Darlington pair is equal to that of \( Q2 \) and accounting for the degeneration induced by the \( R_{E2} \) and \( r_E \) (parasitic emitter resistance) is given by [44]:

\[
g_{mtot} = \frac{g_{m2}}{1 + g_{m2}(R_{E2} + r_E)}
\] (2.59)

The criterion for input matching depends on the value of \( R_F \) and the gain in the same way as in (2.50). With the input matched the gain is given by [44]:

\[
A_v = g_{mtot}(R_F/R_L)
\] (2.60)

In this case, the total output noise can be approximated by:

\[
\overline{V_{n,o}^2} = 4kT(R_F + R_L) + 4kT \left( \frac{2A_v}{2g_{m1}} \right)^2 + \left( \frac{4kT}{R_{E1}} + \frac{4kT}{R_{E2}} + 4kT \frac{g_{m2}}{2} + 4kT \frac{g_{m3}}{2} \right) R o^2
\] (2.61)

\[
R o = \frac{R_L}{(R_F + R_s)}/ \left[ 1/g_{mtot} \left( 1 + \frac{R_F}{R_s} \right) \right]
\] (2.62)

In (2.61) the first term accounts for the noise voltage of the feedback and load resistors, the second for the current shot noise of \( Q1 \) referred to the base, and the last term for the contributions of the current shot noise of \( Q2, Q3 \) and the noise current of \( R_{E2} \) flowing through the output impedance given by (2.62).

![Fig. 2.12: Darlington cascode amplifier.](image)
Note that the equivalent noise voltage at the base of Q1 sees twice the gain of \((2.60)\) (from the base of Q1 to \(V_o\)). The low-frequency noise figure can then be approximated by:

\[
NF = 1 + \frac{4kT(R_F + R_L) + 4kT\frac{2A_v}{2g_{m1}} + \left(\frac{4kT}{R_{E2}} + 4kT\frac{2m_2}{2} + 4kT\frac{2m_3}{2}\right) R_0^2}{4kT R_s (A_v)^2}
\] (2.63)

For the same conditions as before \((R_F = 300 \, \Omega, \, A_v = 5, \, \text{and} \, R_{E2} = 15 \, \Omega\), \((2.63)\) predicts a noise figure higher than 5 dB about 1 dB higher than the case of the HBT-source follower.

In order to estimate the high-frequency noise performance of the Darlington cascode amplifier, we first study this topology without its feedback network. Once we find the noise parameters of the forward amplifier, we can then apply the formulas from the theory of noisy linear networks, in order to include the effect of the feedback network. For this particular case, it is easier if the noise impedance formalism is used [33]. This time the noise voltage source at the input of the two-port network has a correlated component, \(v_c\) with the input noise current:

\[
v_n = v_u + v_c = v_u + Z'_{\text{COR}} i_n \tag{2.64}
\]

where \(Z'_{\text{COR}}\) is the correlation impedance defined as [33]:

\[
Z'_{\text{COR}} = R'_{\text{COR}} + jX'_{\text{COR}} = \frac{v_n i_n}{i_n^2} \tag{2.65}
\]

The Darlington pair can be thought of as an emitter follower (transistor Q1) driving a common-emitter cascode amplifier (transistors Q2-Q3). We assume that all noise sources of the cascode are referred to the base of Q2 via \(\overline{v}_{n2}\) and \(\overline{i}_{n2}\). In the same manner the emitter follower has all noise sources referred to its input (base of Q1) via \(\overline{v}_{n1}\) and \(\overline{i}_{n1}\). It can be shown [33] that the input referred equivalent noise sources of the Darlington cascode amplifier are given by:

\[
\overline{v}_n^2 = \overline{v}_{n1}^2 + |Z_{\text{COR1}}|^2 \overline{i}_{n1}^2 + \frac{1}{|A_{v,EF}|^2} |Z_{\text{in,EF}}|^2 (\overline{v}_{u2}^2 + |Z_{\text{out,EF}} + Z_{\text{COR2}}|^2 \overline{i}_{n2}^2) \tag{2.66}
\]

\[
\overline{i}_n^2 = \overline{i}_{n1}^2 + \frac{1}{|A_{v,EF}|^2} |Z_{\text{in,EF}}|^2 (\overline{v}_{u2}^2 + |Z_{\text{out,EF}} + Z_{\text{COR2}}|^2 \overline{i}_{n2}^2) \tag{2.67}
\]

where \(A_{v,EF}\) represents the gain of the emitter follower, \(Z_{\text{out,EF}}\) its output impedance, \(Z_{\text{in2}}\) the input impedance of the cascode amplifier, \(\overline{v}_{n1}\) the uncorrelated input referred voltage noise source of the emitter follower, \(\overline{i}_{n1}\) its input referred noise current source, \(Z_{\text{COR1}}\) its impedance correlation, \(\overline{v}_{n2}\) the uncorrelated input referred voltage noise source of the cascode, \(\overline{i}_{n2}\) its input referred current noise source, and \(Z_{\text{COR2}}\) the impedance correlation of the cascode. The noise conductance, \(G'_{\text{na}}\), and noise resistance, \(R'_{\text{na}}\), are defined as follows [33]:

\[
G'_{\text{na}} = \frac{\overline{i}_n}{4kT \Delta f} \tag{2.68}
\]

\[
R'_{\text{na}} = \frac{\overline{v}_n^2}{4kT \Delta f} \tag{2.69}
\]
The expressions for noise factor \( F_a \), optimum noise impedance \( Z_{SOPTa} \), and minimum noise factor \( F_{MINa} \) are provided below [33]:

\[
F_a = 1 + \frac{G'_n}{R_s} |Z_{COR} + Z_s|^2 + \frac{R'_n}{R_s} = F_{MIN} + \frac{G'_n}{R_s} |Z_s - Z_{SOPT}|^2
\]  
(2.70)

\[
F_{MINa} = 1 + 2G'_n(R'_{COR} + R_{SOPT})
\]  
(2.71)

\[
Z_{SOPTa} = R_{SOPT} + jX_{SOPT} = \sqrt{R'^2_{COR} + \frac{R'_n}{G'_n} X_{COR}^2}
\]  
(2.72)

After rigorous mathematical analysis, it can be proven that the noise parameters of (2.65), (2.68) and (2.69) depend mainly on the emitter follower device (Q1) [39]:

\[
G'_{na} \approx G_{n1}
\]  
(2.73)

\[
R'_{na} \approx R_{n1}
\]  
(2.74)

\[
Z_{COR} \approx Z_{COR1}
\]  
(2.75)

Translating the noise parameters from noise impedance to the noise admittance formalism and applying negative feedback through \( R_F \), the high-frequency noise parameters of the HBT-TIA become [33]:

\[
R_n \approx R_{na}
\]  
(2.76)

\[
G_n = G_{na} + \frac{1}{R_F}
\]  
(2.77)

\[
Y_{COR} = Y_{CORa} + 1/R_F
\]  
(2.78)

\( Y_{SOPT} \) and \( F_{MIN} \) of the complete Darlington cascode are given by (2.1) and (2.2), respectively.

The real advantage of this topology is in its bandwidth. The \( C_{be} \) of Q1 and Q2 appear in series and so the total input capacitance reduces to \((C_{bc1} + C_{bc2})/2\), similar to the \( f_T \) doubler concept [33]. In addition, the inclusion of the cascode device, Q3, reduces the Miller effect on Q2 improving the bandwidth. At the same time, it shields Q2 from breakdown by limiting its \( V_{CE} \) to less than \( 2 \times V_{BE} \).

\[
\tau_1 = R_{in} C_{in} = (R_s + r_b) \left( \frac{C_{bc1} + C_{bc2}}{2} + 2C_{bc} \right)
\]  
(2.79)

\[
\tau_2 = R_{out} C_{out} = (R_L/\|R_F\|)(C_{cs1} + C_{bc1} + C_{cs3} + C_{bc3})
\]  
(2.80)

The high frequency gain of the Darlington cascode amplifier can be approximated by:

\[
A_v(s) = \frac{g_{m10t}(R_L/\|R_F\|)}{1 + s(R_L/\|R_F\|)(C_{cs1} + C_{bc1} + C_{cs3} + C_{bc3})} A_{v,EF}(s)
\]  
(2.81)

where \( A_{v,EF}(s) \) is the high-frequency gain of the emitter follower that exhibits two poles and one zero.
as shown below [33] (ignoring for the moment $R_{E1}$):

$$A_{v,EF}(s) = \frac{g_{m1} + sC_{bc1}}{R_s(C_{be1}C_L + C_{bc1}C_L + C_{be1}C_{bc1}) + (C_{bc2} + 2C_{be2})s^2 + (g_{m1}R_sC_{be1} + C'_L + C_{be1})s + g_{m1}}$$

(2.82)

$$C_L = C_{be2} + 2C_{bc2}$$

(2.83)

Taking into account the frequency-dependent gain of (2.81) and the negative feedback through $R_F$, the input impedance can be approximated by:

$$Z_{in} = \frac{R_F}{1 - A_v(s)} \frac{1}{sC_{in}} = \frac{R_F}{1 - A_v(s) + s(C_{be1} + 2C_{bc2})}$$

(2.84)

The input impedance of the Darlington stage may look similar to (2.55), however in this case the input of the circuit is located at the base of emitter follower Q1. It can be shown [33] that the input impedance of an emitter follower is:

$$Z_{in,EF} = R_b + \frac{1}{j\omega C_{bc}} / \left[ \frac{1}{j\omega C_{be}} + \frac{1}{j\omega C_L} - \frac{g_m}{C_{be}C_L\omega^2} \right]$$

(2.85)

where $R_b$ represents the total resistance at the base and $C_L$ the load capacitance connected at the emitter ($C_{be1} + C_{be2} + 2C_{bc2}$ in this case). The last term in (2.85) indicates that the real part of the input impedance may become negative leading to instability and ringing in the step response, which may inhibit good input matching at high frequencies. A technique to ameliorate this instability problem is to bias the emitter follower below its peak $f_T$ current density, $J_{p,ft}$, and also connect a resistor instead of a current source at the emitter ($R_{E1}$).

**HBT-TIA**

The analysis of the HBT-TIA shown in Fig. [2.13] is very similar to the one presented for the Darlington cascode. In particular, the gain is given by (2.60). Because of the absence of the Darlington pair and cascode device, the bandwidth is not as high in this case, as confirmed by the increase of the input time constant, $\tau_1$:

$$\tau_1 = R_{in}C_{in} = (R_s + r_b)[C_{be} + (1 - A_v)C_{bc}]$$

(2.86)

$$\tau_2 = R_{out}C_{out} = R_L(C_{cs} + C_{bc})$$

(2.87)

The high-frequency gain in this case is:

$$A_v(s) = \frac{g_m(R_L/\|R_F\|)}{1 + s(R_L/\|R_F\|)((C_{cs1} + C_{bc1})}$$

(2.88)

Because of the existence of just one pole, this topology introduces less phase shift than the Darlington topology and remains stable across frequency. The input impedance across frequency can now be approximated by:

$$Z_{in} = \frac{R_F}{1 - A_v(s)} \frac{1}{sC_{in}} = \frac{R_F}{1 - A_v(s) + s[C_{be} + (1 - A_v)C_{bc}]}$$

(2.89)
Due to the use of a single device, however, this topology can operate from a lower power supply and achieve a lower NF than the Darlington cascode. The output noise at low frequencies in this case is given by:

$$V_{n,o}^2 = 4kT(R_F + R_L) + \left(\frac{4kT}{R_{E2}} + 4kT \frac{g_m}{2}\right) R_o^2$$  \hspace{1cm} (2.90)

$$V_{n,o}^2 = \frac{4kT}{R_L} - \frac{\overline{i_n^2}}{g_{m_{eff}}}$$  \hspace{1cm} (2.91)

where $R_o$ is given by (2.62). To estimate high-frequency noise parameters, the input referred current and voltage noise sources must be found first in the absence of the feedback network. Using the open-circuit output voltage and short-circuit current technique, we get:

$$\overline{i_n^2} = \frac{4kT R_L}{g_{m_{eff}}} + \frac{1}{Z_{in}} \frac{g_{m_{eff}} R_L^2}{2}$$  \hspace{1cm} (2.92)

$$\overline{i_n^2} = \frac{4kTR_L + 4kT R_L + \overline{i_n^2 b_{(g_{m_{eff}}/R_E)}} \frac{1}{2} g_{m_{eff}}^2 R_L^2 + \overline{i_n^2} R_L^2}{Z_{in} g_{m_{eff}} R_L^2}$$  \hspace{1cm} (2.93)

where $g_{m_{eff}} = g_m/[1 + g_m (r_E + R_E)]$, and $Z_{in}$ is the input impedance. The noise parameters of the forward amplifier $R_{na}$, $G_{na}$ and $Y_{CORa}$ can now be calculated from (2.25)-(2.31) and then applied to (2.76)-(2.78) to obtain the noise parameters of the HBT-TIA with feedback.

A collection of state-of-the-art low-noise, broadband amplifiers reported in the literature is provided on table 2.1. Judging from the data on the table, the CMOS LNA with shunt feedback of reference [45] achieves a 60GHz bandwidth, much lower compared to the rest of the amplifiers listed. The rest of the designs have at least 100GHz bandwidth making them suitable candidates for high-performance fiber-optic systems. Despite its high bandwidth, the InP design of reference [46] consumes the most power and does not report noise figure. On the other hand, the staggered gain approach of the 0.12µm SiGe solution presented in [47] results in a relatively high NF of about 11.3 dB. Finally, the Darlington-cascode implementation of [44] shows excellent bandwidth (100 GHz) at low-power (48 mW) with a simulated NF of 8 dB. Nonetheless, it requires two external bias-T networks to operate properly, rendering this solution hard (if not impossible) to integrate on a bigger system on chip.

Fig. 2.13: HBT-TIA schematic.
### Table 2.1: State-of-the-Art Broadband, Low-Noise Amplifiers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>10</td>
<td>102</td>
<td>4.425</td>
<td>2</td>
<td>73</td>
<td>±6</td>
<td>11.3&lt;sup&gt;1&lt;/sup&gt;</td>
<td>0.29</td>
<td>0.12 µm SiGe</td>
</tr>
<tr>
<td>45</td>
<td>9</td>
<td>60</td>
<td>14</td>
<td>1.2</td>
<td>12</td>
<td>-</td>
<td>8</td>
<td>-</td>
<td>45 nm SOI CMOS</td>
</tr>
<tr>
<td>46</td>
<td>14.5</td>
<td>100</td>
<td>7.31&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-2.5</td>
<td>145</td>
<td>±5.5</td>
<td>-</td>
<td>0.21</td>
<td>0.25 µm InP DHBT</td>
</tr>
<tr>
<td>44</td>
<td>12</td>
<td>110&lt;sup&gt;3&lt;/sup&gt;</td>
<td>9.1</td>
<td>2.1</td>
<td>48</td>
<td>±1</td>
<td>8(sim)</td>
<td>0.197</td>
<td>90 nm SiGe BiCMOS</td>
</tr>
</tbody>
</table>

1 based on NF plot provided on paper
2 based on differential gain of 20.5 dB
3 requires external bias-T for operation

### 2.2 Track-and-Hold Amplifier

The track-and-hold (THA) also known as sample-and-hold amplifier (SHA) is a circuit that stores a sample of the input signal for a finite amount of time. In its simplest form it consists of a sampling switch and a capacitor as shown in Fig. 2.14. When the switch is off, the output follows the input (track or acquisition mode). When the switch is off, the instantaneous value of the input voltage is kept on the capacitor (hold mode). The output waveform of the THA operation is depicted in Fig. 2.15. This output can be decomposed into two periodic waveforms of period <i>T<sub>s</sub></i>: one in which the input, <i>x(t)</i>, is multiplied by a square wave (<i>y_1(t)</i>) and one square wave whose amplitude equals the held voltage at the output (<i>y_2(t)</i>)<sup>[14]</sup>:

\[
y_c(t) = y_1(t) + y_2(t) \tag{2.94}
\]

\[
y_1(t) = x(t) \left[ \Pi \left( \frac{2t}{T_s} - \frac{1}{2} \right) * \sum_{k=\infty}^{k=-\infty} \delta(t - kT_s) \right] \tag{2.95}
\]

\[
y_2(t) = \left[ x(t) \sum_{k=-\infty}^{k=\infty} \delta(t - kT_s - \frac{T_s}{2}) \right] * \Pi \left( \frac{2t}{T_s} - \frac{1}{2} \right) \tag{2.96}
\]

In the frequency domain<sup>[2.94]</sup> is written as<sup>[14]</sup>:

\[
Y_c(f) = \sum_{n=-\infty}^{n=\infty} e^{-j\pi n/2} \frac{\sin(n\pi/2)}{n\pi} X \left(f - \frac{n}{T_s}\right) + e^{-3j\pi fT_s/2} \frac{\sin(\pi fT_s/2)}{\pi fT_s} \sum_{n=-\infty}^{n=\infty} X \left(f - \frac{n}{T_s}\right) \tag{2.97}
\]

![Fig. 2.14: The simplest track-and-hold circuit](image)

Sample-and-hold amplifiers have become an essential component of any digitizing system. For ex-
ample, several SHA can be used in parallel to extend the sampling rate of ADCs, a technique called time-interleaving, as shown in Fig. 2.16. The use of SHAs also helps mitigate phenomena that limit the ENOB of A/D converters. In particular [48]:

- Helps minimize capacitive feedthrough on the resistive ladder of a flash ADC that can create an input-dependent offset in the comparators. As shown in Fig. 2.17, the input signal may couple through the gate-source capacitance of the previous stage and disturb the reference voltage level of the resistive ladder used by the comparators in the ADC.

