Energy-Efficient, Wideband Transceiver Architectures and Circuits for High-Speed Communications and Interconnects

by

Jianyun Hu

Submitted in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

Supervised by
Professor Hui Wu

Department of Electrical and Computer Engineering
Arts, Sciences and Engineering
Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York
2012
Curriculum Vitae

Jianyun Hu was born in Shanghai, China on Dec. 1, 1981. He attended Fudan University, Shanghai, China from 1999 to 2006, where he received the Bachelor of Science degree in Electrical Engineering and the Master of Science degree in Microelectronics in 2003 and 2006, respectively. He came to the University of Rochester in the Fall of 2006 and began his graduate studies in the Department of Electrical and Computer Engineering. He pursued his research in RF and analog integrated circuit and system design under the direction of Professor Hui Wu and received the Master of Science degree from the University of Rochester in 2008. He was with Qualcomm Incorporated in the Summer and Fall of 2011, where he worked on circuit design and simulation for cellular RF IC.
Acknowledgments

First of all, I would like to express my deepest appreciation to my advisor, Professor Hui Wu, for his guidance, support, encouragement and suggestions throughout my time at the University of Rochester. During my PhD work, his enthusiasm for scientific research and technical innovations has always inspired me to keep moving forward. He has always enlightened and helped me with his keen insight and extensive experience in the area of integrated circuit and system design. I am deeply grateful to him for his kindness, patience and friendship over the years.

Sincere gratitude is extended to Professor Wendi Heinzelman, Professor Zeljko Ignjatovic and Professor John Marciante for their invaluable time to serve as my dissertation committee members, and for their help and guidance in my research and dissertation. Besides, I would like to thank Prof. Eby Friedman and Ioannis Savidis for their cooperation, helpful discussion and assistance in the free-space optical interconnect project. I would also like to thank Prof. Michael Huang and Dr. Aaron Carpenter for their cooperation, valuable discussion and assistance in the on-chip interconnect for multi-core processors project.

I would like to express my appreciation to the members of Professor Wu’s research group, Shang Wang, Jie Xu, Yong Wang, Jing Gao and Francis Smith, and to the former group members, Dr. Yunliang Zhu, Dr. Lin Zhang, Dr. Berkehan Ciftcioglu. I would also like to thank my friends in Rochester, Dr. Bo Fu, Dr. Yu Song, Dr. Qiaoyan Yu, Dr. Yang Zhang, Dr. Yijing Fu, Min Xu and Jie Zhang and many others.
for their friendship.

I would like to thank Bijoy Chatterjee, Ahmad Bahai, Peter Holloway, Mounir Bohsali, Johnny Yu, Anish Shah, Virginia Abellera, Peter Misich, and Jun Wan of National Semiconductor for their help and support in chip fabrication in the UWB project. I would also like to thank Fujitsu Component America, Inc. for providing the UWB antennas. I would like to thank Tezzaron Semiconductor for their help and support in chip fabrication in the free space optical interconnect project. I would like to thank MOSIS Research Program for their help and support in chip fabrication in the project of on-chip interconnect for multi-core processors.

Finally, I would like to give my special thanks to my parents, my wife and my daughter for their love and support.
Recently with the increasing demand for high-speed communications, wideband systems have become one of the major research focuses for both academia and industry. While wide bandwidth benefits high data-rate communication, compared to the conventional narrow bandwidth system, it poses large design challenges for both transceiver architectures and circuits, especially using the mainstream low cost CMOS and BiCMOS technologies. Besides, wideband systems typically inevitably require large power consumption, which might lead to worse energy-efficiency compared to the narrow-band systems. Therefore, in this thesis, we will focus on the energy-efficient, wideband transceiver architectures and circuits for high-speed communications and interconnects: ultra-wideband impulse radios (IR-UWB), intra-chip free-space optical interconnect, and on-chip electrical interconnect for multi-core processors.

Ultra-wideband communications has become an active research topic with the approval of UWB technology for commercial applications in the 3.1 - 10.6-GHz band by FCC. With such a large bandwidth, UWB technologies promise to offer low-power and high-speed wireless connectivity for future short-range communication systems. In this thesis, we will focus on the energy-efficient, wide-band UWB receiver architecture and circuits. We will first present a new UWB low-noise amplifier with noise cancelation, and use it to investigate the design trade-off for UWB amplifier. Then we will present a new analog correlation receiver architecture. It employs an
energy-efficient correlator called distributed pulse correlator (DPC) for low power ultra-wideband pulse detection. Thanks to the multiple pulsed multipliers time-interleaved in a distributed fashion and built-in local template pulse generation in the DPC, the power consumption and circuit complexity are significantly reduced for the DPC-based analog correlation receiver. The operation and performance of the DPC are analyzed, and the circuit implementation of DPC is discussed in details, especially the most critical component, the pulsed multiplier. A chip prototype of the DPC-based IR-UWB receiver was implemented in a 0.18-μm standard digital CMOS technology. In the measurement, the 8-tap, 10-GSample/s DPC achieves a pulse rate of 250 MHz with an energy efficiency of 40 pJ/pulse, and the whole receiver achieves an energy efficiency of 190 pJ/pulse at the 250-MHz pulse rate. Together with a UWB transmitter and two UWB antennas, the complete IR-UWB communication link is also demonstrated.

The continuous scaling of CMOS technology enables more and more modules to be implemented into a single chip. However, it actually poses challenges in the global interconnect design, especially with the rapid demand for higher-speed communication among more modules. Conventional electrical interconnect inevitably requires significant improvement for this high-speed on-chip global communication. In this thesis, we will investigate the high-speed global interconnect through both electrical and optical options.

Optical interconnects have been recognized as a promising successor to electrical interconnects. They have advantages like large bandwidth, low latency, and less susceptible to noise. We will present a novel optical transceiver architecture and circuits for the free-space optical interconnect for high-speed intra-chip communications. Compared to the conventional embedded-clock and forwarded-clock architectures, the presented shared-clock architecture benefits low power and low design complexity on the clock generation and recovery block and a simple interface between electrics and
optics. An injection-locked oscillator is employed to replace the conventional phase-locked loop as the clock generation block to further improve the energy-efficiency. Due to the high-speed and large bandwidth requirement, bandwidth extension techniques are widely used in the transceiver circuits. The optical transceiver was implemented in a 0.13-μm standard digital CMOS technology. The simulation results show that a 10-Gb/s data rate with 7.1-pJ/b energy-efficiency communication can be achieved.