- Reduces harmonic distortion caused by the non-linear capacitance of the comparators as shown in Fig. 2.17.

- Eliminates errors (bubbles) in the thermometer code of flash ADCs caused by timing mismatches as illustrated in Fig. 2.18.

- Minimizes distortion due to limited bandwidth [49]. For example, if the input stage of a comparator has insufficient bandwidth, then the signal delay will depend on its slew rate. Thus, a full-scale input sampled near the peak will exhibit greater delay than when sampled near the midpoint resulting in a distorted sample as illustrated in Fig. 2.18.

Fig. 2.15: Output waveform of ideal THA [14].

Fig. 2.16: SHAs used in a time-interleaved ADC architecture [48].

Fig. 2.17: SHAs used in a time-interleaved ADC architecture [48].
Fig. 2.17: (a): Capacitive feedthrough in a flash ADC and (b): harmonic distortion due to nonlinear capacitance of the comparators.

(a) Bubble errors due to timing mismatch.  (b) Distortion of the sampled waveform due to inadequate bandwidth.

Fig. 2.18: Errors in ADCs due to timing and bandwidth limitations.
2.2.1 THA performance metrics

The main performance metrics of track-and-hold circuits are listed below:

- **sampling pedestal** or **hold step**: Every time the circuit switches from track to hold mode, there is an error in the output voltage due to charge injection and other phenomena. This error must be kept small and be signal independent to avoid distortion [13].

- **hold mode isolation**: Even when the switch is turned off, there is some signal feedthrough to the output due to capacitive coupling [13]. This effect is more pronounced at high input frequencies.

- **acquisition time**: The time required after entering track mode for the THA to settle within a specified error band to its final value, when a full-scale transition is applied. The acquisition time depends on the bandwidth and slew rate of the input and output buffers of the THA, the value of the hold capacitor, and switch losses [14].

- **hold settling time**: The time required for the THA to settle to its final value after the hold command has been applied. This time mainly depends on the settling behavior of the output buffer following the switch [14].

- **droop rate**: Is the rate of discharge of the voltage stored on the hold capacitor during hold mode. At low frequencies this effect is often negligible in CMOS but prominent in bipolar implementations due to leakage currents from the finite base current of bipolar devices [13]. At high frequencies, however, even the gate current becomes significant and cannot be neglected anymore.

- **aperture jitter** or **aperture uncertainty**: The error due to different effective sampling time from one sampling instant to the next [13]. This error is more pronounced in high-speed THAs, since the large slope of the high-frequency input signal results in significant errors in the sampled output in the presence of sampling clock jitter. The interested reader is referred to [50] for more information.

Fig. 2.19 illustrates some of the errors discussed above.
2.2.2 THA topologies

Track-and-hold designs can be classified into open-loop and closed-loop topologies. An example of an open-loop topology is displayed in Fig. 2.20. In Fig. 2.20, the switch is controlled by the switch driver circuit. When the hold command is applied, the switch is disconnected and the input is sampled on the hold capacitor, $C_H$. The input is buffered by the input buffer which may or may not provide gain. The performance of the input buffer determines the linearity and acquisition time of the THA, while suppressing clock feedthrough that can corrupt the input signal. The output is buffered by an output amplifier which affects the hold settling time and droop rate of the THA. The switch can be implemented using MOSFETs, diodes, CMOS, or bipolar devices. For high-speed designs, diodes and bipolar switches are favored because they require lower clock amplitudes producing sharper transitions and thus more accurate sampling instants. These type of switches can be driven from low-noise CML or ECL circuits resulting in less sampling jitter than MOS implementations [14]. In addition, diode switches exhibit a lower on-resistance and less charge injection than their MOS counterparts [14].

An example of a closed-loop THA is shown in Fig. 2.21. When S1 is closed, S2 is open and the output tracks the input (with a gain determined by $\frac{R_F}{R_I}$ in this case) and any DC errors of the switches are suppressed by the negative feedback [50]. The track mode behavior depends on the slew rate of the opamp. When the hold command is issued, S1 opens storing the instantaneous input voltage on the hold capacitor, and S2 closes shielding the hold capacitor from signal feedthrough while maintaining a constant input impedance. Due to the virtual ground of the opamp, the voltage across S1 is constant during track mode. Thus, S1 injects a constant charge onto $C_{HOLD}$ when entering hold mode generating a signal-independent pedestal error at the output, that does not introduce any nonlinearity. This error appears as an offset voltage and can be corrected with proper calibration techniques [14]. The amplitude of the pedestal error as well as clock feedthrough can be greatly reduced by employing differential switching techniques, since both appear as a common-mode signal [50].

In general, closed-loop architectures provide greater accuracy but can become unstable if enough...
phase margin is not guaranteed in the feedback loop. What is more, they exhibit a slow response and low input bandwidth due to the large gain opamp inside the loop. On the other hand, open-loop topologies are faster providing much higher bandwidth, but their precision is usually limited to about 8 bits due to the linearity of the input and output buffers [14]. Hence, the latter is the topology of choice for high-speed ADC designs.

An interesting implementation of an open-loop THA employing a diode bridge is displayed in Fig. 2.22. During track mode, $V_{trk}$ is high and Q2 carries a current of $2I_B$ (Q1 is turned off). Half of this current is supplied by the current source connected at node $V_2$ and the other half by the diode bridge consisting of $D_1 - D_4$. In this case $D_5$ and $D_6$ are off and the conducting bridge provides a low resistance path from $V_{in}$ to $V_3$. Each diode has an on-resistance of $r_d = \frac{V_T}{I_B/2}$.

During hold mode $V_{hld}$ is held high and Q1 carries $2I_B$ while Q2 is off. Half of the current comes from the top current source and the other half from $V_2$ through $D_5$ and $D_6$ which are now conducting. Since $V_2$ is one diode drop above $V_{out}$, $D_1 - D_4$ become reverse-biased disconnecting the hold capacitor from the input. An advantage of this topology is the the fact that charge injection is minimized because
the charge of $D_1 - D_2$ cancels the charge of $D_3 - D_4$ (to a first-order approximation) [13]. Because of the bipolar pair (Q1-Q2) the current can be switched fast with relatively low clock amplitudes. The use of Schottky diodes, which have no minority carrier storage, can improve the turn-off time of the diode bridge [13]. Finally, since nodes $V_1$ and $V_2$ are held constant and equal to $V_{out}$ by $D_5 - D_6$ signal feedthrough originating from capacitive coupling through the bridge has negligible effect on the stored voltage of the hold capacitor [13].

A disadvantage of the diode bridge is the rather high supply voltage required for its operation. To maintain a high output impedance the current sources of Fig. 2.22 require about 0.4 V assuming a wide-swing cascode current mirror. For high-speed switching Q1, Q2 need at least a $V_{CE}$ of 0.8 V. Assuming a diode drop of about 0.8 V, a supply higher than $V_{CC} = 2 \times 0.8 + 0.8 + 2 \times 0.4 = 3.2 V$ is necessary. To accommodate input swings of the order of 600 mV$_{pp}$, a 3.8 V supply is required. Furthermore, matching of the positive and negative current sources and, more importantly, proper timing of the complementary pulses controlling the bridge can become challenging [51]. An additional drawback may be the droop rate due to leakage currents, if the output buffer is implemented with a bipolar device.

Another popular high-speed THA proposed by Vorenkamp [52] and makes use of a switched emitter follower. The schematic of this idea is provided in Fig. 2.23. During track mode, Q6 and Q12 are conducting turning on the emitter followers Q11 and Q5, respectively. The emitter followers charge the hold capacitor, $C_{hold}$, at a maximum slew rate of $I_2/C_{hold}$. Since the emitter follower presents a light load to the input buffer consisting of Q1 and Q2, the input bandwidth can be quite high. In hold mode, $I_2$ is directed away from the emitter followers and the extra voltage drop on the load resistors helps to turns them off, isolating $C_{hold}$. The use of the cascaded emitter follower output buffers (Q8-Q9 and Q14-Q15) improves the droop rate by $\beta^2$. Even though the emitter followers Q11 and Q5 are turned off, the input signal can pass through $C_{be}$ corrupting the sampled voltage. To mitigate this effect two

![Fig. 2.22: Open-loop THA with diode bridge switch](image)
feedforward capacitors ($C_{ff}$) are introduced. If $C_{ff} = C_{beQ5}$, signal feedthrough can be theoretically canceled \[52\] \[13\]. $C_{ff}$ are implemented as a series of diode connected bipolar devices to imitate the base-collector capacitance of the switched emitter follower, as illustrated in Fig. 2.24.

Even though this topology requires a relatively high supply voltage \[48\], it can achieve superior sampling rates and bandwidth, because the emitter follower can turn off faster than the diode bridge and has less capacitance. A slightly modified version of this topology achieved 40GHz bandwidth at 40 GS/s in a 0.18µm SiGe BiCMOS technology from a 3.6V supply \[54\].

Recently, a new idea based on the use of the base-collector (BC) diode as a switch has been proposed \[55\] \[51\]. In this case, a series sampling diode is introduced between the emitter follower and the hold capacitor as illustrated in Fig. 2.25. Unlike the switched emitter follower topology, the emitter follower is always on and biased by $I_0$ and the diode, $D_1$, switches on and off depending on the pulse applied (T for track mode or H for hold mode). When in track mode, $D_1$ is forward biased by $I_1$ charging the hold capacitor, $C_H$. When in hold mode, $D_1$ turns off and remains reversed biased by $R_L(I_1 + I_2)$.

The BC diode switch offers some advantages over the previously discussed topologies. First of all, it reduces the risk of ringing or oscillation, since the emitter follower does not need to drive a large capacitive load anymore \[51\] ($C_H$ is driven by the BC diode in Fig. 2.25). Secondly, the time constant of the BC diode, $R_{on}C_d$ (where $R_{on}$ is the on-resistance of the forward biased diode and $C_d$ the depletion region capacitance), is much smaller than the BE diode, making the former superior in high-speed applications \[55\]. Finally, in InP double heterojunction bipolar transistors (DHBt) the hole minority charge storage time is much smaller than in single SiGe HBTs, resulting in faster switching operation \[55\]. However, the use of InP DHBt entails a higher power consumption penalty, as high voltage supplies are required (in the order of 5V) as confirmed by recently published designs in such technologies \[51\] \[56\]. In addition,
designs in InP DHBT technology usually employ negative supply voltages \[51\], which increases the complexity and cost of the chip. Hence, the switched emitter follower remains the preferred method of realizing the switch for mass-production purposes. Due to the matching and timing challenges associated with diode bridge-based designs, as well as the reliability and power issues of InP DHBT solutions, the switched emitter follower topology has become the preferred choice for high-speed IC sampling circuits.

A list of recently published THAs above 30 GS/s is provided in Table 2.2.

Based on the data of the table some interesting observations can be made. First of all, all designs are based on the switched emitter follower topology with some minor variations (inclusion of HBT cascode in \[54\] or source followers in the CMOS design of \[57\]). Secondly, all high-bandwidth designs (above 20 GHz) make use of SiGe or InP HBT technology \[54\ 58\]. Among them, the InP implementations achieve the highest reported sampling rates (50 and 70 GS/s), but require rather high supply voltages for the InP devices (-5 V) resulting in large power dissipation (1200 mW). On the other hand, the CMOS design of reference \[57\] achieves the lowest power consumption, but also the lowest bandwidth. It is interesting to note that the CMOS solution uses two time-interleaved samplers, each running at 16.26 GS/s, revealing the limitations in terms of bandwidth and pure sampling rate involved in a CMOS-only approach.
<table>
<thead>
<tr>
<th>Ref.</th>
<th>fs [GS/s]</th>
<th>track BW [GHz]</th>
<th>IIP3@f_m [dBm@GHz]</th>
<th>THD@f_m [dB@GHz]</th>
<th>P_{DC} [mW]</th>
<th>Supply [V]</th>
<th>Area [mm^2]</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>50</td>
<td>27</td>
<td>20.7@6</td>
<td>-29.5@15</td>
<td>1200</td>
<td>-2.5, -5</td>
<td>0.73</td>
<td>250 nm InP DHBT $f_T/f_{MAX} = 370/650$ GHz</td>
</tr>
<tr>
<td>58</td>
<td>70</td>
<td>51</td>
<td>22@5</td>
<td>&lt;-52@2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>InP DHBT $f_T = 320$ GHz</td>
</tr>
<tr>
<td>57</td>
<td>32.5</td>
<td>19.3</td>
<td>N/A</td>
<td>-37@6</td>
<td>192</td>
<td>3</td>
<td>0.36</td>
<td>65 nm CMOS $f_T/f_{MAX} = 210/240$ GHz</td>
</tr>
<tr>
<td>54</td>
<td>40</td>
<td>43</td>
<td>0@19</td>
<td>-29@10</td>
<td>540</td>
<td>3.6</td>
<td>1.1</td>
<td>0.18 µm SiGe BiCMOS $f_T/f_{MAX} = 160$ GHz</td>
</tr>
<tr>
<td>56</td>
<td>30</td>
<td>N/A</td>
<td>19@1</td>
<td>&lt;-59@1</td>
<td>420†</td>
<td>5.5</td>
<td>0.77</td>
<td>250 nm InP DHBT $f_T/f_{MAX} = 300$ GHz</td>
</tr>
</tbody>
</table>

†: excludes output buffer

Table 2.2: State-Of-The-Art Track & Hold Amplifiers.
Chapter 3

Low-Power, Low-Noise Broadband Amplifier Design

This chapter presents the design of two broadband low-power transimpedance amplifiers (TIAs). Both designs employ a single-ended topology, but can be easily converted to differential, if differential operation is necessary. The first one is a 50Ω input-matched, low-power solution for applications with simple NRZ modulation such as silicon photonics. 50Ω matching is necessary for being able to properly test the circuit electrically in the lab. The second one is designed for high performance fiber-optic systems with complex amplitude modulation such as PAM-4, which calls for high linearity and low noise, and therefore employs emitter degeneration. At first, a short comparison study on the low-noise broadband topologies analyzed in Chapter 2 is conducted based on simulation results. Next, a step-by-step methodology for optimal noise figure (NF) is presented and applied to the design of a low-power, 80GHz TIA and a 92GHz, linear TIA. Finally, the chapter concludes with the simulation results of the individual amplifiers.

3.1 TIA topology investigation

The low-noise, broadband topologies of Chapter 2 are illustrated in Fig. 3.1 and were considered for implementation in a 55nm SiGe BiCMOS technology with 9metal BEOL, MiM and MoM capacitors, AMOS varactors, thin and thick oxide MOSFETs with several Vt flavours, and three types of SiGe HBT [42]. The performance of the high-speed HBT and 55nm MOSFET are plotted in Fig. 3.2. The high-speed HBT has a measured fT/fMAX of 335/310 GHz. The peak-fT current density is approximately 1.5 mA/µm of emitter length while the optimal noise figure current density above 10 GHz is 1 mA/µm. The 55nm MOSFET has a measured fT/fMAX of 230/300 GHz when biased at a JpfT of 0.3 mA/µm and its optimal noise current density, Jopt is 0.15 mA/µm.

HBT implementations were preferred over their MOS counterparts because of the following reasons:

- The transconductance of the HBT device is higher for the same bias current: \( g_{mHBT} > g_{mMOS} \).
- The output resistance of the HBT is higher: \( r_{oHBT} > r_{oMOS} \).
- The collector-substrate capacitance of the HBT is lower than the drain-bulk capacitance of the MOS device: \( C_{cs} < C_{db} \) [60].
Chapter 3. Low-Power, Low-Noise Broadband Amplifier Design

Fig. 3.1: Broadband low-noise topologies (a): common-base, (b): regulated cascode, (c): HBT-source follower, (d): Darlington cascode and (e): HBT-TIA

Fig. 3.2: Measured $f_T$ and $f_{MAX}$ of a high-speed HBT (100 nm×4.5 µm) and a 55nm n-MOSFET (40×55 nm×770 nm) [59].

- The effect of flicker noise is less prominent in HBT than MOS devices [13].

In the following sections the above topologies will be examined in terms of bandwidth and NF.

Fig. 3.3 shows the simulation results for these amplifiers, when designed for 14 dB gain and biased at the optimal noise current density, while driving an ideal 50 Ω output buffer that exhibits infinite input impedance and zero capacitive loading. Inductive peaking [33] was employed in all designs to improve the bandwidth and reduce the noise contribution of the feedback resistor at high frequencies. Apart from the common-base and regulated cascode, all other amplifiers were designed with 150mV DC drop on the degeneration resistor, to accommodate input swings up to 300 mVpp.

Although they exhibit the largest bandwidth, the common-base and regulated cascode topologies suffer from poor linearity and relatively high noise. This is caused by the small bias current of only 0.5 mA needed to set the input resistance to 50 Ω. In contrast, the HBT-source follower demonstrates the lowest noise below 75 GHz but also the lowest bandwidth. The use of the source follower, which exhibits a typical gain of 0.7-0.8, forces the designer to increase the gain of the main amplifier to compensate for this loss in gain, resulting in a higher time constant at the gate node that limits the
As expected, the Darlington cascode topology exhibits excellent bandwidth and moderate noise figure, but requires a rather high supply voltage. This is due to the voltage drop on the collector resistor, which must be adjusted to achieve the desired gain. Further, the collector resistor must be able to sustain the sum of the currents of both the emitter follower and the cascode stage. Given the fact that in a modern technology node the maximum allowed current density for a polysilicon resistor is in the order of 0.5 mA/μm of width, wide resistors would have to be employed in the layout, which may result in increased capacitance at the output node. Another issue is that the Darlington introduces more phase shift than a single transistor which may lead to poor phase margin and possibly to oscillations when used in negative feedback configurations. Because of this it proves hard to match this amplifier at the input to 50 Ω, as confirmed by $S_{11}$ simulations in Fig. 3.4. Note that the curve corresponding to the Darlington topology exceeds 0 dB at high frequencies. Finally, the HBT cascode device must be properly decoupled at its base, because it is very prone to oscillations at high frequencies.

The HBT-TIA, on the other hand, achieves the lowest overall noise figure up to 100 GHz due to its decreased number of components, and with bandwidth comparable to that of the Darlington cascode. Among its advantages is the fact that it can operate from a lower supply voltage. For example, in the case provided in Fig. 3.1 the Darlington cascode operates from a 5.2V supply and draws 12 mA, while
Table 3.1: Power summary of broadband topologies in Fig. 3.1.

<table>
<thead>
<tr>
<th>Topology</th>
<th>Supply [V]</th>
<th>Current [mA]</th>
<th>Power [mW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>common-base</td>
<td>2.2</td>
<td>1.1</td>
<td>2.42</td>
</tr>
<tr>
<td>regulated cascode</td>
<td>3.3</td>
<td>2</td>
<td>6.6</td>
</tr>
<tr>
<td>HBT-SF</td>
<td>3</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>darlington cascode</td>
<td>5.25</td>
<td>12</td>
<td>63</td>
</tr>
<tr>
<td>HBT-TIA</td>
<td>3.2</td>
<td>8</td>
<td>25.6</td>
</tr>
</tbody>
</table>

The HBT-TIA consumes only 8 mA from a 3.2V supply.

Table 3.1 summarizes the power consumption of the amplifiers in the comparison. The common-base and regulated cascode consume very little current because of the input matching requirement. The current value deviates from 0.5 mA because it accounts for the parasitic emitter resistance of the HBT, $r_E$. To maintain speed and low-noise they are biased at the optimal noise current density of 1 mA/µm which results in small device sizes and, thus, poor fanout, which may prove inadequate to drive the subsequent stages in a fiberoptic receiver. From the feedback topologies it is clear that the proposed HBT-TIA has the lowest power consumption, while simultaneously meeting the input linearity constraint, providing matched input to 50 Ω and not suffering from poor fanout.

### 3.2 80 GHz Low-Power TIA

In this section a step-by-step methodology for the design of an 80GHz SiGe HBT TIA with over 13dB gain is presented that results in optimal noise figure across the entire bandwidth with the least possible power consumption.

#### 3.2.1 Design

Fig. 3.5 shows the TIA schematic with component values. It consists of the shunt-shunt feedback amplifier with inductive peaking both at the collector node and the feedback path, and a 50Ω output driver. The feedback inductor, $L_F$, also helps reduce the thermal noise of the feedback resistor, $R_F$, at high frequencies [33]. Biasing is accomplished via $R_F$, which connects the collector with the base of the HBT, forcing the device to draw a current equal to:

$$I_E = \frac{V_{CC} - V_{BE}}{R_C + \frac{R_F}{\beta+1}}$$

(3.1)

As a result of this biasing scheme, the amplifier can operate from just 1.15V supply. However, its current value depends on $\beta$ of the transistor and is very sensitive to PVT variation as we will see later in Fig. 3.10. In a differential implementation, the current can be accurately set by a MOS current source, which would not require more than 0.2V voltage headroom. This single-ended version was just a test vehicle to explore the lowest noise and broadest bandwidth performance.

The algorithmic design of this TIA is presented below:

i) Both HBTs were biased at the optimal noise current density for the desired 3dB bandwidth of 80 GHz, which was found through simulation to be 1 mA/µm of emitter length. This value also matches with previous measured data of this technology.
ii) The feedback resistor, $R_F$, was set to 202 $\Omega$ to minimize noise at low frequencies without sacrificing the bandwidth. As explained in Chapter 2, a large value of $R_F$ ensures a low $F_{MIN}$ for the shunt-shunt feedback amplifier, but at the same time increases its input time constant ($R_FC_{in}$) reducing its bandwidth. The resistor value is best found through simulation, as it depends both on the bandwidth and NF requirements. Usually, a value less than 250 $\Omega$ provides low noise and adequate bandwidth for most applications. The choice of $R_F$ also fixes the required loop gain for 50 $\Omega$ input matching [33]:

$$R_{in} = 50 \Omega = \frac{R_F}{A+1}$$

where $A$ is the gain of the amplifier.

iii) The HBT size and bias current were swept in tandem in simulation, while keeping the current density constant, until the real part of the optimum noise impedance, $R_{SOPT}$, became 50 $\Omega$ at the 3dB frequency. Since the optimal noise current density is 1 mA/$\mu m$ and two emitter stripes of 3 $\mu m$ were found to provide the desired $R_{SOPT}$, the current was set to 6 mA. This transistor size and bias current minimize the noise contribution of the HBT device at high frequencies.

iv) With the device size and current fixed, $R_C$ was selected to provide the required gain, which according to equation (3.2) should be close to 3. The gain of the shunt feedback topology is given by [33]:

$$A = \frac{R_C}{R_F} - g_{meff} R_C \approx - \frac{g_{meff} R_C}{1 + \frac{R_C}{R_F}}$$

where $g_{meff} = \frac{g_m}{1 + g_m r_E}$ is the effective transconductance. Because of the absence of any emitter degeneration the accuracy of the gain depends heavily on the transistor and its parasitic emitter resistance and so $R_C$ was determined in simulation after parasitic extraction of the HBT devices.

v) In the absence of a current source, the bias current of the first stage is determined from the feedback loop. Since all resistor values have been fixed the choice of voltage supply gives the desired 6 mA of bias current:
\[ V_{CC} = I_E \left( R_C + \frac{R_F}{\beta + 1} \right) + V_{BE} \approx 1.15 \text{ V} \] (3.4)

Because the base draws negligible current, there is almost no voltage drop on the feedback resistor and so the first stage acts as a current mirror for the output buffer in a ratio of 2:1.

vi) Inductive peaking was applied at the base and collector nodes according to (3.5) in order to maximize the bandwidth. In order to ensure minimum group delay variation the inductor values were initially chosen according to [33]:

\[ L = \frac{R^2 C}{3.1} \] (3.5)

where R and C represent the total resistance and capacitance, respectively, at the offending node. Using equation (3.5) as a good starting point, the final values were optimized after parasitic extraction using computer simulation.

Minimizing the layout footprint of the entire circuit is critical in maximizing the bandwidth at mm-wave frequencies. Special attention was given to the inductor layout to ensure that the self resonant frequency is at least 200 GHz, suitable for 100 GHz operation. The 237pH stacked inductor was realized in the top 3 metal layers using 2 turns with minimum spacing and minimum metal width to increase its inductance. The footprint was kept small (11 µm×11 µm) to maintain the SRF above 200 GHz.

The 150pH and 120pH inductors were designed in a similar manner in the top 3 metal layers and occupy 10 µm×9 µm and 9 µm×9 µm, respectively. Grounded substrate PTAPs were introduced between inductors to improve isolation. Excepting the inductor regions, an uninterrupted multi-metal ground and supply mesh covers the entire chip to minimize ground resistance and inductance, supply plane inductance, and to provide adequate supply decoupling from DC to 200 GHz.

Post layout simulations show a 3dB bandwidth of 80 GHz and a noise figure of less than 6dB all the way up to 100 GHz, resulting in an equivalent input noise of 0.28 mVrms, as demonstrated in Fig. 3.6. Since there is no emitter degeneration in this design, the maximum linear input swing at the base of the bipolar is of the order of 10 mVpp, which severely limits the input linearity of this TIA in particular. The main source of nonlinearity comes from the 50Ω output driver which sees the amplified signal at the output of the TIA stage and does not have any sort of linearization scheme such as negative feedback or emitter degeneration. The linearity of the TIA was characterized in simulation through the 1dB compression point, \( P_{1dB} \), and the total harmonic distortion, THD, which is a better suited metric for broadband circuits. Both simulations are provided below in Fig. 3.7.

Group delay simulation results are provided in Fig. 3.8 which is an important parameter of broadband amplifiers. A group delay variation of less than ±1 ps can be observed, which is better than the best reported value of Table 2.1 with state-of-the-art broadband amplifiers. Finally, the transient performance was characterized through eye diagram simulations at small input amplitudes (as low as 5 mVpp) up to 80 Gb/s as illustrated in Fig. 3.9.

Due to the absence of emitter degeneration, the gain of the amplifier is very susceptible to PVT variations and has been verified in simulation in Fig. 3.10 for the conditions included in Table 3.2.
Fig. 3.6: Simulated S-parameters and noise figure. Vertical dashed line indicates 3dB bandwidth.

Fig. 3.7: Simulated 1dB compression point (left) and THD (right) at different frequencies and input power levels.

Fig. 3.8: Simulated group delay.
Fig. 3.9: 80 Gbps simulated eyes with 5 mV_{pp} input.

Fig. 3.10: Simulated $|S_{21}|$ over process corners.

<table>
<thead>
<tr>
<th></th>
<th>typical</th>
<th>slow</th>
<th>fast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transistor</td>
<td>typical</td>
<td>slow</td>
<td>fast</td>
</tr>
<tr>
<td>Resistance</td>
<td>typical</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Capacitance</td>
<td>typical</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Temperature</td>
<td>65°C</td>
<td>125°C</td>
<td>25°C</td>
</tr>
<tr>
<td>Supply</td>
<td>typical</td>
<td>typical-10%</td>
<td>typical+10%</td>
</tr>
</tbody>
</table>

Table 3.2: Process Corners
3.3 92 GHz Linear TIA

In this section the design of a 92GHz SiGe HBT TIA with 13dB gain and 200 mVpp input linearity is presented. This amplifier is well suited for PAM-4 or discrete multitone (DMT) 100Gb/s high-performance fiberoptic receivers.

3.3.1 Design

The schematic of this amplifier is shown in [3.11] and was designed in the same 55nm SiGe BiCMOS technology presented earlier in this chapter. It consists of a common-emitter HBT amplifier with resistive feedback and resistive emitter degeneration for improved linearity and a 50 Ω matched output driver. The bandwidth is improved by inductive shunt peaking at the collector and series peaking at the base node. Again the feedback resistor alleviates the need for any current source, thus allowing operation for a lower supply voltage. The current now is set by (3.7). Note that in this case the supply is higher to accommodate enough voltage drop on the degeneration resistor, in order to satisfy the input linearity requirements.

![Fig. 3.11: Linear TIA schematic with component values](image)

The circuit operates from a 2.3V supply and can be easily converted to a differential equivalent by introducing a low V_{DS} MOS current source, as explained in section 3.2. The algorithmic design of this TIA is presented below:

i) Both HBTs were biased at the optimal noise current density, ensuring the minimum possible $F_{MIN}$ for the desired 3dB bandwidth of 92 GHz, which was found through simulation to be 1 mA/μm of emitter length.

ii) The feedback resistor, $R_F$, was set to 220 Ω to minimize noise at low frequencies without compromising the bandwidth. This choice sets the required loop gain for 50 Ω input matching according to equation (3.2).

iii) The DC voltage drop on the degeneration resistor, $R_E$, was set to 200 mV for an input linear swing of 400 mVpp.

iv) Keeping the voltage drop on $R_E$ and the current density constant, the HBT size and bias current were swept in tandem, until the real part of the optimum noise impedance became 50 Ω at 92 GHz.
An optimal emitter length of $12 \mu m$ provided noise matching at 92 GHz, which was realized from two stripes of $6 \mu m$ each. Since the devices must be biased at $1 mA/\mu m$ for low noise, the DC bias current of the TIA stage was fixed at $12 mA$.

$v$) With the device size, current density and value of $R_E$ fixed, $R_C$ was selected to provide the required gain, which in this case can be expressed as [33]:

$$A = \frac{-R_C}{R_E + r_E + \frac{1}{g_m}} + \frac{R_F - R_E - r_E}{R_F + R_C}$$  \hspace{1cm} (3.6)

Starting from (3.6) the load resistance that resulted in good input matching without increasing the time constant at the intermediate node, was found to be $100 \Omega$.

$vi$) Now that all component values have been fixed, the bias current is set ultimately by the voltage supply:

$$I_E = \frac{V_{CC} - V_{BE}}{R_C + R_E + \frac{R_F}{R_F+1}}$$  \hspace{1cm} (3.7)

For $12 mA$ the above equation yields $V_{CC} \approx 2.3 V$. Note that the output driver bias current is in a ratio 4:3 with respect to the TIA.

$vii$) Inductive peaking was applied at the base and collector nodes according to [3.3] in order to maximize the bandwidth. The feedback inductor is now connected in series with the base to resonate with the input capacitance. This has been proven to result in 25% bandwidth boosting [14] and results in lower inductance values compared with placing the inductor in the feedback path, as first shown in Fig. 3.5. On the other hand, since the inductor is no longer in series with the feedback resistor, $R_F$, it does not suppress its noise current contribution. In the layout the feedback resistor was placed at the center of the inductor to reduce interconnect inductance, as shown in Fig. 3.12. Since the inductor is connected in series with the base of the transistor, its quality factor ($Q$) had to be improved to reduce its noise contribution. A $20 \mu m \times 20 \mu m$ spiral inductor realized in the top copper layer (metal 8) with 2 turns and $2 \mu m$ wide spirals with minimum spacing resulted in a $90 \mu H$ design with $Q=12$ at 92 GHz and an SRF over 200 GHz. A degeneration capacitor of $130 fF$ was placed across $R_E$ to further boost the high frequency response and suppress the noise of $R_E$ by introducing a zero at 81 GHz.

Post layout simulations in Fig. 3.13 show a 3dB bandwidth of 92 GHz and a noise figure of less than 7dB all the way up to 100 GHz, resulting in an equivalent input noise of $0.3 mV_{rms}$. The input referred noise current integrated up to 100 GHz is $5.6 \mu A_{rms}$. Thanks to the emitter degeneration resistor, the linearity of this design has improved significantly (about 15 dB) from the previous design (see Fig. 3.7) judging from the IP$_{1dB}$ and THD simulations included in Fig. 3.14. The simulated input referred 1 dB compression point at 1 GHz is $-10.67$ dBm and the total harmonic distortion $-19.64$ dB at the 1 dB compression point. The linearity is mainly limited by the output buffer, which has the same degeneration of 200 mV DC as the TIA stage.

Simulations of group delay, transimpedance gain, and eye diagrams are illustrated in Fig. 3.15 and 3.17 respectively. Group delay variation remains close to $\pm 1$ ps. Eye diagrams were simulated at different input amplitudes and data rates. Fig. 3.16 and 3.17 plot the simulated eyes for 5 and $50 mV_{pp}$ input amplitudes, respectively, confirming the high sensitivity and improved linearity of this topology.
Fig. 3.12: Layout detail of the feedback network.

Fig. 3.13: Simulated S-parameters and noise figure of linear TIA. Vertical dashed line indicates 3dB bandwidth.
Fig. 3.14: Simulated 1dB compression point (left) and THD (right) at different frequencies and input power levels for the linear TIA.

Fig. 3.15: Simulated group delay of linear TIA.

Fig. 3.16: 120 Gbps $2^{31} - 1$ simulated eyes with 5 mVpp input.
The presence of emitter degeneration makes the amplifier more robust in terms of PVT variations as confirmed by simulations in Fig. 3.18 for the conditions included in Table 3.2. It is clear that the gain variation is less pronounced compared to the design without degeneration (see Fig. 3.10). The robustness of this design against PVT variations is also confirmed by the small spread of the standard deviation of the gain, as shown in Table 3.3 for 200 Monte Carlo simulation runs.

<table>
<thead>
<tr>
<th>Design</th>
<th>Typical</th>
<th>Slow</th>
<th>Fast</th>
</tr>
</thead>
<tbody>
<tr>
<td>80GHz Low-Power TIA</td>
<td>$\sigma = 6.1$ dB</td>
<td>$\sigma = 3.5$ dB</td>
<td>$\sigma = 9.94$ dB</td>
</tr>
<tr>
<td>92GHz Linear TIA</td>
<td>$\sigma = 0.137$ dB</td>
<td>$\sigma = 0.15$ dB</td>
<td>$\sigma = 0.058$ dB</td>
</tr>
</tbody>
</table>

Table 3.3: $|S_{21}|$ standard deviation ($\sigma$) after 200 Monte Carlo simulations.
Fig. 3.18: Simulated $|S_{21}|$ of linear TIA over process corners.
Chapter 4

Track & Hold Amplifier with New Quasi-CML MOS-HBT Switch

In this chapter the design of a wideband, high-sampling rate track-and-hold amplifier is presented that satisfies the requirements of future fiber-optic receivers listed in Table 4.1. The circuit builds on the concept of the switched emitter follower topology, which was briefly discussed in Chapter 2 and combines the benefits of the MOS and HBT devices available in a BiCMOS technology to achieve higher sampling rates than reported in the published data of Table 2.2. The introduction of the new quasi-CML MOS-HBT sampling switch, which will be discussed in detail later, also alleviates the need for a current source leading to low-supply and lower power compared to previous designs. As proof of concept, the design of a low-power, 90GS/s, 40GHz bandwidth THA will be described based on the new sampling switch.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>7 bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample rate</td>
<td>75 GS/s</td>
</tr>
<tr>
<td>Input bandwidth</td>
<td>&gt;30 GHz</td>
</tr>
<tr>
<td>Input amplitude</td>
<td>300 mVpp per side</td>
</tr>
<tr>
<td>Power</td>
<td>&lt;300 mW</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of THA specifications

4.1 Design Considerations

Among the sampling topologies discussed in Chapter 2 the switched emitter follower is the best candidate for highest speed applications because of its high switching speed and high linearity performance. The schematic of the original topology as first proposed in [52] is repeated in Fig. 4.1 for convenience. This topology serves as the base for most published THAs with sampling rates above 30 GS/s [53, 54, 63, 53] proof of its suitability for applications that require fast switching speed, such as the front-end of ADCs for future 100Gb/s fiber-optic communication systems.