For the electrical interconnect, we will present a novel on-chip interconnect system for multi-core chips using transmission lines as shared media in this thesis. It supports both point-to-point and broadcasting communications. Compared to network-on-chip approaches, it offers significant advantages in circuit complexity, energy efficiency and link latency. To demonstrate the scheme, a chip prototype with two 20-mm transmission lines running in parallel and multiple transmitters/receivers (including 2:1 serializer/1:2 deserializer) was implemented in a 130-nm SiGe BiCMOS technology. The transmission lines are designed with Ground-Signal-Signal-Ground configuration and patterned ground shields to exhibit low latency, small attenuation, generate less crosstalk, and provide high bandwidth density. The transceivers are designed and optimized to achieve good energy efficiency at the target data rate of 25 Gb/s. On the transmitter side, an efficient and low power pre-emphasis technique is applied to compensate for the transmission line’s frequency-dependent loss. On the receiver side, latched samplers are adopted for high sensitivity. To eliminate the insertion loss caused by a dedicated isolation switch, both the transmitter and receiver are designed to be internally switched in/out from the transmission lines. The prototype can successfully demonstrate point-to-point and broadcasting communications, and can achieve a data rate of 25.4 Gb/s with an energy efficiency of 1.67 pJ/b in the measurement.
Contents

Curriculum Vitae ii
Acknowledgments iii
Abstract v
List of Tables xii
List of Figures xiii
Foreword 1

1 Motivation 3
   1.1 High-Speed Communication Systems 3
   1.2 Ultra-Wideband Impulse-Radios (IR-UWB) 6
   1.3 Intra-Chip Free-Space Optical Interconnect (FSOI) 9
   1.4 On-Chip Interconnects for Multi-core processors 12
   1.5 Dissertation Organization 14

2 Wideband Circuit Techniques 17
   2.1 Challenges for Energy-Efficient Wideband Circuits 17
   2.2 Wideband Circuit Techniques 19
2.2.1 Inductive Peaking .......................................................... 19
2.2.2 Active Inductive Peaking .................................................. 22
2.2.3 Negative Impedance Compensation ..................................... 23
2.2.4 Negative Miller Capacitance .............................................. 25
2.2.5 \( f_T \) Doubler ............................................................... 26
2.2.6 Inverse Scaling ............................................................ 26
2.2.7 Capacitive/Resistive Degeneration ..................................... 28
2.2.8 Shunt-Series Amplifier ................................................... 30
2.2.9 Cherry-Hooper Amplifier ................................................ 31
2.2.10 Active Feedback .......................................................... 33
2.2.11 Distributed Circuits ...................................................... 34
2.2.12 Current Mode Logic Circuits ............................................ 36

3 Ultra-Wideband Impulse-Radio Receiver 38

3.1 IR-UWB Transceiver Architectures ........................................ 38
  3.1.1 Transmitter .............................................................. 39
  3.1.2 Non-Coherent Receiver ................................................ 40
  3.1.3 Direct Conversion Receiver .......................................... 41
  3.1.4 Analog Correlation Receiver ........................................ 42

3.2 Proposed IR-UWB Receiver Architecture .............................. 44

3.3 Wideband Low-Noise Amplifier (LNA) ................................. 45
  3.3.1 Topologies of Wideband LNA ...................................... 45
  3.3.2 A Wideband LNA Design Example .................................. 46

3.4 Distributed Pulse Correlator .............................................. 56
  3.4.1 Rationale ............................................................... 56
  3.4.2 Distributed Pulse Correlator Architecture ......................... 57
  3.4.3 Built-In Template Pulse Generation ................................. 59
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.4</td>
<td>Performance Analysis</td>
<td>61</td>
</tr>
<tr>
<td>3.5</td>
<td>IR-UWB Receiver Prototype Implementation</td>
<td>63</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Ultra-Wideband LNA</td>
<td>63</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Distributed Pulse Correlator</td>
<td>69</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Pulsed Multiplier</td>
<td>72</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Timing Control Circuit</td>
<td>76</td>
</tr>
<tr>
<td>3.5.5</td>
<td>Variable Gain Amplifier</td>
<td>77</td>
</tr>
<tr>
<td>3.5.6</td>
<td>Analog-to-Digital Converter</td>
<td>78</td>
</tr>
<tr>
<td>3.5.7</td>
<td>Measurement Results</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>Intra-Chip Free-Space Optical Interconnect Transceiver</td>
<td>88</td>
</tr>
<tr>
<td>4.1</td>
<td>Optical Transceiver Architectures</td>
<td>88</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Embedded-Clock Transceiver</td>
<td>88</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Forwarded-Clock Transceiver</td>
<td>90</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Shared-Clock Transceiver</td>
<td>91</td>
</tr>
<tr>
<td>4.2</td>
<td>Proposed FSOI transceiver Architecture</td>
<td>92</td>
</tr>
<tr>
<td>4.3</td>
<td>FSOI Transmitter Prototype Implementation</td>
<td>96</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Laser Driver</td>
<td>96</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Serializer</td>
<td>99</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Clock Generator</td>
<td>101</td>
</tr>
<tr>
<td>4.4</td>
<td>FSOI Receiver Prototype Implementation</td>
<td>104</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Transimpedance Amplifier</td>
<td>104</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Limiting Amplifier</td>
<td>107</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Decision Circuit</td>
<td>109</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Phase Selector</td>
<td>112</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Deserializer</td>
<td>112</td>
</tr>
<tr>
<td>4.5</td>
<td>Measurement and simulation Results</td>
<td>114</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Performance summary of wideband LNA</td>
<td>55</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparisons between pulsed multiplier and Gilbert cell</td>
<td>75</td>
</tr>
<tr>
<td>3.3</td>
<td>Summary of IR-UWB receiver performance</td>
<td>84</td>
</tr>
<tr>
<td>3.4</td>
<td>Performance comparisons with other IR-UWB receiver</td>
<td>85</td>
</tr>
<tr>
<td>5.1</td>
<td>Performance summary of the on-chip interconnect system</td>
<td>149</td>
</tr>
<tr>
<td>5.2</td>
<td>Performance comparisons with other on-chip interconnect systems</td>
<td>150</td>
</tr>
</tbody>
</table>
# List of Figures

1.1 Data Rate Trends. .................................................. 4
1.2 Wireless band allocation. .......................................... 5
1.3 FCC power emission mask. ....................................... 7
1.4 MB-OFDM UWB band allocation. ............................... 8
1.5 (a) The 3-D integrated chip stack of the overall FSOI system; (b) Top view of 4×4 core, where the VCSEL arrays are in the center and the photodetectors are on the periphery within each core. ........ 11
1.6 An on-chip interconnect system in a 64-core processor with 16 communication nodes (4 cores sharing a node). .................. 13

2.1 A single-stage common-source amplifier. ...................... 18
2.2 Inductive shunt peaking technique. ............................. 20
2.3 Inductive shunt and series peaking technique. ................. 21
2.4 Active inductive peaking technique (a) replacement of the inductor by a transistor and a resistor; (b) an example. ............... 23
2.5 Negative impedance compensation. ............................. 24
2.6 Negative Miller capacitance. ...................................... 25
2.7 (a) Simple differential stage; (b) $f_T$ doubler. .................. 27
2.8 Inverse Scaling Technique. ........................................ 27
2.9 Capacitive/resistive degeneration. .............................. 28
2.10 Shunt-series amplifier. ............................................ 30
2.11 (a) Simple two-stage cascaded amplifier (b) Cherry-Hooper amplifier;
(c) Differential Cherry-Hooper amplifier. ......................... 32
2.12 Active Feedback Technique. .................................... 34
2.13 Distributed topology. .............................................. 34
2.14 Current mode logic circuit topology. ......................... 36
3.1 Conventional IR-UWB transmitter architectures. ............. 39
3.2 Non-Coherent IR-UWB receiver architecture. .................. 40
3.3 Direct Conversion IR-UWB receiver architecture. ............ 41
3.4 Analog Correlation IR-UWB receiver architecture. ........... 42
3.5 System architectures of the IR-UWB receiver with the proposed DPC. 45
3.6 Possible wideband LNA topologies: (a) Distributed amplifier; (b) Common-gate amplifier; (c) Passive input matching; (d) Resistive feedback amplifier. ............................................ 46
3.7 Schematic of the wideband resistive-feedback noise-canceling LNA. . 47
3.8 Wideband noise cancelation. ..................................... 48
3.9 Simulated results of noise figure with and without noise cancelation. . 50
3.10 Power gain versus frequency with different $L_g$. ................ 51
3.11 Chip photo of the prototype wideband LNA. ................... 52
3.12 Measured power gain of the wideband LNA. ................... 52
3.13 Measured return loss and reverse isolation of the wideband LNA. ... 53
3.14 Measured and simulated noise figure of the wideband LNA. ....... 53
3.15 Measured IIP3 of the wideband LNA at 3GHz. .................. 54
3.16 Measured IIP3 of the wideband LNA at different frequencies. ...... 54
3.17 (a) Generic architecture of a distributed pulse correlator (DPC); (b) Waveform illustration of the DPC. ................................. 58
3.18 A DPC can be considered as a transversal finite impulse response filter.

3.19 Simulation results: (a) comparison of input pulse and template pulse; normalized SNR with regard to (b) template tap coefficient; (c) shape factor (related to pulse width); (d) timing drift or asynchronization; (e) individual tap timing error.

3.20 Schematic of the UWB LNA in the receiver prototype.

3.21 Power gain versus frequency with different $L_d$.

3.22 Measured and simulated power gain of the UWB LNA.

3.23 Measured return loss and reverse isolation of the UWB LNA.

3.24 Measured and simulated noise figure of the UWB LNA.

3.25 Measured IIP3 of the UWB LNA at 6 GHz.

3.26 Measured group delay of the UWB LNA.

3.27 Schematic of the DPC in the receiver prototype.

3.28 Simulation results of the transmission line: (a) attenuation with different number of taps; (b) delay from each tap to the succeeding taps.

3.29 Schematic of (a) a Gilbert cell, (b) the pulsed multiplier in a DPC (simplified schematic), and (c) a switched transconductor mixer.