Even though the differential bipolar pair of Q7-Q6 and Q12-Q13 in Fig. 4.1 requires theoretically just about $4 \times V_T$ to fully switch [62] (in practice to account for temperature and process variations as well as for the drop on the parasitic emitter resistance, a swing of 200-300 mV is typically chosen [64]),
its switching frequency can be vastly improved by employing a more sophisticated architecture, as shown in the 40GS/s THA in Fig. 4.2.

In this design a bipolar cascode is used to reduce the effect of the Miller capacitance and boost the sampling rate. However, the addition of an extra HBT device for the cascode raises the minimum supply voltage this topology can operate from, and also requires special attention to the layout to guarantee stability of the bipolar cascode at high frequencies by providing plenty of decoupling at the base of the cascode device. Therefore, it becomes evident that the design of the switch is crucial both for fast as well as low-power operation.

4.1.1 MOS-HBT cascode

We saw that the use of the cascode in the switch controlling the emitter follower overcomes the issue related to the Miller capacitance. However, not all cascode topologies are made equal. In a BiCMOS process the designer has the flexibility to choose among a wide variety of such topologies as illustrated...
Fig. 4.3: BiCMOS cascode topologies: (a) MOS-MOS, (b) HBT-HBT, (c) HBT-MOS, and (d) MOS-HBT

in Fig. 4.3, namely [60]: MOS-MOS, HBT-HBT, HBT-MOS, and MOS-HBT (or BiCMOS).

The delay of the above cascode topologies was first studied in [60]. The results of the study based on the open-circuit time constant (OCTC) technique [13] are summarized below, where $k$ denotes the fanout and $C_{\text{int}}$ the interconnect capacitance associated with the output node.

\begin{align}
\tau_{\text{MOS-MOS}} &\approx \Delta V \left( C_{gd} + C_{db} + C_{\text{int}} \right) \frac{1}{I_{\text{tail}}} + \left( k + \frac{R_G}{R_L} \right) \Delta V \left( C_{gs} + 2C_{gd} \frac{1}{I_{\text{tail}}} \right) + \frac{C_{gs} + C_{db} + C_{gd} + C_{sb}}{g_{m,MOS}} \quad (4.1) \\
\tau_{\text{HBT-HBT}} &\approx \Delta V \left( C_{bc} + C_{cs} + C_{\text{int}} \right) \frac{1}{I_{\text{tail}}} + \left( k + \frac{R_B}{R_L} \right) \Delta V \left( C_{z} + 2C_{bc} \frac{1}{I_{\text{tail}}} \right) + \frac{C_{z} + C_{cs} + C_{bc}}{g_{m,HBT}} \quad (4.2) \\
\tau_{\text{HBT-MOS}} &\approx \Delta V \left( C_{gd} + C_{db} + C_{\text{int}} \right) \frac{1}{I_{\text{tail}}} + \left( k + \frac{R_B}{R_L} \right) \Delta V \left( C_{gs} + \left( 1 + \frac{g_{m,HBT}}{g_{m,MOS}} \right) C_{bc} \right) + \frac{C_{gs} + C_{cs} + C_{bc} + C_{sb}}{g_{m,MOS}} \quad (4.3) \\
\tau_{\text{BiCMOS}} &\approx \Delta V \left( C_{bc} + C_{cs} + C_{\text{int}} \right) \frac{1}{I_{\text{tail}}} + \left( k + \frac{R_G}{R_L} \right) \Delta V \left( C_{gs} + \left( 1 + \frac{g_{m,MOS}}{g_{m,HBT}} \right) C_{gd} \right) + \frac{C_{gs} + C_{db} + C_{gd}}{g_{m,HBT}} \quad (4.4)
\end{align}

In equations (4.1) - (4.4), the first two terms account for the output and input time constant, respectively. The former improves when a SiGe HBT is used as the cascode device, since the collector-to-substrate capacitance, $C_{cs}$, tends to be much smaller than the drain-to-bulk capacitance of the n-MOS, $C_{db}$. This is more pronounced in nanoscale CMOS nodes, where $C_{db}$ is comparable with the gate-to-source capacitance, $C_{gs}$. On the other hand, the latter term related to the input time constant can be minimized by appropriately sizing the finger width of the MOSFET and by using contacts on both sides of the gate, in order to reduce the gate resistance of the MOS transconductor device. The gate
As was the case for the HBT and MOS INVs analyzed earlier, the first two terms in equations (3.13)-(3.16) with a loading factor of $\frac{R_{\text{CON}}}{N_{\text{CON}}N_F}$

\[
R_G = \frac{R_{\text{shg}} W_F}{3N_F L} + \frac{R_{\text{CON}}}{N_{\text{CON}}N_F}
\]  

where $R_{\text{shg}}$ represents the sheet resistance per square of the polysilicon gate, $R_{\text{CON}}$ the metal-to-poly via resistance, $N_F$ the number of fingers, $W_F$ the finger width, $L$ the physical gate length of the device, and $N_{\text{CON}}$ the number of metal-to-poly vias per gate contact. In a differential implementation there is a minimum voltage swing required to fully switch the current to one side or the other. This is critical in high-speed logic blocks, such as CML inverter stages, or high-speed analog blocks such as track-and-hold amplifiers. When biased for maximum switching speed, the minimum voltage swing required to fully switch a bipolar differential pair is less than that of a MOS pair. However, to ensure that the switching operation is robust across PVT at least 200 mV are necessary in the bipolar case. As the gate length shrinks in new CMOS nodes, so does the minimum overdrive voltage for switching [65] and as a result, the minimum required voltage swing for the bipolar and MOS pair become almost comparable in the most advanced 55nm SiGe BiCMOS nodes.

Finally, the third term in the above equations refers to the intermediate node between the two devices in the cascode and is smaller for topologies with SiGe HBT devices at the output, because their transconductance, $g_{m,HBT}$, is always higher than $g_{m,MOS}$ for the same bias current. Furthermore, because the low-$V_T$ GP 55nm MOSFETS have smaller $V_{GS}$ at the peak $f_T$ bias than the corresponding $V_{BE}$ of the SiGe HBT, the use of a MOS device allows for a lower supply voltage as opposed to an HBT-only implementation, without compromising the bandwidth. Fig. 4.4 plots the time constants of the different cascode topologies for a 130nm and a 55nm SiGe BiCMOS technology based on equations (4.1) - (4.4). A tail current of 6 mA was assumed with a voltage swing, $\Delta V$, of 300 mV for an HBT input and 400 mV for the 55nm n-MOS input, which corresponds to the minimum voltage swing that ensures full switching across PVT (the 130nm n-MOS requires 500 mV). The gate resistance, $R_G$, is given by (4.5) and the base-emitter capacitance is calculated from:

\[
C_{be} = g_{m,eff} r_T + C'_{je} w_E l_E
\]

where $g_{m,eff}$ is the effective transconductance accounting for the parasitic emitter resistance: $\frac{g_m}{1 + g_m r_T / (w_E l_E)}$. 

Fig. 4.4: Calculated time constants: (a) 130nm SiGe BiCMOS [60] and (b) 55nm SiGe BiCMOS nodes.
\( \tau_F \) the total transit time, \( C_{je}' \) the junction depletion capacitance per emitter area, \( w_E \) the emitter width, and \( l_E \) the emitter length. It is clear that the MOS-HBT combination minimizes all time constants leading to the highest bandwidth topology irrespective of the technology node. Interestingly, the bandwidth of the MOS-HBT cascode has almost doubled from the 130nm node in the new 55nm SiGe BiCMOS node. In order to confirm that the bandwidth advantage explained above is valid irrespective of technology nodes, the results of the simulation of the different cascode topologies are displayed in Fig. 4.5.

Because of their higher transconductance the HBT-HBT and HBT-MOS cascode topologies have higher gain, as shown in Fig. 4.5. To illustrate that the advantage of the MOS-HBT cascode is not associated with just a gain-bandwidth trade-off, the measured maximum available gain, MAG, of the different combinations when biased at the same current is depicted in Fig. 4.6. Little to no difference between the MOS-HBT and the HBT-HBT cascode can be observed at high frequencies both in the 130nm and the 55nm SiGe BiCMOS node, confirming that the MOS-HBT combination maximizes the gain-bandwidth product.
Fig. 4.6: Measured MAG of cascode topologies in: (a) a 130nm [60] and (b) a 55nmSiGe BiCMOS technology [59].

### 4.1.2 Quasi-CML Topology

Apart from the MOS-HBT cascode, another modification to the sampling switch is the elimination of the tail current switch entirely. This idea has been first proposed for CML logic latches [67, 68]. This can save about 0.2-0.3 V, which allows operation from a lower supply voltage. A quasi-CML BiCMOS switch is illustrated in Fig. 4.7.

Because of the absence of a tail source, care must be taken to ensure the devices are optimally biased for high-speed switching [33]. The differential switch is biased in a similar fashion as a class-AB stage. In particular, when the differential pair is balanced, both MOSFETs are biased at 0.15 mA/µm and when fully switched, the drain current swings up to 0.3 mA/µm, thus maintaining operation close to the peak of the $f_T$ curve, which maximizes the switching speed. Similarly, the HBT is biased at $0.75I_{fTHBT}$ and switches to $1.5I_{fTHBT}$. A high $V_T$ MOS device ensures that one side of the differential pair is turned off even when the clock signal applied is as low as 0.4 V for the 55nm BiCMOS technology node. On the other hand, this stage provides no common-mode rejection due to the absence of the tail current source.
4.2 Proposed Design

The block diagram of the proposed THA is shown in Fig. 4.8. It consists of a linear input buffer, a switched emitter follower (SEF) based on the MOS-HBT sampling switch, a 50Ω driver, and a clock amplifier. The combination of the input buffer and the SEF basically make up the track-and-hold amplifier. Each of these blocks is discussed separately in the following subsections.

![Block diagram of the proposed track-and-hold amplifier.](image)

**Fig. 4.8: Block diagram of the proposed track-and-hold amplifier.**

### 4.2.1 Linear Input Buffer

The linear input buffer plays a very important role in the operation of the THA, as it determines the linearity of the overall system. This buffer should closely follow the input signal while providing a linear gain, as any distortion due to the non-linearity of the input buffer will directly affect the analog value stored on the hold capacitor during hold-mode. As we will see later, the design of this buffer and especially the choice of the voltage drop on its load resistor, also impacts the performance of the switch in the hold-phase.

Different possible topologies have been proposed in the literature, which could serve as eligible candidates for a linear input buffer, such as differential pair with emitter degeneration, diode loaded differential pair [69], Quinn’s cascomp [70], Caprio’s quad [71], Karanicolas’ folded diode loaded differential pair [72], and Miki differential pair [73]. These topologies are illustrated in Fig. 4.9. A study of the suitability of these amplifiers for the input stage of track-and-hold amplifier circuits has been conducted in [69]. Except for the classic differential pair with emitter degeneration and the diode loaded differential pair, the study concluded that all other topologies were inappropriate either due to circuit complexity, poor linearity or high power requirements. Besides the topologies illustrated in Fig. 4.9 a MOS-only and a MOS-HBT differential pair biased for maximum linearity were tested. Both topologies were discarded, as the former lacked bandwidth, while the latter required a rather high supply voltage (exceeding 2.5 V).

Since no diodes were available in the design kit used in this thesis, and a diode connected HBT would introduce more parasitic capacitance at the output than a simple polysilicon resistor, a differential pair with emitter degeneration was selected as the best option for the linear input buffer of the THA.

The schematic of the linear input buffer is shown in Fig. 4.10. Because the input buffer must interface to 50Ω for testing purposes, 50Ω resistors were connected from the 1.8V supply to the base of Q2 and
Q3. 1.8 V was the lowest supply voltage that the CML clock amplifier network could operate from. To account for the voltage drop on the load resistors, the next higher standard supply of 2.5 V was chosen for the MOS-HBT switch. This choice of power supplies also allows for sufficient drop on the input transistors, in order to avoid the transistor entering the saturation region and degrading the bandwidth. In contrast to [54] a bipolar current source was chosen for the buffer, in order to increase the common mode rejection ratio (CMRR) of this stage, as the THA would be tested single-endedly in the lab.

The input buffer must satisfy two main requirements: adequate bandwidth in track-mode and high linearity. Unity gain was chosen for this amplifier, in order to satisfy the linearity requirement of the complete THA from a low supply voltage and maintain high bandwidth operation. This choice, however, comes at the expense of noise. Since the buffer provides no gain, the thermal noise of subsequent stages is not attenuated raising the noise referred at the input of the circuit. It has been demonstrated that thermal noise \(4kTR\) is the main noise contributor in a sampled circuit [74]. Due to this noise getting folded in the band of interest, the aliased power assuming a single-side-band spectrum is given by [75, 76]:

\[
P_{als|dB} = 10 \log \left( \frac{\pi}{2} \frac{EBW}{0.5 f_s} \right)
\]  

(4.7)

where \(EBW\) is the effective noise bandwidth and \(f_s\) the sampling frequency. \(EBW\) can be easily determined by the noise transfer function in simulations, when the THA is configured in track-mode. The use of a low-noise amplifier before the input buffer can effectively reduce the noise contribution of all
The introduction of an LNA would call for higher linearity by the input buffer, which would lead to higher power consumption. In addition, the bandwidth of the system would be compromised. Since the main goals of this THA design were linearity, bandwidth, and low-power operation instead of low-noise, the design of such an LNA was not pursued.

In the case of a bipolar differential pair with degeneration the small-signal gain is given by:

$$A_v = \frac{g_m R_L}{1 + g_m (R_E + r'_E/(w_E l_E))} \approx 1 \quad (4.8)$$

where $g_m = \frac{I_T}{2 V_T}$ is the small signal transconductance of the HBT, $R_L$ is the load resistor, $R_E$ is the degeneration resistor, $r'_E$ the parasitic emitter resistance in $\Omega \mu m^2$, $l_E$ the emitter length, and $w_E$ the emitter width. For the values depicted in Fig. 4.10 the simulator gave a small-signal gain value very close to 1.

Bandwidth is important, so that the input buffer can quickly follow the input signal after the hold-phase is over. To process signals with $> 32$GHz bandwidth, the minimum requirement for 64GBaud fiber-optic systems, a sampling frequency of over $80$ GHz is needed. In practical systems larger bandwidth is budgeted for. Using the OCTC technique and ignoring the feedforward capacitors, $C_{ff}$, shown in Fig. 4.14, the bandwidth of the input buffer can be approximated by:

$$f_{3dB} = \frac{1}{2\pi (\tau_{in} + \tau_{out})} \quad (4.9)$$

$$\tau_{in} = \frac{R_s}{2} (C_{bc2} + 2C_{bc1}) \quad (4.10)$$

$$\tau_{out} = R_L (C_{bc2} + C_{cs2} + C_{bc1} + C_{bc1} + C_{bc4} + C_{cs4}) \quad (4.11)$$

where in (4.10) the 50Ω source resistance, $R_s$, appears in parallel with the 50Ω resistance used for input matching in Fig. 4.10. Due to the unity gain of the buffer and the Miller effect $C_{bc2}$ is multiplied by 2. Using measured data of the 55nm SiGe BiCMOS technology, (4.9) yields 51.38 GHz. To compensate
for device and interconnect capacitance after parasitic extraction as well as for the existence of the feedforward capacitors, inductive peaking was employed in this design to Fig. 4.2. Since the simulated bandwidth of the whole THA was found to be 40 GHz in the worst corner after parasitic extraction, extra emitter followers were not required at the input, as they would only increase the total power of the circuit unnecessarily (corners are defined in Table 3.2). As alluded to before, no cascode was used in the buffer as that would result in higher supply and ultimately higher power.

When the output swing quantities are referred to the input through $A_v$, the maximum peak-to-peak input linear swing is given by the minimum of the above requirements [54]:

$$
Swing_{max} = \min \left\{ \frac{I_T R_L}{A_v}, \frac{2(V_{CEQ2,3} - V_{CEsatQ2,3})}{A_v}, 2V_T(1 + g_m R_E) \right\}
$$

(4.12)

The linearity target for the THA was set at 300 mV$\text{pp}$ per side, which corresponds roughly to 150 mV of DC voltage drop on the emitter degeneration resistors, $R_E$, of the differential pair in Fig. 4.10. The linearity of the pair depends on the input and output swing, respectively. The output swing must be small enough so that the output voltage at the collector node does not clip to the supply during the positive transition, and does not reduce the collector to emitter voltage below $V_{CEsat}$ during the negative transition, therefore saturating the device. On the other hand, the small signal approximation is valid at the input as long as the input signal swing does not exceed $V_T + I_T R_L / 2 R_E$. All these conditions are reflected in (4.12) and for the selected values of Fig. 4.10 result in a maximum input linear swing of 385 mV$\text{pp}$, which satisfies the linearity specifications.

### 4.2.2 MOS-HBT switch

The topology of the switch is shown in Fig. 4.8. It employs MOS-HBT cascode switches biased as a quasi-CML stage, which ensures high switching speed from a low voltage supply. In contrast to previously published designs [54, 52], there is no switched output buffer, even though this buffer helps improve droop rate by a factor of $\beta$. The reason behind this design choice was threefold:

1. First, the beta of the SiGe HBT, unlike that of Si bipolar and InP HBTS is over 500, so buffering is not needed because the base current is very small and there will be no drooping. The same applies if the output buffer is with MoSFETs. The drooping is also caused by lack of bandwidth in the output buffer.

2. Second, the switching buffer would load the clock amplifier and compromise switching speed. Alternatively, the clock amplifier would need to consume more power to drive four instead of two MOS-HBT switches.

3. Third, the addition of the switching buffers would double the power consumption of the THA.

The design of the switch proves to be quite challenging, since the designer is faced with conflicting requirements between the phases when the switch is activated (track-mode) and deactivated (hold-mode), which will be described below.

#### Track-mode

Fig. 4.11 shows the equivalent circuit when the THA operates in track-mode. During this mode, the track signal is held high while the hold signal is kept low, thus transistors Q1 are turned on, acting as
emitter followers and charging capacitor $C_H$. The accuracy to which the THA follows the input depends on the linear input buffer and the linearity of the switch. In principle the accuracy of the THA for a sinusoidal input can be quantified as the effective number of bits (ENOB)\cite{77}:

$$ENOB = \frac{SINAD - 1.76\,dB}{6.02}$$

(4.13)

where SINAD stands for Signal-to-Noise-and-Distortion. Although the emitter follower is considered to be the most linear of the three single stage amplifier configurations (common emitter, common base, emitter follower or common collector), its linearity is severely affected by the modulation of the base-emitter junction due to the input voltage swing. As a result, this modulation of the current, $i_{C_H}$, flowing into the hold capacitor would cause distortion in the output voltage equal to \cite{52}:

$$\Delta V_{out} = \Delta V_{in} - V_T \ln \left( \frac{I_{T2} + 2i_{C_H}}{I_{T2}} \right)$$

(4.14)

where $V_T$ is the thermal voltage equal to $kT/q$. Given that for a sinusoidal input $V_{in} = A\sin(2\pi ft + \phi)$ the maximum current through the capacitor is \cite{54}:

$$i_{C_{H\,\max}} = 2\pi f ACH$$

(4.15)
it can be shown that the second and third harmonic distortion are given by [52]:

\[ HD_2 = \frac{V_T(2\pi f_{AC_H})^2}{4I_{T2}(AI_{T2}/2 + V_T2\pi f_{AC_H})} \]  

\[ HD_3 = \frac{V_T(2\pi f_{AC_H})^3}{12I_{T2}(AI_{T2}/2 + V_T2\pi f_{AC_H})} \]

Since the topology of the THA is differential, it is expected that the second order distortion products will be suppressed. It is evident from equation (4.17) that the only ways to reduce the distortion is either by increasing the bias current, \( I_{T2} \), or by reducing the value of the hold capacitor, \( C_H \). Increasing the bias current would inevitably translate to higher power consumption. In addition, the size of the emitter follower devices, Q1, would have to be increased proportionally to ensure linear operation close to the peak \( f_T \) current density, and consequently the bandwidth of the THA would decrease. Reducing the hold capacitor, on the other hand, introduces significant hold-mode distortion.

Another important parameter is the time needed for the output of the track-and-hold amplifier to reach the input within a certain error band, when it enters track-mode. This is known as acquisition time [78] and depends on the slew rate, \( SR \), of the THA:

\[ SR = \frac{I_{T2}}{C_H} \]  

Since the maximum swing of the THA is 300 mVpp, the slew rate required for settling within half a period of the 90GHz clock is approximately 54 mV/ps. For 7-bit accuracy equation (4.13) gives a SINAD of 43.9 dB. Assuming that the SINAD is dictated mainly by third-order distortion and thus can be approximated reasonably well by the total harmonic distortion \( \text{SINAD} \approx \text{THD} = -43.9 \) dB (the negative sign is due to the definition of THD in [13]), the value of the bias current and hold capacitance can be optimized based on eq. (4.17) and (4.18) to simultaneously satisfy the slew rate and harmonic distortion requirement for a 36GHz input signal sampled at 90 GS/s. Solving the system of equations (4.17) and (4.18) under the constraint of the lowest possible power consumption that meets these requirements, resulted in a bias current of 4 mA and a hold capacitance of 60 fF. These choices give an \( HD_3 \) of -49.79 dB and a slew rate of 66 mV/ps.

Emitter follower devices Q1 were biased below their peak \( f_T \) current density (at 1 mA/\( \mu \)m) to ensure linear operation when in track-mode and avoid oscillations at high frequencies. Because these transistors are driving a relatively large capacitance from their emitters, there is always a risk of oscillation at high frequencies, which shows up as negative resistance at the base of the HBT device. Extensive S-parameter and transient simulations confirmed the absence of oscillations. The devices comprising the MOS-HBT switch were sized such that the MOSFETs reach 0.3 mA/\( \mu \)m and the HBTs 1.5 times their peak \( f_T \) current density when fully on.

Hold-mode

Fig. 4.12 shows the circuit operating in hold mode. In this case node "hold" is kept high and the current is directed away from the base of the emitter followers through M1 and M3, respectively. Transistors Q1 are off and \( C_H \) is isolated. Some important design considerations are:

- Adequate isolation of the hold capacitor from the input signal.
Fig. 4.12: THA in hold-mode. Deactivated parts of the circuit are displayed in grey.

- Minimize capacitive coupling of the input signal to the hold capacitor through the base-emitter capacitance of the emitter follower.
- Hold-step also known as sampling pedestal [13].
- Gradual discharge of the hold capacitor also known as droop-rate [13].

As shown in Fig. 4.12 when the hold signal is applied, all of the current $I_{T2}$ flows through M1 generating an additional voltage drop on the load resistor, $R_L$, of the linear input buffer. This voltage drop lowers the base voltage of Q1 to a point where it is completely turned off. This extra voltage drop can be approximated from [52]:

$$\Delta V_b, Q1 = R_L I_{T2} + V_T \ln \left( \frac{2I_{T2} + I_{T1}}{I_{T1}} \right)$$

(4.19)

For the values given in the schematic of Fig. 4.12 the $V_{BE}$ of Q1 drops by about 0.45 V, which is half the required $V_{BE,\text{ON}}$ of 0.9 V for this technology node. Given the fact that the reverse saturation current, $I_S$, is of the order of $10^{-15}$ A, the Ebers-Moll model [62] predicts a current $I_S e^{\frac{V_{BE}}{V_T}}$, which is of the order of a few nA and the switch is off. Therefore, the value of $R_L$ should be chosen carefully, so as to provide enough IR drop in hold-mode, but without increasing the output time constant of the linear input buffer during track-mode.

Even when the switch is nonconducting, there can be significant signal feedthrough from the linear input buffer at high frequencies due to the base-emitter parasitic capacitance. This capacitance along with $C_H$ form a capacitive divider resulting in [62]:

\[ \Delta V_b, Q1 = R_L I_{T2} + V_T \ln \left( \frac{2I_{T2} + I_{T1}}{I_{T1}} \right) \]
\[ V_{fth} = \frac{C_{be,Q1}}{C_H + C_{be,Q1}} V_{in} \]  \hspace{1cm} (4.20)

For the 60fF hold capacitor and the calculated \( C_{be,Q1} \) from (4.6), (4.20) results in 29\% feedthrough. In order to reduce hold-mode feedthrough equation (4.20) suggests maximizing the value of \( C_H \) or minimizing the value of \( C_{be,Q1} \). Due to the differential nature of the circuit two feedforward cross-coupled capacitors, \( C_{ff} \) can be employed to reduce this feedthrough to [52]:

\[ V'_{fth} = \frac{C_{be,Q1}}{C_H + C_{be,Q1}} \left( 1 - \frac{C_{ff}}{C_{be,Q1}} V_{in} \right) \]  \hspace{1cm} (4.21)

Theoretically hold-mode feedthrough can be completely canceled by choosing the value of the feedforward capacitors equal to the hold capacitors. Usually, the feedforward capacitors are implemented as diode-connected HBTs [52, 53]. This arrangement has the benefit that the feedthrough cancellation tracks in theory with process and temperature variations. In practice, since the feedthrough capacitance transistors are not biased, they do not have the same \( C_{be} \) as \( Q1 \). However, complete cancellation by using this scheme can negatively impact the bandwidth of the switch, because their effective value doubles due to the Miller effect [54], as confirmed by simulations for different feedthrough capacitance values plotted in Fig. 4.13. Hence, in this design the feedthrough cancellation capacitors have been realized with overlapping metal stripes in the top two metal layers. This scheme reduces parasitic coupling to the substrate and minimizes layout footprint. The value of 10fF was found through simulation after parasitic extraction of the laid-out track-and-hold amplifier.

![Fig. 4.13: Effect of feedthrough capacitor, C_{ff}, on bandwidth.](image)

During the transition from track to hold-mode, some charge is dumped on the hold capacitor distorting the input value being held. The amount of the voltage change, \( \Delta V_H \text{step} \) due to this effect depends, at least to a first order, on the value of the hold capacitor and should be kept lower than the voltage
Chapter 4. Track & Hold Amplifier with New Quasi-CML MOS-HBT Switch

corresponding to the least significant bit ($V_{\text{LSB}}$) \[69\]:

$$\Delta V_{\text{H step}} \approx \frac{I_T T_{\text{off}}}{C_H} < 1 V_{\text{LSB}} \quad (4.22)$$

where $T_{\text{off}}$ is the time required for the switch to turn off. It can be shown\[52\] that theoretically hold pedestal errors can be eliminated when the output is taken differentially because equal charges are being dumped on the hold capacitors. Nevertheless, in a real circuit these errors cannot be eliminated completely due to the nonlinear dependence of the base-emitter junction on the input voltage being applied on Q1. In practice, perfect cancellation is impossible because the two switches start turning off from slightly different bias points, since one sees the input signal out of phase with respect to the other. On top of that, slight clock skew between them causes inevitably some errors. As can be observed from equation (4.22), a sufficiently large value of $C_H$ helps minimize the hold-step. This requirement comes in direct contradiction with the low hold capacitance value predicted by equation (4.17) for low-distortion operation in track-mode.

While the sampled value is being held, charge is gradually leaking from the hold capacitor changing the value initially stored. This effect is more pronounced in single-ended designs and reduces basically to a common-mode effect in differential implementations \[52\]. The discharge of the hold capacitance is mainly due to the bias input current of the stage following the THA. In the case of an output driver with an HBT differential pair, the amount of droop rate (DR) can be quantified from \[69\]:

$$DR = \frac{\Delta V_H}{\Delta T} = \frac{I_B}{C_H} \quad (4.23)$$

where $I_B$ is the base current of the next stage. Given the fact that in modern SiGe BiCMOS processes the $\beta$ of the HBT device exceeds 500, droop rate is not a big concern at high sampling rates because the value of the hold capacitors gets refreshed very frequently (unless an extremely small hold capacitance value is used). For the 50\,$\Omega$ output driver depicted in Fig. 4.18, the base current, $I_B$, equals to $I_E/(\beta+1) = 8\,\mu\text{A}$ resulting in a droop rate of 0.13\,$\text{V/ns}$. At this rate the minimum sampling rate that yields an error less than 1 LSB (300\,$\text{mV}/2^7$) is 30\,$\text{GHz}$. Droop rate was found to be at acceptable levels, so no extra measures were taken to mitigate this effect. If droop rate, however, proves to be a serious problem, an output buffer based on a MOS differential pair can help mitigate this effect, because the gate draws practically no current. Other solutions include switched buffers that are enabled with the sampling clock \[52\,\text{[54],53}\].

The layout of the THA core consisting of the switched emitter follower device Q1, the hold capacitors, $C_H$, the MOS-HBT switches, and the feedthrough cancellation capacitors, $C_{ff}$, is shown in Fig. 4.14. A compact layout is of paramount importance at very high frequencies to minimize parasitic capacitances. All active devices have been placed at minimum distance from each other. As mentioned earlier, the hold capacitors were formed as MIM capacitors in the upper layers of the metal stack to reduce substrate noise coupling. The small feed-forward capacitors, whose role is to reduce clock feedthrough, are implemented by overlapping metal stripes in the top two metal layers. Careful attention was paid to symmetry to avoid mismatches in the signal paths and minimize systematic clock skew.
4.2.3 Clock Amplifier

If the THA is intended to work only around a specific sampling frequency, then a tuned clock amplifier network can be used that provides maximum gain at the frequency of interest. This choice means that the clock amplifier can be AC-coupled to the MOS-HBT switch of the switched emitter follower core in Fig. 4.12. Apart from lower power consumption, this configuration makes the operation of the switch less sensitive to PVT variations, since the gate DC voltage of the quasi-CML MOS-HBT cascode does not depend on the common-mode output voltage of the preceding emitter follower circuit [66]. However, in this case, it would be impossible to test the THA over different sampling rates, so a broadband DC-coupled clock amplifier was chosen for this design.

![Clock Amplifier Schematic](image)
The schematic of the proposed clock amplifier is shown in Fig. 4.15. It consists of three cascaded differential common-emitter amplifier stages and an emitter follower to drive the MOS-HBT switch of the THA. The common-emitter topology was chosen because it can operate with less than 1.8V supply. To compensate for its lower bandwidth, inductive peaking was employed. Because it is very difficult to provide perfectly differential broadband signals up to 100GHz in a lab test environment, the input to the clock amplifier is single-ended. This implies that the clock amplifier must have very good common mode rejection up to 100GHz, an extremely difficult design challenge. To provide adequate common mode rejection at these frequencies several cascaded amplifier stages were necessary. Each stage is scaled accordingly to reduce power consumption, improve bandwidth and provide the required voltage swing to the gates of the MOS-HBT switches of the THA.

Starting from the last stage in the clock amplifier chain, an 8mA emitter follower was employed that acts as a level shifter providing the proper common-mode level to the gates of the MOS-HBT switches. An 8mA bias current was chosen because it was found in simulations that this way the follower provided sufficient drive and bandwidth for fully switching the switched emitter follower. Since the MOS-HBT differential pair requires at least 400mV<sub>pp</sub> in the worst case corner to fully switch the current to one side, a 6mA bipolar CML inverter with about 600mV<sub>pp</sub> output swing precedes the emitter follower (more swing than the minimum of 400mV<sub>pp</sub> was budgeted for to account for PVT). The 6mA bipolar CML inverter is driven by a 4mA version with reduced output swing (250mV<sub>pp</sub>) which is the minimum swing to ensure full switching of the bipolar pair of the 6mA CML inverter over PVT variations. The values of the load resistors and inductors have been scaled down to account for the larger bias current. This results in smaller capacitance both at the input and output of this stage improving the bandwidth of the cascade. Because the common mode rejection provided by the 4mA and 6mA bipolar stages was found to be inadequate, a 2mA bipolar CML inverter was introduced at the input. The maximum fanout between stages is just 2 as required for high-speed designs.

The HBTs were sized such that they reach 1.5 times their peak f<sub>T</sub> current density when fully switched. Notice that bipolar tail current sources were employed everywhere in the clock amplifier chain to increase common mode rejection. The DC voltage drop on the emitter resistors of the tail sources is 150mV, which has been proven to make the bias current value more stable over temperature and process variations [33]. All load resistors and the degeneration resistors of the tail current sources were made out of the same type of 12.5Ω resistor, so that the design is less sensitive to process variations. Each tail current source consists of a parallel combination of the same 1mA bipolar current source.

Since the clock is the highest speed signal in the entire system, the layout of the clock amplifier was optimized to ensure minimal clock skew and large bandwidth. The inductors were designed as multi-layer stacked spirals with minimum spacing to increase inductance and a maximum footprint less than 10µm×10µm to keep the SRF above 200 GHz as required for 100GHz operation. At high frequencies every piece of interconnect introduces RC delays slowing down the circuit, so the length of the cascade of the CML inverters had to be minimized by placing the current sources symmetrically on the sides and passing the load resistors at the center of the inductors as shown in Fig. 4.16. This arrangement reduces the amount of interconnect along the signal path and does not affect the magnetic field of the peaking inductors. All inductors and interconnects were modeled using the electromagnetic simulator EMX by Integrand®.
Fig. 4.16: Layout of the clock amplifier.
The results of the AC simulations after parasitic extraction confirm that the clock amplifier has broadband gain from DC up to approximately 100 GHz, when connected to the switched emitter follower core as shown in Fig. 4.17. However, no flat region is observed in the AC response, implying that the bandwidth can be further increased with more inductive peaking. A redesign of the amplifier with larger inductor values yields an improved response as confirmed by the dashed line in Fig. 4.17 (due to the limited time available the improved version of the clock amplifier was not ready before the submission deadline, and thus was not included in the tapeout).

![Fig. 4.17: AC simulation results of the clock amplifier after parasitic extraction.](image)

### 4.2.4 50Ω Output Driver

The output of the track-and-hold amplifier is buffered via a 50Ω driver for easy interfacing with the 50Ω testing environment. In a real system implementation the THA will most likely drive a capacitive load like the input of a comparator in a SAR ADC for example. Some important design considerations for the output driver are:

- Output match to 50Ω.
- Minimum linearity better than that of the THA.
- Minimum bandwidth better than that of the THA. This also helps avoid droop degradation among other things.
- Low input current to avoid droop degradation.
- Low-power operation from 1.8V supply.