3.30 Schematic of the pulsed multiplier with built-in impulse generator.

3.31 Schematic of the timing control block in the receiver prototype.

3.32 Schematic of the VGA in the receiver prototype.

3.33 Block Diagram of the ADC in the receiver prototype.

3.34 Chip micrograph of the prototype IR-UWB receiver.

3.35 Test setup for the prototype receiver characterization.

3.36 Measured pulse response of the LNA.

3.37 Measured and simulated normalized coefficient with different bias current.
3.38 Measured DPC outputs with (a) correct timing and templates; (b) different timing offset; (c) different templates. ............... 83
3.39 Measurement results of the receiver with modulated data. .......... 84
3.40 Measurement setup for the UWB impulse radio test. ............... 85
3.41 Measurement results of the transceiver with different pulse rate and different transmitted waveforms. ......................... 87
3.42 Measurement results of the transceiver with modulated data. ...... 87
4.1 Conventional optical communication system with embedded clock. . 89
4.2 Optical communication system with forwarded clock. ............... 91
4.3 (a) Optical communication system with shared clock; (b) Implementation of a shared-clock optical transceiver. ................. 93
4.4 An FOSI system with shared clock. .......................... 94
4.5 An FOSI system with ILO-Based clock generation and shared clock. 95
4.6 Laser Driver. ............................................. 96
4.7 Laser driver circuit schematics (a) pre-amplifier; (b) main amplifier; (c) pre-emphasis. ............................................. 97
4.8 Simulation results of the laser driver with (a) pre-emphasis off; (b) pre-emphasis on. ............................................. 98
4.9 (a) 1:64 serializer; (b) 2:1 multiplexer cell. ........................ 100
4.10 (a) Conventional CML-based latch; (b) CML-based selector; (c) CMOS digital latch. ............................................. 101
4.11 Phase-Locked Loop. ......................................... 102
4.12 (a) ILO topology; (b) Delay cell; (c) Ring oscillator core. ........ 103
4.13 Transimpedance amplifier. ................................... 105
4.14 Simulated transimpedance gain of the active transimpedance amplifier. 106
4.15 Limiting amplifier topology. .................................. 107
4.16 Limiting amplifier: (a) input stage; (b) gain stage. ............... 108
4.17 Simulated voltage gain of the limiting amplifier. ............... 109
4.18 (a) DFF-based decision circuit; (b) Modified CML-latch used in the
decision circuit (where the tail transistor in conventional CML-latch is
removed). ......................................................... 110
4.19 Simulated results of the decision circuit: (a) Time-domain waveforms;
(b) Input sensitivity of the decision circuit. ......................... 111
4.20 (a) 1:64 deserializer; (b) 1:2 demultiplexer cell. .................... 113
4.21 Simulated results of the serializer and deserializer. ............... 115
4.22 (a) Chip photo of the optical transceivers; (b) Layout of optical transceivers.
116
4.23 Simulated results of (a) transmitter; (b) Receiver. ................. 117
4.24 Measurement and simulation results of: (a) Laser driver; (b) Receiver
front-end (passive TIA and limiting amplifier); (c) Receiver front-end
(active TIA and limiting amplifier); ........................................ 119
4.25 (a) Back metal illustration (not to scale, dielectric not shown); (b)
Simulation results of effect on bandwidth of on-chip transmission lines,
TSV and back metal. ......................................................... 121
4.26 An FSOI system with latched sampler. ............................... 122
4.27 A 4-channel FSOI system with shared ILO-Based clock generation and
phase recovery. ......................................................... 123

5.1 Different transmission-line-based on-chip interconnect transceiver im-
plementation: (a) Carrier-based; (b) Carrier-based with multiple ac-
cess technique (FDMA is used for illustration); (c) Direct transmission. 125
5.2 Interconnect prototype architecture. ................................. 127
5.3 Parallel-running differential coplanar waveguides with a Ground-Signal-Signal-Ground (GSSG) configuration. ........................................... 128
5.4 Simulated transmission line attenuation with different lengths. .... 129
5.5 Architecture of the transmitter for the on-chip interconnect system. 129
5.6 (a) Topology of the transmitter multiplexer; (b) Schematic of the selector; (c) Schematic of the latch. ........................................... 130
5.7 (a) Schematic of the transmitter driver; (b) Illustration of pre-emphasis scheme. ........................................... 132
5.8 Architecture of the receiver for the on-chip interconnect system. ... 132
5.9 Schematic of the pre-amplifier in receiver. ............................... 133
5.10 Simulated input sensitivity of the latch. ................................. 134
5.11 Simulated insertion loss and isolation of an isolation switch. ....... 135
5.12 Simulated isolation performance of the driver. ......................... 136
5.13 Simulated isolation performance of the pre-amplifier. ............... 136
5.14 Transmission line with multiple communication nodes. ............. 137
5.15 Chip photo of the interconnect prototype. .............................. 139
5.16 Measured and simulated voltage gain: (a) Transmitter driver; (b) Receiver pre-amplifier. ........................................... 140
5.17 Test setup for the prototype chip characterization. .................... 141
5.18 Measured and simulated outputs of Rx5 pre-amplifier at the data rate of 25.4 Gb/s. ........................................... 142
5.19 Measured outputs of Rx6 latches with simulated inputs of Tx1 serializer at the data rate of 25.4 Gb/s. ........................................... 143
5.20 Measured eye diagrams at the Rx6 outputs. ............................ 143
5.21 Measured eye-diagrams at amplifier output of Rx5 with pre-emphasis control. Pre-emphasis off: (a) 10.16 Gb/s; (c) 15.24 Gb/s; (e) 20.32 Gb/s; (g) 25.4 Gb/s. Pre-emphasis on: (b) 10.16 Gb/s; (d) 15.24 Gb/s; (f) 20.32 Gb/s; (h) 25.4 Gb/s. .................................................. 145

5.22 Measured eye-diagrams at amplifier output of Rx2 with pre-emphasis control. Pre-emphasis off: (a) 10.16 Gb/s; (c) 15.24 Gb/s; (e) 20.32 Gb/s; (g) 25.4 Gb/s. Pre-emphasis on: (b) 10.16 Gb/s; (d) 15.24 Gb/s; (f) 20.32 Gb/s; (h) 25.4 Gb/s. .................................................. 146

5.23 Measured broadcasting communication of Tx1 with different number of receivers turning on: (a) Rx2 output; (b) Rx3 output; (c) Rx5 output; (d) Rx6 output. .................................................. 147

5.24 Measured crosstalk performance for the multi-channel interconnect system. .................................................. 148

5.25 Measured results of (a) Tx1 to Rx5 with termination switch on @ 20.32 Gb/s; (b) Tx1 to Rx5 with termination switch off @ 20.32 Gb/s. (c) Tx4 to Rx5 with termination switch on @ 12.7 Gb/s; (d) Tx4 to Rx5 with termination switch off @ 12.7 Gb/s. .................................................. 150

6.1 Schematic of a receiver with 8 latched samplers. .................. 152

6.2 Simulated on-chip interconnect system at 40 Gb/s. .................. 153
Foreword

This thesis discusses energy-efficient, wideband transceiver architectures and circuits for high-speed communications and interconnects: ultra-wideband impulse-radio, intra-chip free-space optical interconnect, and on-chip electrical interconnect for multi-core processors. I performed all the analysis, implementation and measurement for all the transceiver architectures and circuits for all three examples in this thesis unless specified below.

In Chapter 3, I present a new analog correlation IR-UWB receiver. I collaborated with Yunliang Zhu and Shang Wang from Prof. Wu’s group in this work. I performed all the analysis, implementation and measurement of the receiver architecture and circuits, except that Yunliang Zhu contributed to the implementation of the correlator and Shang Wang contributed to the implementation and measurement of the delay-locked loop. The IR-UWB transceiver architecture analysis has been published in IEEE Upstate New York Workshop on Communications, Sensors and Networking in 2007. The IR-UWB receiver has been published in IEEE Radio Frequency Integrated Circuit (RFIC) Symposium in 2009. The IR-UWB transceiver demonstration has been published in IEEE International Conference on Ultra-wideband (ICUWB) in 2009. I am the first author of all three papers.

In Chapter 3 Section 3, I present a new ultra-wideband low-noise amplifier with noise cancelation. I performed all the analysis, implementation and measurement of the low-noise amplifier except that Yunliang Zhu helped to perform the measurement
on noise figure. The work has been published in *8th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems* in 2008, and I am the first author.

In Chapter 4, I present transceiver architectures and circuits for the intra-chip free-space optical interconnect. I collaborated with Berkehan Ciftcioglu and Jie Xu from Prof. Wu’s group. I performed all the analysis, implementation and measurement of the optical transceiver architectures and circuits, except that Berkehan Ciftcioglu contributed to the implementation of the injection-locked oscillator, Jie Xu contributed to the implementation of the PRBS generator for test purpose. Part of the work on the optical transceiver and circuits has been published in *37th International Symposium on Computer Architecture (ISCA)* in 2010, *IEEE Photonic Technology Letters* in 2011, *SPIE Photonics West* in 2012, and *Optics Express* in 2012. I am one of the co-authors of all these papers.

In Chapter 5, I present a new transceiver architecture and circuits for an on-chip electrical interconnect system for multi-core processors. I collaborated with Jie Xu from Prof. Wu’s group. I performed all the analysis, implementation and measurement of the interconnect system, transceiver architecture and circuits, except that Jie Xu contributed to the implementation of the transmission line and the PRBS generator (for test purposes), interconnect system layout and measurement. The work has been submitted to *IEEE Custom Integrated Circuits Conference (CICC)* in 2012, and I am the first author of the paper. I also collaborated with Aaron Carpenter from Prof. Michael Huang’s group on the on-chip interconnect system architecture. This part of work has been published in *38th International Symposium on Computer Architecture (ISCA)* in 2011, *International Symposium on Low Power Electronics and Design (ISLPED)* in 2011, *39th International Symposium on Computer Architecture (ISCA)* in 2012, and *IEEE Journal for Emerging and Selected Topics in Circuits and Systems (JETCAS)* in 2012. I am the second author of all four papers.
Chapter 1

Motivation

1.1 High-Speed Communication Systems

The last few years have witnessed a rapid increase in demand for high-speed communications. As shown in Fig. 1.1, which shows the data rate trends of several wireline, wireless and optical technologies, all these technologies go through great increments in data rate. With the higher data rate, more data can be transferred within a certain amount of time, or a certain amount of data can be transferred within less time. This advantage enables many useful features, such as sharing files within much less time, and browsing the internet and streaming video using cell phones.

According to the Shannon theorem [1], the channel capacity is proportional to the bandwidth of the channel with a given signal-to-noise (SNR) ratio, therefore a large channel bandwidth typically has the benefit of a high data rate. For example, in order to achieve a high data rate for short-range wireless communications, wideband radios, such as ultra-wideband (UWB) radio and 60 GHz millimeter-wave (mm-wave) radio, have emerged recently. As shown in Fig. 1.2, UWB radio has a bandwidth of 7.5 GHz from 3.1 to 10.6 GHz, and 60 GHz mm-wave radio has a bandwidth of 7 GHz
from 57 to 64 GHz. Compared to narrowband radios, the large bandwidth of these two radios provides the possibility of a high communication data rate of hundreds of Mb/s to several Gb/s [2–4].

The large bandwidth associated with the high data rate, however, poses difficulties on the design of system architectures and front-end circuits. For example, the UWB system mentioned above, which occupies the bandwidth of 7.5 GHz from 3.1 to 10.6 GHz, requires the architecture and front-end circuits to have the capability to handle such a large bandwidth. Besides, for a high-speed optical communication system operating at a data rate of 10 Gb/s, a minimum bandwidth of 7 GHz is required for the front-end circuits [5]. These large bandwidth requirements inevitably face more design challenges than a narrowband system, especially with the consideration on other important design parameters, such as gain, noise, supply voltage, current consumption, linearity and input/output impedance. All these challenges bring the demand of investigation on wideband system architecture and circuit techniques.

As mentioned above, power consumption is an important design consideration. However, the circuit techniques for bandwidth extension inevitably require more
power consumption. Energy efficiency is often used to characterize the power consumption with a given data rate for a high-speed system. It is a very important parameter for high-speed communications. The high energy efficiency can minimize energy use, leading to a longer operation time for mobile devices using only batteries. Besides, an energy-efficient system can also help to reduce the heat dissipation and associated cooling and packaging costs. High energy efficiency is especially important for recently emerged multi-channel or parallel communications, since in these systems, the energy use can be easily boosted. Therefore, in order to achieve a high-speed communication system with high energy efficiency, we must make efforts on both system architecture and circuits.

Wideband, high-frequency, high-speed integrated circuits have been mainly occupied by Silicon bipolar and BiCMOS and III/V compound semiconductor technologies. Compared to these technologies, CMOS technology, widely used in digital circuits, has the disadvantages of lower cut-off frequency and relatively large parasitic capacitances. However, the low-cost and high-yield characteristics of CMOS technology make it an attractive candidate. Additionally, CMOS technology also brings the possibility of a highly integrated system-on-chip including high-speed front-ends, analog-to-digital and digital-to-analog data converters, and digital baseband processors, which further reduce chip and package costs. Thanks to the continual scaling of
CMOS technology, much higher cut-off frequency can be achieved in CMOS technology nowadays. Combined with effective wideband techniques, such as the widely used inductive peaking technique [6], more and more researches based on CMOS technology have been emerging recently in the field of wideband, high-frequency, high-speed integrated circuits [7–11]. BiCMOS technology, mentioned above as an option for conventional wideband, high-frequency, high-speed circuits, is still a good candidate for these applications, especially with improvements in low cost and high yield. Therefore, in this thesis, CMOS and BiCMOS technologies will both be our options to explore the wideband communication systems.