The output driver is depicted in Fig. 4.18. The DC output voltage of the switched emitter follower preceding the driver is about 1.2 V. Since the maximum DC voltage allowed by the technology on any of the terminals of a MOS device cannot exceed 1 V, a MOS-HBT output driver could not be used, even though it would improve the droop rate. As a result, an HBT differential pair with emitter degeneration had to be employed. Because $V_{BE,ON} = 0.9$ V, the degeneration resistor, $R_E$, was put between two current sources. Although this scheme requires more power, it leaves about 0.3 V as headroom, which is plenty for the low-$V_T$ MOS devices comprising the current sources. The load resistors, $R_L$, were set to 50 $\Omega$ for output matching reasons allowing for an output swing of 200 mV$_{pp}$. Accounting for the parasitic emitter resistance, $r_E$, the overall gain of this stage is approximately equal to 1. The HBT devices are biased close to their peak $f_T$ current density of 1.5 mA/$\mu$m and peaking inductors were employed to increase the bandwidth of this stage when driving a 50$\Omega$ load.

![Schematic of the 50$\Omega$ output driver.](image)

It can be proven that the third order distortion of the degenerated bipolar pair is [69]:

$$HD_3 = \frac{16V_T\Delta V_{in}}{12I_E^3(4V_T/I_E + R_E)^3} \quad (4.24)$$

where $I_E$ is the emitter current per side (4 mA) and $\Delta V_{in}$ the differential input voltage. A simpler approximate expression for the linearity of this stage is given by [61]: $I_E R_E$. The DC voltage drop on $R_E$ is $I_E R_E = 88$ mV which falls short of the 300 mV$_{pp}$ (or 150 mV per side) input linearity requirement. In order to increase the linearity and output swing, the current would have to be increased because $R_L$ is fixed to 50 $\Omega$ and the value of $R_E$ needs to stay in a certain ratio to $R_L$ to keep the gain equal $\geq 1$. 
Keeping the current density constant, this would inevitably result in bigger device sizes and lead to bandwidth limitations because the switched emitter follower device, Q1, is rather small (only $2 \mu m$). A solution would be to use a higher power output driver and an intermediate chain of appropriately scaled buffers from the output of the THA to the input of the driver. Apart from the significant increase in total power consumption, maintaining sufficient linearity when more stages are added can prove to be quite challenging and was not pursued due to limited time before the tapeout deadline.

The layout of the fabricated THA is illustrated in Fig. 4.19. The analog differential input is on the left and the output of the right hand side. The single ended clock is coming from the south perpendicular to the signal path to minimize coupling between them. The 1.8 and 2.5V supply pads are on the top along with a bias pad that adjusts the reference current for all mirrors on the chip (not shown in Fig. 4.19). Since the clock is DC-coupled the THA can be set to hold or track-mode simply by changing the DC voltage on the clock pad. If the clock is connected through a decoupling capacitor, the outputs of the clock amplifier force the MOS-HBT switches of the THA in a balanced state, where the current is split evenly between the two branches.
Fig. 4.19: Layout of the proposed THA (pads are not shown).
4.3 Simulations

This section presents some post-layout simulations of the designed track-and-hold amplifier. Unless otherwise specified, the simulations were run under typical corner and 65 °C.

4.3.1 Small signal S-parameters in track-mode

The simulated S-parameters in track-mode are plotted in Fig. 4.20. The single-ended gain denoted by \(|S_{21}|\) starts at about -5.3 dB and has a small signal 3dB bandwidth of 40 GHz in the slow corner (corners are defined in Table 3.2). If the signal was tested differentially, the gain would increase by 6 dB resulting in 0.7 dB. Input matching denoted by \(|S_{11}|\) remains better than -10 dB up to 75 GHz and \(|S_{22}|\) (output matching) is lower than -10 dB up to 50 GHz. The matching of the single-ended clock is given by \(|S_{55}|\) and stays better than -15 dB across the entire range of simulated frequencies. The variation of the small signal bandwidth as quantified by \(S_{21}\) is plotted in Fig. 4.21.

![Fig. 4.20: Single-ended simulated S-parameters of the track-and-hold amplifier in track-mode.](image-url)
4.3.2 Noise in track-mode and hold-mode

Fig. 4.22 shows the simulated noise transfer function, when the circuit is set in track-mode, and Fig. 4.23 the equivalent input noise when it is sampling at various sample rates. The simulated integrated equivalent input noise voltage from 100 kHz to 50 GHz is 950 $\mu$V$_{\text{rms}}$. Table 4.2 summarizes the main noise contributors reported from a noise analysis simulation in track-mode. The effect of the sampled noise is captured by a periodic steady state (PSS) simulation and the results are plotted in Fig. 4.23 for various sampling rates. Notice that because of the noise being folded into the signal band, the input referred noise when sampling exceeds the noise in track-mode.

<table>
<thead>
<tr>
<th>Component</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_L$ (input buffer)</td>
<td>6.65%</td>
</tr>
<tr>
<td>$R_E$ (input buffer)</td>
<td>4.29%</td>
</tr>
<tr>
<td>50 $\Omega$ load (output buffer)</td>
<td>4%</td>
</tr>
<tr>
<td>SEF (Q1)</td>
<td>3.35%</td>
</tr>
</tbody>
</table>

Table 4.2: Top 4 noise contributors.
Fig. 4.22: Simulated noise transfer function after extraction of layout parasitics.

Fig. 4.23: Simulated equivalent input referred noise for various sampling rates.
4.3.3 Hold Pedestal

To simulate the effect of hold pedestal a methodology presented in [69] was followed, which is summarized below:

1. A DC voltage was applied as the input to the THA. This eliminates any issues related to signal feedthrough through $C_{be}$.

2. The THA was set in track-mode at the beginning of the simulation and later switched to hold-mode.

3. The voltage step that appears after entering track-mode is the pedestal error of the circuit.

4. The process is repeated for different input DC values.

Table 4.3 summarizes the hold step error for different inputs.

<table>
<thead>
<tr>
<th>$V_{in}$ [mV]</th>
<th>Pedestal [mV]</th>
<th>LSB error</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>5.98</td>
<td>1.28</td>
</tr>
<tr>
<td>200</td>
<td>3.28</td>
<td>0.55</td>
</tr>
<tr>
<td>100</td>
<td>1.64</td>
<td>0.28</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-100</td>
<td>1.64</td>
<td>0.28</td>
</tr>
<tr>
<td>-200</td>
<td>3.28</td>
<td>0.55</td>
</tr>
<tr>
<td>-300</td>
<td>5.98</td>
<td>1.28</td>
</tr>
</tbody>
</table>

Table 4.3: Pedestal error for various differential inputs.

4.3.4 Large Signal Transient and Periodic Steady State (PSS)

The circuit operation was verified using transient simulations for various input frequencies and sampling rates. Differential and single-ended simulated output waveforms for a sinusoidal input signal at 18 GHz with an amplitude of 300 mV$_{pp}$ sampled at 90 GS/s is illustrated in Fig. 4.24 and 4.25, respectively. The waveforms before and after the 50Ω output driver are shown to indicate the effect of bandwidth and linearity of the driver. The output of the driver looks "rounded-off" due to insufficient bandwidth.

Plotting the sampled spectrum of a signal requires attention when setting up the Discrete Fourier Transform (DFT) simulation in Cadence. An important source of error has to do with the interpolation and end points of the waveforms as explained [79]. Errors like these reduce the dynamic range of the FFT. A technique that avoids these errors, increases the spectral resolution of the FFT and obviates the need for windowing is coherent sampling [80]. During transient simulation care must be taken to ensure that the simulator saves the sampled points at equally spaced time intervals (FFT assumes that all samples are equally spaced in time) and that these samples correspond to the analog values at the middle of the hold interval of the THA, so that any settling phenomena have died out. Using the above technique the sampled data was saved from Cadence and further processed in Matlab [81]. A 15.95GHz input coherently sampled at 64 GHz is plotted in Fig. 4.26 and Fig. 4.27, respectively.

Different spectra were simulated for different input frequencies and sampling rates. Fig. 4.28 shows the benefits of oversampling; the SNR is improved by roughly 1.6 dB when the same signal is sampled
Fig. 4.24: Differential output of a full-scale (300 mV_{pp} per side) 18GHz sinusoidal signal sampled at 90 GS/s. The dashed sinusoid corresponds to the sampling clock, the solid line to the differential output of the switched emitter follower, and the dashed line to the output of the output driver.

Fig. 4.25: Single-ended outputs of a full-scale (300 mV_{pp} per side) 18GHz sinusoidal signal sampled at 90 GS/s. The dashed sinusoid corresponds to the sampling clock, the solid lines to the outputs of the switched emitter follower, and the dashed lines to the outputs of the output driver.

at 80 GS/s instead of 64 GS/s. This effect is commonly employed in ΔΣ ADCs to reduce the in-band quantization noise.

To estimate the large signal bandwidth (also known as sampling bandwidth [82]) of the THA, a 50mV input signal at frequencies ranging from 1 GHz to 40 GHz were sampled at various frequencies (40 GHz-90 GHz). The amplitude was kept small to ensure linear operation of the amplifiers used in the circuit. The large signal bandwidth is estimated as the input frequency where the harmonic power drops
Chapter 4. Track & Hold Amplifier with New Quasi-CML MOS-HBT Switch

Fig. 4.26: Time-domain waveform of a 15.95GHz sinusoidal input sampled at 64 GS/s.

Fig. 4.27: 4096-point FFT spectrum of a 15.95GHz sinusoidal input sampled at 64 GS/s.
Fig. 4.28: Improvement of SNR due to oversampling: 4096-point FFT of a 20GHz sinusoid coherently sampled at (a) 64 GS/s (bin: 1283) and (b) 80 GS/s (bin: 1021).
Fig. 4.29: Hanning window 1024-point FFT for different input frequencies sampled at 40 GS/s.

by 3 dB. It is clear from the simulation results in Fig. 4.29 and 4.30 that higher sampling rates offer greater bandwidth. For example, in the 40 GS/s case of Fig. 4.29, the 3 dB drop occurs before 25 GHz, while in the 40 GS/s case of Fig. 4.30, it occurs before 40 GHz.

Fig. 4.30: Hanning window 1024-point FFT for different input frequencies sampled at 90 GS/s.

Finally, the linearity of the THA circuit was assessed through PSS simulations at different input frequencies and sample rates. The THA achieves a 1dB input compression point ($P_{1dB}$) of -4.2 dBm as plotted in Fig. 4.31 for a 10 GHz input signal sampled at 90 GS/s.
Fig. 4.31: Simulated 1dB input compression point ($P_{1dB}$) for a 10GHz input signal sampled at 90 GS/s (extracted from -50 dBm).
4.3.5 Summary of Results

The table below summarizes some of the simulations results of the track-and-hold amplifier.

### Table 4.4: Simulated DC parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Conditions</th>
<th>Min</th>
<th>Typ</th>
<th>Max</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{2p5}$</td>
<td>2.5V supply</td>
<td></td>
<td>2.25</td>
<td>2.5</td>
<td>2.75</td>
<td>V</td>
</tr>
<tr>
<td>$V_{1p8}$</td>
<td>1.8V supply</td>
<td></td>
<td>1.65</td>
<td>1.8</td>
<td>1.9</td>
<td>V</td>
</tr>
<tr>
<td>$I_{ref}$</td>
<td>DC bias current</td>
<td></td>
<td>0.95</td>
<td>1</td>
<td>1.05</td>
<td>mA</td>
</tr>
<tr>
<td>$P_{DC}$</td>
<td>power dissipation</td>
<td></td>
<td>62</td>
<td>89</td>
<td>122</td>
<td>mW</td>
</tr>
<tr>
<td>$T$</td>
<td>temperature</td>
<td></td>
<td>25</td>
<td>65</td>
<td>125</td>
<td>°C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Conditions</th>
<th>Min</th>
<th>Typ</th>
<th>Max</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_s$</td>
<td>sampling rate</td>
<td></td>
<td>10</td>
<td>64</td>
<td>90</td>
<td>GHz</td>
</tr>
<tr>
<td>$f_{3dB}$</td>
<td>3dB small signal bandwidth</td>
<td>track-mode</td>
<td>40</td>
<td>49</td>
<td>57</td>
<td>GHz</td>
</tr>
<tr>
<td>Insertion loss</td>
<td>loss from input to output</td>
<td>track-mode differentially</td>
<td>-1</td>
<td>dB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{11}$</td>
<td>input return loss</td>
<td>up to 50 GHz</td>
<td>-14</td>
<td>dB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{33}$</td>
<td>output return loss</td>
<td>up to 50 GHz</td>
<td>-10</td>
<td>dB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{44}$</td>
<td>CLK return loss</td>
<td>up to 75 GHz</td>
<td>-15</td>
<td>dB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$IIP_{35GHz}$</td>
<td>Intercept Point</td>
<td>switching-mode @ 35GHz input</td>
<td>+6</td>
<td>+8</td>
<td>dBm</td>
<td></td>
</tr>
<tr>
<td>$P_{1dB_{35GHz}}$</td>
<td>Input referred compression point</td>
<td>switching-mode @ 35GHz input</td>
<td>-4</td>
<td>-2</td>
<td>dBm</td>
<td></td>
</tr>
<tr>
<td>LO leakage</td>
<td></td>
<td>switching-mode</td>
<td>-60</td>
<td>dBc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO amplitude $10GHz$</td>
<td>amplitude at 10GHz clock</td>
<td>switching-mode</td>
<td>-16</td>
<td>-10</td>
<td>dBm</td>
<td></td>
</tr>
<tr>
<td>LO amplitude $75GHz$</td>
<td>amplitude at 75GHz clock</td>
<td>switching-mode</td>
<td>-4</td>
<td>0</td>
<td>dBm</td>
<td></td>
</tr>
<tr>
<td>$V_{n,in}$</td>
<td>integrated equivalent input referred noise</td>
<td>track-mode</td>
<td>0.95</td>
<td>mV$_{rms}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differential Droop Rate</td>
<td>droop rate due to leakage</td>
<td>switching-mode</td>
<td>0.67</td>
<td>mV/µs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SFDR_{41GS/s}$</td>
<td>SFDR at 64 GS/s</td>
<td>10 GHz full scale differential input, 64 GHz clock</td>
<td>42</td>
<td>dB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SFDR_{10GS/s}$</td>
<td>SFDR at 10 GS/s</td>
<td>35 GHz full scale differential input, 10 GHz clock</td>
<td>30</td>
<td>dB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SFDR_{75GS/s}$</td>
<td>SFDR at 75 GS/s</td>
<td>35 GHz full scale differential input, 75 GHz clock</td>
<td>28</td>
<td>dB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.5: Simulated AC parameters.
Chapter 5

ADC Front-end

This chapter discusses the use of the track-and-hold amplifier presented in Chapter 4 in an ADC front-end sampling system. In the beginning, some background information on time-interleaved systems is provided. Next, the idea of quadrature sampling clocks is introduced to extend the sampling rate capabilities of the system. Finally, a 128GS/s sampling front-end prototype is described that could be used as part of a time-interleaved SAR ADC. This version was designed solely for assessing the layout complexity and loading effect of the multiple THAs on the bandwidth performance. However, it does not offer true time-interleaved functionality as it lacks a complete multiphase, clock distribution network.

5.1 Time-interleaving Considerations

Fig. 5.1 illustrates a giga-sampling front-end for an ADC system. By using multiple THAs in a master-slave configuration (forming effectively a sample-and-hold amplifier similar to the one presented in [51]) slower ADCs can be time-interleaved to increase the sampling rate of the A/D system. Ultimately, its sampling rate would be dictated by the maximum rate of operation of the master S/H. It has been shown [83] that the use of a master can significantly relax the timing skew and bandwidth mismatch requirements of the subsequent samplers, because the input has already been stored by the master. Such timing skew errors are more pronounced near the zero-crossing regions of the input signal and show up as phase modulated noise ($f_{PM}$) in the time domain [84]. This type of noise degrades the SNR with increasing input frequency and in the frequency domain causes periodic spurs at fractions of the sampling frequency as predicted by [84]:

$$f_{PM} = \pm f_{in} + \frac{k}{N} f_s, \quad k = 1, 2,...$$  \hspace{1cm} (5.1)

where $f_{in}$ is the input signal frequency, $N$ the number of interleaved stages, and $f_s$ the sampling frequency. Even though moderate timing skew errors and bandwidth mismatches can be tolerated when a master S/H is used, the bandwidth of each individual slave sampler must be sufficient to capture the information stored in the master, otherwise the slave would not have settled to the correct value before the master saves the next sample [85]. Since the sampled output spectrum of the master contains information up to $f_s/2$, using the previously demonstrated THA with 40GHz bandwidth a sampling front-end with a maximum rate of 80GS/s (80GHz clocked master THA) can be built that satisfies the
Nyquist criterion. The timing waveforms of a front-end similar to the one depicted in Fig. 5.1 are shown in Fig. 5.2, where it is assumed that the held signal has settled to its final value in a half period of the sampling clock.