The large bandwidth required for high-speed communication, together with the requirements on energy efficiency and low-cost technologies like CMOS/BiCMOS, inevitably bring challenges to the design of high-speed communication systems. In this work, we will focus on the high-speed wireless communication and on-chip interconnects, where ultra-wideband impulse radios (IR-UWB), intra-chip free-space optical interconnect (FSOI) and on-chip electrical interconnect for multi-core processors are used to investigate energy-efficient, wideband transceiver (i.e., transmitter and receiver) architectures and circuits in CMOS/BiCMOS technology for high-speed communication systems.

1.2 Ultra-Wideband Impulse-Radios (IR-UWB)

Ultra-wideband (UWB) communication is emerging as the next-generation short-range wireless technology, particularly for wireless personal area networks [12]. As shown in Fig. 1.3, UWB communications exhibit two unique characteristics compared to the conventional narrow-band wireless system: (a) large available bandwidth (3.1 - 10.6 GHz in U.S.), \(^1\) and (b) low transmit power (peak power -41.3 dBm/MHz

\(^1\)Officially, UWB is defined as any wireless transmission scheme occupying an absolute bandwidth of more than 500 MHz, or a fractional bandwidth larger than 20%. 
outdoors in U.S.) [13], which is required to mitigate interference with the co-existence with existing wireless system across this wide frequency range. The large bandwidth of UWB not only enables high data rates, but also provides the flexibility in radio implementation.

Currently, there are two major categories of UWB systems. Multi-band orthogonal frequency-division multiplexing (OFDM) approaches (MB-OFDM UWB) divide the large UWB bandwidth into multiple sub-bands (i.e., Group 1-5). As is shown in Fig. 1.4, Groups 1-4 consist of three sub-bands of 528 MHz, and Group 5 consists of 2 sub-bands of 528 MHz. Due to the coexistence with 802.11a system which operates between 5 and 6 GHz, MB-OFDM UWB usually avoids the strong interference by not using Group 2. Since each sub-band has a bandwidth of only 528 MHz, multi-band system can operate within each sub-band as a conventional narrow-band radio with advanced modulation schemes, e.g., orthogonal frequency division multiplexing (OFDM) and quadrature phase shift keying [14]. However, complex digital baseband is required in the radio implementation to deal with these advanced modulation schemes. Further, a complex wideband frequency synthesizer with multi-GHz
tuning range is needed to generate the local carriers that cover the multiple sub-bands \[15,16\]. Both requirements will result in significant increases in hardware cost and power consumption, while not truly utilizing the large UWB bandwidth.

![MB-OFDM UWB band allocation.](image)

On the other hand, by using short-duration pulses, impulse radio UWB (IR-UWB) utilizes a much larger portion of the available UWB bandwidth, without dividing it into sub-bands \[17\]. Signals are directly modulated on the ultrashort UWB pulses, and no carrier is required, which illustrates the truly wideband nature of IR-UWB. Relatively simple modulation schemes are employed such as pulse-position modulation (PPM), on-off keying (OOK), and binary phase-shift keying (BPSK). Compared to the carrier-based MB-OFDM UWB, carrier-less IR-UWB requires no up/down frequency conversion or complex frequency synthesizer, which leads to much simpler transceiver architectures and lower power consumption. Further, the sub-nanosecond UWB pulses yield good timing resolution, and hence IR-UWB can potentially provide ranging and locating capabilities. To solve the issue of interference located between 5 and 6 GHz, IR-UWB can utilize a notch filter between the antenna and a low noise amplifier (LNA), or simply use an antenna with a notch filter or an LNA with a notch filter.

Due to the large bandwidth and stringent requirements on transmit power spectrum, however, it remains challenging to design the analog front-end circuitry that
can efficiently generate, modulate, and detect the ultrashort IR-UWB pulses using mainstream digital CMOS technologies. CMOS compatibility is important because of the cost advantage and system-on-chip integration considerations. Further, it is highly desirable for IR-UWB systems to be reconfigurable in order to (a) optimize the system performance on demand based on channel environments, (b) accommodate process, voltage and temperature (PVT) variations, and (c) satisfy different regulatory requirements.

Our research goal is to develop energy-efficient, high-speed and reconfigurable UWB impulse radios using mainstream digital CMOS technologies. Recently, we have demonstrated a low-power, reconfigurable IR-UWB transmitter based on a digitally-assisted distributed circuit technique called distributed waveform generator (DWG) [18–20]. We also implemented an energy-efficient reconfigurable pulse detection technique distributed pulse correlator (DPC) [21]. In this work, we will consummate the work in [21]. We will have a detailed analysis on both the architecture and circuit of the DPC, including the core building block of the DPC (i.e. pulsed multiplier), and then propose an energy-efficient, high-speed and reconfigurable IR-UWB receiver based on the DPC [22,23].

1.3 Intra-Chip Free-Space Optical Interconnect (FSOI)

Thanks to the continual scaling of CMOS technology, systems are becoming more and more compact. Alternatively, within the same chip area, scaling makes it possible to build more complex systems with better performance and more functions (e.g., multi-core microprocessors). While scaling benefits better transistor performance, it actually poses challenges in global interconnect design, especially for complex systems with multiple modules.
First, with technology scaling, most local interconnects shrink. Global interconnects, however, still remain long length, especially when connecting two modules far away from each other. Long-distance interconnects may degrade system performance due to the limitation of electrical interconnects on attenuation, bandwidth and latency. Additional techniques, such as equalization or adding repeaters, can be applied to mitigate these issues. These techniques, however, inevitably consume more power and increase system complexity. Secondly, most modules communicate between each other, this makes the inter-module interconnect arrangement difficult, especially with the increasing number of module. Thirdly, electrical interconnects are susceptible to environment noise, such as substrate noise or coupling from neighboring interconnects. This issue becomes even worse when the density of the interconnect increase. In this thesis, we will investigate these issues with two solutions, one based on electrical interconnects, the other based on optical interconnects. We will first discuss the latter option.

Optical interconnects have been recognized as the promising successor to electrical interconnects. They have several main advantages over the electrical counterpart, such as larger bandwidth, lower latency and less susceptible to noise [24, 25]. Researches have demonstrated Gb/s optical interconnects for chip-to-chip and board-to-board communications [26–29]. In [30], a free-space optical interconnect (FSOI) was proposed for high-speed intra-chip communication.

As illustrated in Fig. 1.5-(a) [30, 31], the FSOI system is based on a 3-D fashion. The bottom layer is the electronic part of the FSOI system, where the I/O transceivers are implemented in this CMOS electronic layer. Note that the multicore microprocessor is also built in this layer and on the same chip with the I/O transceivers, which benefits for high integration and low cost. The middle layer is the GaAs photonics layer, where all the electro-optic (E/O) and opto-electronic (O/E) conversions happen. It includes vertical-cavity surface-emitting lasers (VCSELs) [32]
and photodetectors (PDs), respectively converting signals from electrical form to optical form and the opposite operation. The top layer is the optical part of the FSOI system, and there are micro-lenses and micro-mirrors, focusing and guiding the light beam emitted by the VCSELs and detected by the PDs.
Due to the advantages of optical interconnect in bandwidth, latency and noise, FSOI systems overcome the drawback of electrical links for on-chip global interconnects, and benefits large bandwidth, low power consumption and minimal system complexity. As shown in the topview in Fig. 1.5-(b) [30], the FSOI also provides an all-to-all direct communication link between each processor core, and easily scales with the increasing number of core, which overcomes the issue of electrical interconnects when arranging a large amount of inter-core interconnects.

All three FSOI components, I/O transceiver in the electronics layer, E/O and O/E converters in the GaAs photonics layer, and micro-devices in the free-space optical layer, are critical to system performance. In this thesis, we will focus on the I/O transceiver in the electronics layer. As mentioned earlier, high-speed optical I/Os have been under active research recently. However, most of them target chip-to-chip and board-to-board interconnects. For the high-speed intra-chip multi-core interconnects, since the I/O transceiver is required in each core, it has stringent requirements on power consumption and chip area to mitigate the system overhead on power and cost. Therefore energy/area-efficient system architecture and circuits are demanded for this purpose. In this thesis, we will investigate conventional optical/serial-link system architectures, and then develop the energy/area-efficient I/O transceiver architecture and circuits for this high-speed FSOI system.

1.4 On-Chip Interconnects for Multi-core processors

In the previous section, we discussed the optical solution, and here we will focus on the electrical option. As mentioned above, the continued technology scaling enables systems with higher performance, more functionality and more complexity
to be integrated into a single chip (e.g. multi-core processors). However, with the number of cores increases, the global interconnects in a multi-core processor becomes increasingly critical to its performance and energy efficiency. Network-on-chip (NoC) has been proposed to replace conventional buses thanks to its advantages in bandwidth scalability and modularity. However, packet switching in NoC requires routers that consist of complex circuitry, occupy large chip area and consume significant power [33,34]. In addition, repeated packet relaying adds latency to communication, and degrades the performance significantly. Transmission lines were proposed for point-to-point on-chip communications previously to achieve wide bandwidth and low latency [35,36], and later extended to multipoint-to-multipoint communications [37].

In this work, we will explore a transmission-line based on-chip interconnect system for future multi-core processors on the architecture, system and circuit levels. This packet-switching-free interconnect can provide large bandwidth, high throughput and extremely low latency over a long distance.

![Figure 1.6: An on-chip interconnect system in a 64-core processor with 16 communication nodes (4 cores sharing a node).](image-url)

In the proposed interconnect system, as shown in Fig. 1.6, multiple transmission
lines run in parallel across all the cores, and serve as the shared media for both point-to-point communications and broadcasting. Each transmission line is connected to a transmitter and a receiver at each node without any router or switches, and properly terminated at both ends. Isolation switches, which are similar to T/R switches in wireless transceivers [38,39], are used between the transmitter/receiver and the transmission line to isolate all but the communicating nodes from the transmission line. Therefore, even though all nodes are connected to the transmission line, the line can still operate as a point-to-point media. At the same time, the transmission line can be easily configured into broadcasting media by switching on multiple communicating modes, which enables a transmitter to transmit data to multiple receivers.

Throughput comparable to NoC can be achieved because of its capability of operating at higher clock speed and by taking advantage of the low latency more efficiently on the architectural level [33,34]. For example, the whole interconnect system can achieve a total throughput of 1 Tb/s by using 25 transmission lines, each operating at 40 Gb/s. Such a transmission line link without repeaters eliminates skew and jitter, saves power consumption, and reduces the link latency significantly. With proper transmission line and transceiver design, a sub-nanosecond cross-chip latency can be achieved. For instance, with a low latency of 6 ps/mm, the maximum latency is less than 600 ps between two end nodes of the transmission lines (Node 1 and 16) in a 2.5 cm × 2.5 cm, 64-core chip (Fig. 1.6). That translates into 3 computing cycles for a 5-GHz processor, half of the time for a packet-switching mesh network.

1.5 Dissertation Organization

In this thesis, energy-efficient wideband transceiver architectures and circuits are developed for high-speed communications and interconnects. We will focus on the analysis and implementation of the ultra-wideband impulse-radio system, intra-chip
free-space optical interconnect system, and on-chip interconnect system for multi-core processors.

In Chapter 2, we will first discuss the fundamental bandwidth limitation factors, and then review major wideband circuit techniques, including inductive peaking, negative impedance compensation, negative Miller capacitance, $f_T$ doublers, capacitive/resistive degeneration, Cherry-Hooper amplifier, active feedback, inverse scaling, and distributed circuits. In addition, we will review the current mode logic circuit topology, which is widely used in high-speed circuits.

In Chapter 3, we will review IR-UWB transceiver architectures. After a brief review of the transmitter architectures, we will focus on the review of the receiver architectures, including non-coherent, direct conversion, and analog coherent architectures. The analysis and design of the IR-UWB receiver architecture and circuits will be presented. An energy-efficient IR-UWB receiver architecture will be proposed, and we will focus on the analysis of two important building blocks: wideband low-noise amplifier and distributed pulse correlator. After that, we will present the implementation of the circuits, and conclude the IR-UWB receiver implementation with measurement results.