It is evident by Fig. 5.2 that a new sample is captured every $T_s$ seconds, where $T_s$ is the sampling period of the master S/H. Can we do even better than this? The answer is yes! By using multiple master samplers in parallel we can sample several times during the same sampling period and thus increase the overall sampling rate of the A/D system. This can be achieved by using a multiphase clock for the first rank of masters, so that each sampler captures one sample at a time within the clock period. An example of such an arrangement that doubles the effective sampling rate is shown in Fig. 5.3, where the role of the multiphase clock is assumed by a quadrature hybrid tuned to the frequency of interest. The hybrid produces two versions of the sampling clock that are 90° out of phase. The in-phase clock, $clk_i$, drives one master and synchronizes its subsequent slave samplers, while the quadrature clock, $clk_q$, drives the second master after $T_s/4$. The ideal timing waveforms of such a system are shown in Fig. 5.4.
Fig. 5.2: Timing waveforms of an ideal time-interleaved front-end similar to Fig. 5.1. Crosses represent the samples captured within one period of the sampling clock $f_s$. 
Fig. 5.3: ADC front-end with quadrature sampling clocks for doubling the effective sampling rate.
Fig. 5.4: Timing waveforms of quadrature sampling clock architecture in Fig. 5.3.
5.1.1 Circuit Description

To estimate the input bandwidth and layout complexity of such a front-end for high-speed ADCs, a test circuit was designed that incorporates two master THAs, 18 slave THAs and a quadrature hybrid tuned at 64 GS/s. The conceptual block diagram is provided in Fig. 5.5. In contrast to the system shown in Fig. 5.1 and 5.3, the master and slave samplers are implemented as THAs. This means that the master THA and one of the slaves at each time form a sample-and-hold (S/H) circuit. However, this arrangement has no implication on the bandwidth of the front-end, since each THA sees the same loading as in Fig. 5.1. The quadrature hybrid produces the I and Q clocks which drive the two master track-and-hold amplifiers. Each amplifier is loaded by nine identical slaves. In a high-speed digital acquisition system each slave would drive one of the 9 lanes in a 7-bit, time-interleaved SAR ADC in a similar manner as explained in [83]. The slave as well as the master THAs are made up of the same THA discussed in Chapter 4. This arrangement presents a worst case scenario; the size and power of the slave track-and-hold amplifiers can be reduced to ease the loading on the master and will be pursued in a future version.

Out of the 9 slave THAs in every path (I or Q path) only one slave is fully active in the sense that it can be independently set either to track or hold-mode. The rest 8 of them are set to hold-mode by default via resistive dividers. This scheme reflects the real operation of such a time-interleaved front-end, where only one slave tracks the voltage kept by the master, while the rest are in hold-mode until their subsequent ADCs finish the conversion of the previous analog samples. On the other hand, since 8 out of 9 slave THAs are deactivated, the front-end cannot operate in a time-interleaved manner. In the layout each master THA sits at the center of the array of slave THAs as shown in Fig. 5.6. Notice that in the I-path the active slave lies at the bottom of the top array of slaves, which represents the longest path from the master to the slave THA. Likewise, the active slave THA in the Q-path lies at the top of the bottom array of slaves. The output of the active slave THAs is buffered by the same 50Ω driver as described previously in 4.2.4.

Looking at Fig. 5.6 the front-end breakout occupies 1070 µm × 895 µm. The 64GHz clock and the single-ended analog inputs come from the left-hand side, while the single-ended I and Q outputs are on the right-hand side (the other ends are terminated on chip via a 50Ω resistor, in order to reduce the number of output pads). The long piece of interconnect connecting the quadrature hybrid to the master and slave THAs was modelled as an RLC network based on the metal layer parameters, such as resistivity, conductor thickness, dielectric constant, inductance per length, etc, that the design kit reported. Although this breakout was designed solely for the purposes of bandwidth evaluation and layout complexity estimation, real-time operation can be tested by inserting a power divider off-chip, splitting the input signal to the I and Q-path, respectively. This way within one period of the 64GHz sampling clock one sample is captured by the I-path samplers (top half) and another is captured a quarter of a period later by the Q-path (bottom half). The chip uses a 1.8 and 2.5V supplies. Because the current consumption through the 2.5V was 219 mA, which exceeded the maximum allowable DC per pad (approximately 100 mA per pad), two power pads had to be used for this supply. Additionally, two bias pads provide the 1mA stable current reference for biasing the circuitry in the I and Q-path, respectively. In order to be able to individually control the state of the master and slave samplers of the front-end, two separate pads are provided (track_master and track_slave) which can be used to set both to track-mode for S-parameter measurements. During normal operation the DC level of these external pads is adjusted to 1.8 V, so that the inputs of the clock amplifiers driving the master and slave THAs are
perfectly balanced. This ensures that the master and slave THA will switch 180° out of phase according to the sampling clock applied. The clock amplifiers used are identical to the one described in Chapter 4. A short description of the I/O pads is provided in table 5.1.

Fig. 5.5: ADC front-end block diagram.
Fig. 5.6: ADC front-end layout. Size is: 1070 µm × 895 µm.
<table>
<thead>
<tr>
<th>Pin Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>inI</td>
<td>50Ω analog input</td>
<td>input sampled by in-phase clock. AC-coupled 300 mV pp max.</td>
</tr>
<tr>
<td>inQ</td>
<td>50Ω analog input</td>
<td>input sampled by quadrature clock. AC-coupled 300 mV pp max.</td>
</tr>
<tr>
<td>clk</td>
<td>50Ω sampling clock input</td>
<td>AC-coupled single-ended clock at least 300 mV pp swing.</td>
</tr>
<tr>
<td>track_master</td>
<td></td>
<td>DC-coupled: 1.8 V for normal operation, 1.5 V for track-mode</td>
</tr>
<tr>
<td>track_slave</td>
<td></td>
<td>DC-coupled: 1.8 V for normal operation, 2.1 V for track-mode</td>
</tr>
<tr>
<td>sIp</td>
<td>50Ω analog output</td>
<td>input for setting master THAs in track-mode</td>
</tr>
<tr>
<td>sQp</td>
<td>50Ω analog output</td>
<td>input for setting active slave THAs in track-mode</td>
</tr>
<tr>
<td>v2p5</td>
<td>I/O</td>
<td>single-ended output of I-path</td>
</tr>
<tr>
<td>v1p8</td>
<td>I/O</td>
<td>single-ended output of Q-path</td>
</tr>
<tr>
<td>IrefQ</td>
<td>I/O</td>
<td>1mA stable current reference for Q-path</td>
</tr>
<tr>
<td>IrefI</td>
<td>I/O</td>
<td>1mA stable current reference for I-path</td>
</tr>
<tr>
<td>GND</td>
<td>I/O</td>
<td>Analog ground</td>
</tr>
</tbody>
</table>

Table 5.1: ADC front-end I/O pin description.

5.1.2 Quadrature Hybrid

The schematic of the quadrature hybrid is shown in Fig. 5.7 and is based on the lumped hybrid coupler design methodology first presented in [86]. It consists of two monolithic coupled inductors modeled as a transformer in EMX electromagnetic simulator software and four 17fF MOM capacitors. Since the hybrid is a 3-terminal device, the other end of the quadrature coil is grounded through a 50Ω resistor. Special attention was given to the symmetry of the layout to avoid systematic phase mismatches as shown in Fig. 5.8. Post layout S-parameter simulations reveal a perfect 90° phase difference under typical conditions at exactly 64 GHz as illustrated in Fig. 5.9. It is evident from the simulations that the phase error remains at most ±5° from 55 GHz to 75 GHz suggesting that the operation of the hybrid can be relatively wideband.

![Fig. 5.7: 64GHz quadrature hybrid schematic](image)
Fig. 5.8: 64GHz quadrature hybrid. Size is: 100µm×110µm.

Fig. 5.9: Simulated phase difference of quadrature hybrid.
5.1.3 Simulations

Setting both the master and slave track-and-hold amplifiers in track-mode the insertion loss from inI to sIp (S_{31}) and from inQ to sQp (S_{42}) was simulated under typical corner at 65°C. Simulation results confirm the symmetry of the two paths (I and Q) and report a 3dB bandwidth of approximately 33.8 GHz as shown in Fig. 5.10. An example of the simulated coherently sampled spectrum of a 10.016GHz sinusoid sampled at 64 GS/s on the I-path is provided in Fig. 5.11.

![Simulated S-parameters](image)

Fig. 5.10: Simulated 3dB bandwidth from S-parameters.
Fig. 5.11: 4096-point FFT simulated coherently sampled spectrum of a 10.016GHz sinusoidal input sampled at 64 GS/s (bin: 641) on the I-path.
Chapter 6

Experimental Results

This chapter describes the experimental setups and shows the measured results of the TIA designs presented in Chapter 3, the quasi-CML track-and-hold amplifier and ADC front-end test circuit from Chapter 4. The chapter finally compares the measured performance of the circuits with simulation results and against state-of-the-art designs reported in the literature. The technology used for all circuits is ST Microelectronics 55nm SiGe BiCMOS technology [42], which was described in Chapter 3. Due to issues with proper operation of the thermal chuck in the lab at the time of testing, all measurements discussed in this chapter were conducted at room temperature (25°C). Approximately 5 - 7 dies were tested for each circuit and negligible variation in their measured performance was observed both in S-parameters and time-domain measurements.

6.1 TIA Breakouts

6.1.1 80 GHz Low-Power TIA

The micrograph of the manufactured amplifier is shown in Fig. 6.1. This version, without emitter degeneration, is for low-cost datacomm applications with NRZ signals beyond 100 Gb/s. It has a pad-limited area of 0.138 mm² and an active area of just 0.023 mm², while consuming only 9 mA from a 1.2 V supply when biased at its optimum NF current density ($J_{OPT}$).

S-parameter measurements were conducted in the DC-170 GHz range using a Keysight N5227A PNA [87] and OML [88] extender modules. The laboratory setups are shown in Fig. 6.2 and 6.3 below. Fig. 6.4 compares the simulated and measured S-parameters and noise figure across all frequency bands. The amplifier exhibits 13 dB gain and a 3dB bandwidth of 80 GHz. The discontinuities in the measurements close to 70 GHz and 110 GHz are because of the change in the measurement setup from DC to W-band and W-band to D-band, respectively. The measured phase of $S_{21}$ in the low frequency range was used to calculate the group delay of the amplifier and is presented in Fig. 6.5. The latter exhibits a maximum variation of ±3 ps from DC to 67 GHz and agrees with the simulation results.
Fig. 6.1: Fabricated TIA die micrograph.

Fig. 6.2: S-parameter setup for the DC-67 GHz range.

Fig. 6.3: S-parameter setup for the W and D-band.
Fig. 6.4: Comparison of the simulate and measured S-parameters and noise figure.

Fig. 6.5: Measured group delay.
Noise figure measurements were conducted with an Agilent N8975A Noise Figure Analyzer (NFA) up to 26.5 GHz [89]. An Agilent K88 downconverter was used along with the NFA to perform noise frequency measurements in the range 77-88.5 GHz. The block diagrams of both setups are reproduced in Fig. 6.6 and 6.7 respectively. The Elva W-band noise source comes with a built-in isolator to minimize reflections. The NFA was first calibrated on a through-port and loss compensation was applied to eliminate any noise associated with the cables and the downconverter. Even though there is a discrepancy between simulated and measured noise figure due to excessive ground resistance in the layout, the measured noise figure does not exceed 7dB.

Fig. 6.6: Noise setup up to 26.5 GHz.

Fig. 6.7: Noise setup for W-band.

The linearity of the amplifier was measured simply to compare with the linear version discussed in the next section which employs emitter degeneration. In order to characterize the linearity of the amplifier, a Keysight E4448 spectrum analyzer [90] was used in conjunction with a Keysight E8257D signal source.
Finally, eye diagram measurements were carried out at 40 Gbps using a Centellax TG1P4A PRBS generator and an Agilent Infinium DCA-X 86100D sampling oscilloscope. 40 Gbps is the highest data rate the PRBS source can operate at. The $2^{31} - 1$ output eyes are shown in Fig. 6.10. The eyes remain open even at 4 mV$_{pp}$ input amplitude and the noise floor is actually determined by the sensitivity of the oscilloscope. On the other hand, duty cycle distortion can be observed in the measured eye diagram at large signal amplitude due to the non-linear behavior of the circuit, as illustrated in Fig. 6.10(f) and 6.10(h).

The performance of the TIA is summarized in Table 6.1.
Chapter 6. Experimental Results

(a) 40 Gbps input eyes at 40 dB attenuation.
(b) 40 Gbps output eyes at 40 dB attenuation.
(c) 40 Gbps input eyes at 37 dB attenuation.
(d) 40 Gbps output eyes at 37 dB attenuation.
(e) 40 Gbps input eyes at 6 dB attenuation.
(f) 40 Gbps output eyes at 6 dB attenuation.
(g) 40 Gbps input eyes at 3 dB attenuation.
(h) 40 Gbps output eyes at 3 dB attenuation.

Fig. 6.10: $2^{31} - 1$ eye diagram measurement results.
Table 6.1: Summary of low-power TIA performance at room temperature.

<table>
<thead>
<tr>
<th></th>
<th>Simulated</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC gain [dB]</td>
<td>15.5</td>
<td>13.5</td>
</tr>
<tr>
<td>3dB bandwidth [GHz]</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Noise figure [dB]</td>
<td>&lt;5</td>
<td>&lt;7</td>
</tr>
<tr>
<td>Group Delay variation [ps]</td>
<td>±3</td>
<td>±3</td>
</tr>
<tr>
<td>$S_{11}$ [dB]</td>
<td>&lt;-10</td>
<td>&lt;-12</td>
</tr>
</tbody>
</table>

6.1.2 92 GHz Linear TIA

The photomicrograph of the manufactured amplifier is shown in Fig. 6.11. It has a pad-limited area of 0.138 mm² and an active area of just 0.023 mm², while consuming 21 mA from a 2.3 V supply when biased at its optimum NF current density. The characterization of this design was done with the same measurement setups demonstrated in section 3.2. The experimental results are provided in the figures below.

![Fig. 6.11: Fabricated linear TIA die micrograph.](image)

Fig. 6.12 illustrates the simulated and measured S-parameters and noise figure across all frequency bands. The amplifier exhibits 13 dB gain and a 3dB bandwidth of 92 GHz. The discontinuities present close to 70 GHz and 110 GHz are because of the change in the measurement setup from DC to W-band and W-band to D-band respectively. The measured group delay of the amplifier is shown in Fig. 6.13 and exhibits a variation of just ±1.5 ps up to 67 GHz. The group delay measurements at W-band are noisier, but follow closely the expected trend.

The measured 50 Ω noise figure is less than 6 dB up to 88 GHz, as can be observed in Fig. 6.12. To the author’s best knowledge, this is a record for silicon-based broadband amplifiers. The power gain and the measured noise figure at several frequencies are plotted in Fig. 6.14 as a function of the HBT collector current density per emitter length. As can be seen, the minimum noise figure current density ($J_{OPT}$) is 1 mA/µm and does not change with frequency, while the power gain is maximized at 1.5 mA/µm, which is also the peak $f_T/f_{MAX}$ current density of the HBT. However, there is practically no degradation of the power gain at the minimum noise figure current density.
Fig. 6.12: Comparison of the simulated and measured S-parameters and noise figure.

Linearity measurements were conducted with a Keysight N9030A PXA Signal Analyzer. For intermodulation tests HP83712B and HP83650B signal sources were used. Fig. 6.15 shows the measured total harmonic distortion and output power at different frequencies as a function of the input signal amplitude. The measured $P_{1dB}$ is -9 dBm mainly limited by the linearity of the output driver that has the same amount of 200 mV degeneration as the TIA stage. Again, good agreement can be observed between measurements and simulations. The measured $IIP_3$ is shown in Fig. 6.16 as a function of frequency and remains larger than +1 dBm up to 20 GHz.

Fig. 6.17 shows the eye diagram measurements that were carried out by varying the input signal over 36 dB using different attenuators. The eyes remain open even at 4 mV$_{pp}$ input amplitude and again the sensitivity of the oscilloscope is the limiting factor. To the author’s best knowledge, this kind of input sensitivity, about 6 dB better than that of a 40 Gbps InP TIALA [94] is a record for 50 Ω matched broadband amplifiers at 40 Gbps and confirms the low noise figure. The eyes start to look distorted at attenuation levels above 7 dB, which corresponds to an input signal amplitude of approximately 200 mV$_{pp}$. Lower duty cycle distortion is observed at the same input signal level than the TIA without degeneration (Fig. 6.10f and 6.10h). Further, the linear TIA was experimentally tested at higher data rates up to 120 Gbps as shown in Fig. 6.18 A SHF 12104A Bit Pattern Generator (BPG) provided two 60 Gbps data streams which were multiplexed up to 120 Gbps thanks to a SHF 603A 2:1 MUX. The data rate of the BPG was varied from 36-60 Gbps and the output amplitude of the multiplexer was set to 500 mV$_{pp}$. A block diagram and a real picture of the measurement setup are shown in Fig. 6.18 and Fig. 6.19 respectively, while the measurement results are plotted in Fig. 6.20.
Fig. 6.13: Measured group delay.

Fig. 6.14: Measured noise figure vs. current density at different frequencies.
Fig. 6.15: Measured THD and output power as a function of input power. The dashed line indicates $P_{1dB}$ at -9 dBm.

Fig. 6.16: Measured IIP3 as a function of frequency.
Fig. 6.17: 2\(^{31} - 1\) eye diagram measurement results for the linear TIA.
Fig. 6.18: Setup of the 120Gbps eye measurement setup.
Fig. 6.19: Picture of the 120Gb/s measurement setup.
Fig. 6.20: 72-120 Gbps eye diagram measurements for the linear TIA.
Finally the performance of the TIA is summarized in the following table.

Table 6.2: Summary of linear TIA performance at room temperature.

<table>
<thead>
<tr>
<th>Simulated</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC gain [dB]</td>
<td>13.5</td>
</tr>
<tr>
<td>3dB bandwidth [GHz]</td>
<td>92</td>
</tr>
<tr>
<td>Noise figure [dB]</td>
<td>&lt;6</td>
</tr>
<tr>
<td>Group Delay variation [ps]</td>
<td>±1.5</td>
</tr>
<tr>
<td>$S_{11}$ [dB]</td>
<td>&lt;10 (up to 35 GHz)</td>
</tr>
</tbody>
</table>

### 6.1.3 Comparison with State-of-the-Art

This section compares the both TIAs presented in this chapter with other broadband amplifiers reported in the literature. To draw a comprehensive comparison, designs from different processes such as CMOS, SiGe and III-V technologies have been included. At this juncture, it is important to note that, to the author’s best knowledge, this is the first broadband, low-noise amplifier that has been tested up to 120 Gb/s in any technology.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>10</td>
<td>102</td>
<td>4.425</td>
<td>2</td>
<td>73</td>
<td>±6</td>
<td>11.3$^1$</td>
<td>0.29</td>
<td>0.12µm SiGe</td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>60</td>
<td>14</td>
<td>1.2</td>
<td>12</td>
<td>-</td>
<td>8</td>
<td>-</td>
<td>45nm SOI CMOS</td>
</tr>
<tr>
<td>46</td>
<td>14.5</td>
<td>100</td>
<td>7.31$^2$</td>
<td>-2.5</td>
<td>145</td>
<td>±5.5</td>
<td>-</td>
<td>0.21</td>
<td>0.25µm InP DHBT</td>
</tr>
<tr>
<td>44</td>
<td>12</td>
<td>110$^3$</td>
<td>9.1</td>
<td>2.1</td>
<td>48</td>
<td>±1</td>
<td>8(sim)</td>
<td>0.197</td>
<td>90nm SiGe BiCMOS</td>
</tr>
<tr>
<td>This work</td>
<td>13</td>
<td>92</td>
<td>8.56</td>
<td>2.3</td>
<td>48</td>
<td>±1.5</td>
<td>6</td>
<td><strong>0.138</strong></td>
<td>55nm SiGe BiCMOS</td>
</tr>
<tr>
<td>This work</td>
<td>13.5</td>
<td>80</td>
<td><strong>35</strong></td>
<td>1.2</td>
<td><strong>10.8</strong></td>
<td>±3</td>
<td>7</td>
<td><strong>0.138</strong></td>
<td>55nm SiGe BiCMOS</td>
</tr>
</tbody>
</table>

$^1$ based on NF plot provided on paper

$^2$ based on differential gain of 20.5 dB

$^3$ requires external bias-T for operation

Table 6.3: Comparison with State-of-the-Art Broadband Amplifiers.
6.2 90 GS/s Quasi-CML Track & Hold Amplifier Breakout

The manufactured track-and-hold breakout is displayed in Fig. 6.21. It occupies an area of 940\(\mu\)m\(\times\)520\(\mu\)m and consumes a total of 86.65 mW from a 2.5 and a 1.8V supplies.

![Micrograph of the fabricated track and hold amplifier. Die area is: 940\(\mu\)m\(\times\)520\(\mu\)m.](image)

The track-mode behavior was assessed through S-parameter measurements up to 67 GHz with the setup depicted in Fig. 6.22. The circuit has a differential gain of +1 dB in track mode with a 3dB bandwidth of 40 GHz (from input pad to driver output pad), while S\(_{11}\) remains better than -15 dB up to 67 GHz. Its large signal operation during track-mode was verified through eye measurements at different rates and PRBS patterns as shown in Fig. 6.24.

Large-signal operation in sampling mode was measured with a 70GHz bandwidth sampling oscilloscope in time domain and with a spectrum analyzer in frequency domain up to 50 GHz using the testbench displayed in Fig. 6.25. A low phase-noise 67GHz signal source and a W-band multiplier were used to provide clock signals in the 40-110 GHz range. The 10MHz trigger output of the low frequency source was used to synchronize the sampling oscilloscope. Although the THA was verified to be sampling up to 108GHz clock signals, the performance is degraded above 92 GHz because of the limited bandwidth of the setup at the output of the circuit, which introduces a droop in the sampled output not seen at smaller sampling frequencies. Fig. 6.26 shows the measured differential output waveforms for single-ended sinusoidal inputs at different frequencies sampled at rates ranging from 40 to 108 GS/s, while Fig. 6.27 displays the single-ended outputs of an 18GHz signal sampled at 90 GS/s.
Fig. 6.22: S-parameter setup of THA breakout. Clock input is DC biased to fix the circuit in track-mode.

Fig. 6.28 compiles the measured IIP$_3$ at different sampling rates and the measured P$_{1dB}$ at 90 GS/s. The total harmonic distortion (THD) was measured at a low input power level where the third harmonic exceeds the integrated noise floor by 10 dB, and also at P$_{1dB}$. The integrated noise floor was estimated to be -60 dBm for a signal bandwidth of 40 GHz. As shown in Fig. 6.29, a THD of -44.4 dB and a SFDR of 50 dB were measured for a -16 dBm input sinusoid at 1.01 GHz sampled at 75 GS/s, which corresponds to the input power level where the third harmonic exceeds the integrated noise floor by 10 dB. For small-signal input sinusoids up to 15 GHz, the THD and SFDR with 40 GS/s and 90 GS/s sampling rates remain better than -33 dB and 40 dB, respectively. THD and SFDR degrade to -25 dB and 25 dB, respectively, under large-signal excitation at P$_{1dB}$, as illustrated in Fig. 6.30. On the same figure, the results of a beat frequency test at $f_{in} = f_s + \Delta f = 19$ GHz + 2 MHz are provided (due to the unavailability of a second signal source higher than 20 GHz, a sampling frequency of 19 GHz was used). The measured difference between the fundamental and the second and third harmonics are 44.2 dB and 54 dB, respectively. Some more spectra measured at the P$_{1dB}$ compression point are plotted in Fig. 6.31.
Fig. 6.23: Measured and simulated single-ended S-parameters in track mode. Vertical dashed line indicates the 3dB bandwidth.

Fig. 6.24: Measured single-ended PRBS $2^7 - 1$ output eye at 56 Gb/s in track-mode.
Chapter 6. Experimental Results

Fig. 6.25: THA breakout setup for time-domain and spectrum measurements.

Fig. 6.26: Measured differential output waveforms of a single-ended sinusoidal input at 8 GHz sampled at 40 GS/s (top left), 12 GHz sampled at 60 GS/s (top right), 18 GHz sampled at 90 GS/s (bottom left), and 12 GHz sampled at 108 GS/s.
Fig. 6.27: Measured single-ended outputs of an 18GHz signal sampled at 90 GS/s.

Fig. 6.28: Measured IIP₃ vs. input frequency for different sampling rates and measured P₁dB at 90 GS/s as a function of input frequency.
Fig. 6.29: Measured single-ended output spectrum for a 1.01GHz small signal input sampled at 75 GS/s (top) and SFDR and THD measured as a function of input frequency at 40 GS/s and 90 GS/s.
Fig. 6.30: Beat frequency test of a 19.002GHz sinusoid sampled at 19 GHz (top) and measured THD at 90 GS/s for input signal power at the 1dB compression point, $P_{1dB}$, and at the input level, where the third harmonic exceeds the noise floor by 10 dB (bottom).
Fig. 6.31: Measured single-ended spectra of 19.91GHz (top) and 39.9GHz (bottom) sinusoidal signals at the $P_{1dB}$ sampled at 90 GS/s.
6.2.1 Comparison with State-of-the-Art Track & Hold Amplifiers

A breakdown of the total power consumption of the circuit along with a comparison table with previously published high-speed THAs at different technologies are provided in table 6.4 and 6.5 respectively. To the best of the author’s knowledge, this is the first THA ever reported at 90 GS/s with a power consumption below 100 mW.

<table>
<thead>
<tr>
<th>Stage</th>
<th>$P_{DC}$ [mW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias</td>
<td>6.25</td>
</tr>
<tr>
<td>clock amplifier</td>
<td>36</td>
</tr>
<tr>
<td>50Ω output driver</td>
<td>14.4</td>
</tr>
<tr>
<td>input buffer</td>
<td>10</td>
</tr>
<tr>
<td>switched EF</td>
<td>20</td>
</tr>
<tr>
<td>total</td>
<td>86.65</td>
</tr>
</tbody>
</table>

Table 6.4: Breakdown of THA power consumption at room temperature.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>$f_s$ [GS/s]</th>
<th>track BW [GHz]</th>
<th>$I_{IP3}@f_{in}$ [dBm@GHz]</th>
<th>$THD@f_{in}$ [dB@GHz]</th>
<th>$P_{DC}$ [mW]</th>
<th>Supply [V]</th>
<th>Area [mm²]</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>50</td>
<td>27</td>
<td>20.7@6</td>
<td>-29.5@15</td>
<td>1200</td>
<td>-2.5, -5</td>
<td>0.73</td>
<td>250 nm InP DHBT $f_T/f_{MAX} = 370/650$ GHz</td>
</tr>
<tr>
<td>58</td>
<td>70</td>
<td>51</td>
<td>22@5</td>
<td>&lt;-52@2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>InP DHBT $f_T = 320$ GHz</td>
</tr>
<tr>
<td>57</td>
<td>32.5</td>
<td>19.3</td>
<td>N/A</td>
<td>-37@6</td>
<td>192</td>
<td>3</td>
<td>0.36</td>
<td>65 nm CMOS $f_T/f_{MAX} = 210/240$ GHz</td>
</tr>
<tr>
<td>54</td>
<td>40</td>
<td>43</td>
<td>0@19</td>
<td>-29@10</td>
<td>540</td>
<td>3.6</td>
<td>1.1</td>
<td>0.18 µm SiGe BiCMOS $f_T/f_{MAX} = 160$ GHz</td>
</tr>
<tr>
<td>56</td>
<td>30</td>
<td>N/A</td>
<td>19@1</td>
<td>&lt;-59@1</td>
<td>420↑</td>
<td>5.5</td>
<td>0.77</td>
<td>250 nm InP DHBT $f_T/f_{MAX} = 300$ GHz</td>
</tr>
<tr>
<td>This work</td>
<td>90</td>
<td>40</td>
<td>8@1</td>
<td>-49@1</td>
<td>87</td>
<td>2.5, 1.8</td>
<td>0.49</td>
<td>55 nm SiGe BiCMOS $f_T/f_{MAX} = 330/350$ GHz</td>
</tr>
</tbody>
</table>

↑: excludes output buffer

Table 6.5: Comparison With State-Of-The-Art Track & Hold Amplifiers.
6.3 ADC front-end Breakout

The micrograph of the fabricated ADC front-end test circuit is shown in Fig. 6.32 and consumes a total of 722 mW from a 2.5 V and a 1.8 V supplies.

The circuit was biased as shown in Fig. 6.33 in order to set both the master and slave THAs in track-mode. The GSG probe was connected initially to the input of the in-phase path, inI, and the transmission coefficient, $S_{21}$, was measured. Next, the probe was connected to inQ and the coefficient for the quadrature path was measured and found to be exactly equal to that of the in-phase path, proving the layout symmetry of the design. The measured S-parameters are plotted in Fig. 6.34 against the simulated results. A measured small-signal 3dB bandwidth of 16.5 GHz was achieved, which is almost 50% lower than the simulated value of 33.8 GHz. This large difference is likely due to inaccuracies in the model used for the long interconnect between the master and the active slave THA (see Fig. 5.6), and can be remedied by creating a more accurate EM model of the interconnect between the master
and slave THAs. Due to time constraints the circuit has not been fully characterized (time domain and spectrum measurements need to be done).

![Fig. 6.33: ADC front-end S-parameter setup.](image1)

![Fig. 6.34: Measured vs. simulated S-parameters of ADC front-end.](image2)
Chapter 7

Conclusion

7.1 Summary

This thesis presented the analysis and design of mm-wave broadband circuits capable of satisfying the stringent requirements of 100+GS/s ADCs in future fiberoptic receivers. Among the blocks of such receivers, the main focus of this work was on broadband amplifiers and sampling circuits preceding the ADC which require special attention due to their high demands in terms of bandwidth, noise, linearity, sampling frequency, and power consumption.

Based on a comparison study of possible low-noise broadband amplifier topologies found in the literature, the single-transistor shunt feedback (TIA) topology has yielded the best compromise between high-speed operation, low-noise, and low power consumption. The theory of noise in linear time invariant feedback systems has been revisited and a simple methodology for low-noise broadband shunt feedback amplifiers has been derived that results in the lowest possible power consumption taking also into account the input linearity requirements. Using this methodology two TIA circuit breakouts were designed, fabricated and measured: a low-power version suitable only for processing NRZ signals, and a linear version suitable for high performance fiberoptic receivers employing PAM-4 (or higher order) or DMT modulation. Unlike in current systems which use NRZ and QPSK signals, next generation ones have complex multi-level modulation formats and need linear low-noise broadband amplifiers in the receiver. The linear TIA circuit achieved 92GHz bandwidth with a maximum 50Ω NF of 6dB up to 88 GHz, which to the best of the author’s knowledge constitutes the lowest noise and highest data rate broadband amplifier in any technology.

Taking advantage of the superiority of the MOS-HBT cascode topology and eliminating the current source, a new Quasi-CML switched emitter follower topology has been proposed that supports GS/s operation at low power consumption. To verify the feasibility of this concept a track-and-hold amplifier (THA) was fabricated that demonstrated 90GS/s operation consuming 87mW from a 1.8V and 2.5V supplies, respectively. Combined with a tracking bandwidth of 40 GHz this design surpasses previously published THAs in InP technologies [51, 58]. This THA was integrated in an ADC sampling front-end system capable of 128GS/s operation.
7.2 Contributions and Publications

The original contributions of this thesis can be summarized in the following:

1. The highest data rate, lowest noise figure and highest sensitivity TIA in any technology and associated design methodology.

2. A record low-power and high-speed BiCMOS track-and-hold amplifier based on a new MOS-HBT quasi-CML switch with 40GHz tracking bandwidth, 90 GS/s rate and 87mW power consumption from 2.5V and 1.8V supplies.

The research conducted in this thesis has resulted in the following publications:


7.3 Future Work

The quasi-CML BiCMOS track-and-hold amplifier presented in this thesis can be incorporated in a full SAR ADC with 56-75GHz sampling clock and ×4 quadrature time-interleaving to enable sampling rates beyond 300 GS/s. Fig. 7.1 illustrates the concept. 25% non-overlapping sampling clocks are used to drive the THAs of each lane. This arrangement quadruples the sampling rate of the sub-ADCs used in the system. Fig. 7.2 shows the behavioral simulation results of a time-interleaved ADC with quadrature sampling clocks.

The fastest reported ADC to date is a 90GS/s time-interleaved SAR implemented in 32 nm CMOS technology by IBM [22]. To achieve this record sampling speed, it employs 64 interleaved lanes and has a maximum input bandwidth of 22 GHz. By exploiting the speed of the SiGe HBT in combination with the low-power of MOS devices in modern BiCMOS processes, a high bandwidth linear sampling front-end can be developed for driving multiple ADC lanes. In addition, the number of interleaved lanes can be reduced considerably by allowing each SAR ADC to work at much higher clock rates resulting in reduced power dissipation and improved input bandwidth. This can be achieved by switching from CMOS to low-power, high-speed quasi-CML implementations for the logic in every SAR lane. Quasi-CML has been proven to be a viable solution for high-speed data converters, as confirmed by a recent design of a 75GS/s 8bit DAC [66].

At the heart of the SAR logic illustrated in Fig. 7.3a lies the settable and resettable D flip-flop (SR D-FF) which makes up the successive approximation register. The code register block in Fig. 7.3b consist of SR D-FF as explained in [95]. In high-speed CML logic, it is important to ensure low-power operation while maintaining the highest possible switching speed. The SR D-FF consists of a master 1.8V quasi-CML latch similar to the one discussed in [66] and a slave 2.5V quasi-CML settable and resettable latch. A low-voltage, quasi-CML latch topology proposed in this thesis is displayed in Fig. 7.4. The use of stacked MOS devices for the CLK and reset (R) inputs ensures proper operation from just 2.5V and combined with inductive peaking it can operate with 75GHz clocks. The operation is as follows: when CLK is high, then the latch is transparent. When CLK is low, then there are three possibilities: if reset (R) and set (S) are low, the current value is latched in the latching pair (third from the left-hand side). If now S is high, while R is low, then the output is set to logic one, otherwise (R...
Fig. 7.1: \( \times 4 \) time-interleaved ADC system with 25\% duty cycle sampling clocks.

Fig. 7.2: Behavioral simulation results of the first two SAR ADC lanes.
is high), the latch is reset. Notice that both set and reset can only be activated on the complementary clock (\(\text{CLK}\)) and that reset has the highest priority. The outputs are buffered via two 0.5mA emitter followers that provide the appropriate DC-shift required by the DAC following the coder block.

Fig. 7.3: (a): Conceptual block diagram of a SAR ADC \[12\] and (b): successive approximation register \[95\].

Fig. 7.4: Schematic of the proposed SR D latch.
Bibliography


[69] V. Papanikolaou, “A Comparator and Track and Hold For Use In a 1 GS/s, 10 Bit Analog To Digital Converter,” Ph.D. dissertation, University of Toronto.