In Chapter 4, we will focus on the optical interconnect system. We will first review conventional optical transceiver architectures, where we will review the embedded-clock architecture, which is widely used in optical systems. After that, two relatively new architectures will be reviewed: forwarded-clock and shared-clock architectures. Then the design of FSOI transceiver architecture and circuits will be described. After proposing energy-efficient FSOI transceiver architectures, we will discuss the implementation of the FSOI circuits, including laser driver, transimpedance amplifier, limiting amplifier, decision circuit, serializer and deserializer. The simulation and measurement results of the circuits will be presented.

In Chapter 5, the on-chip interconnect system for multi-core processors will be
presented. After reviewing the existing on-chip interconnect systems, we will propose a new on-chip interconnect topology, featuring high-throughput and low-latency and supporting both point-to-point and broadcasting communications. Then we will present the analysis and implementation of the interconnect system, including the system architecture and the circuits. After that, we will present measurement results of the on-chip interconnect system.

Finally, Chapter 6 will be the discussions and conclusions.
Chapter 2

Wideband Circuit Techniques

2.1 Challenges for Energy-Efficient Wideband Circuits

Before reviewing each wideband technique, let us first examine the challenges for energy-efficient wideband circuits. As illustrated in Fig. 2.1, a single-stage common-source amplifier is used for this examination, where $C_L$ is the total load capacitances including output parasitic capacitance of $M_1$ and the input capacitances of the subsequent stage, $R_d$ is the load resistor. Intuitively speaking, due to the decrease of the impedance of the capacitance $C_L$ with the increasing frequency, the gain $\left( g_{m1}(R_d/(1/sC_L)) \right)$ rolls-off at high frequencies and hence the bandwidth is limited. To be more specific, we can derive the input and output poles of the amplifier as follows

$$\omega_{in} = \frac{1}{R_s(C_{gs} + (1 + g_{m1}R_d)C_{gd})} \quad (2.1)$$
\[ \omega_{out} = \frac{1}{R_d C_L} \] (2.2)

where \( R_s \) is the source resistor (not shown in the plot). The term \((1 + g_m R_d) C_{gd}\) in \( \omega_{in} \) is caused by Miller effect. As we can see, \( C_{gd} \) is approximately amplified by the gain of the amplifier, which leads to a much larger value than \( C_{gd} \) itself. In a typical amplifier, \( C_L \) is much larger than other capacitances, and load resistor \( R_d \) is larger than the source resistor \( R_s \), therefore the output pole \( \omega_{out} \) is normally much smaller than the input pole \( \omega_{in} \) and hence is the dominant factor limiting the bandwidth of the amplifier. The input pole \( \omega_{in} \) is then the second dominant factor. Depending on its value, it may also affect the bandwidth in different extent.

Different wideband techniques targeting load capacitance and input capacitance have been developed to achieve a large bandwidth. In addition, resistor, even transconductance of the transistor, can also be potentially improved for a large bandwidth. However, bandwidth extension techniques often inevitably require extra current consumption. We will review these wideband techniques and discuss the trade-off between bandwidth extension and lower power consumption.

![Figure 2.1: A single-stage common-source amplifier.](image)
2.2 Wideband Circuit Techniques

Wideband circuit techniques have been extensively used to extend the circuit bandwidth. In this Section, we will review the major wideband circuit techniques, including inductive peaking, active inductive peaking, negative impedance compensation, negative Miller capacitance, $f_T$ doubler, inverse scaling, capacitive/resistive degeneration, Cherry-Hooper amplifier, active feedback, and distributed circuits. In the end, current mode logic circuit topology, widely used in high-speed circuits operating in large-signal mode, will also be reviewed.

2.2.1 Inductive Peaking

Inductive Shunt Peaking

As discussed earlier, the impedance of the load capacitance decreases with frequency, making the gain of the amplifier roll-off at high frequencies. Therefore a straightforward approach to resolve this issue is to add an inductor in shunt with the load capacitance, where the inductor can provide impedance increasing with frequency. Called inductive (shunt) peaking, this classical technique has been extensively used for bandwidth extension [40–42]. Fig. 2.2 depicts a practical implementation of this technique. With the increasing frequency, the increasing impedance of the inductor can compensate the decrease of the impedance of the load capacitance, therefore the total load impedance can maintain a constant value across a wide bandwidth, pushing the roll-off of the gain to higher frequency and leading to a constant gain over a wider bandwidth. In terms of the impedance of the RLC network:

$$Z_{RLC}(s) = \frac{(sL + R)}{1/sC} = \frac{R(s(L/R) + 1)}{s^2LC + sRC + 1}$$ (2.3)
as we can see, the inductor essentially introduces a zero, which cancels the main pole of the amplifier depending on the load capacitance.

The analysis above is based on the frequency-domain. Same conclusion can be got from the time-domain point of view. When a step signal is applied to the input, the inductor suppresses the current flowing through the resistor at the beginning, therefore the load capacitance discharges faster and hence the output voltage decreases faster, which implies a large bandwidth.

The maximum bandwidth can be improved to about 1.85 times, however, a large peak (about 20%) in the frequency response accompanies this increased bandwidth. Therefore depending on additional requirements on either maximum bandwidth, or maximally flat frequency response, or best group delay, the inductor values need to be chosen carefully to satisfy different requirements [6].

![Inductive shunt peaking technique](image)

Figure 2.2: Inductive shunt peaking technique.

**Inductive Series Peaking**

In addition to the shunt-connection with the capacitance, an inductor can also be placed in series with the capacitor. As illustrated in Fig. 2.3, the inductor is placed between the load capacitance $C_L$ and the transistor $M_1$, note that the $C_L$
here excludes $C_1$, the parasitic capacitances of the transistor $M_1$. By placing the series inductor, the load capacitance $C_L$ and the parasitic capacitance of $M_1$ ($C_1$) are isolated, therefore each node has a smaller capacitance, leading to a larger bandwidth.

On the other hand, from the time-domain point of view, when the step signal comes, the additional series inductor separating the $C_L$ and $C_1$ suppresses the current flowing through the $C_L$ at the beginning, therefore the $C_1$ discharges faster and the drain voltage decreases faster. After that, the current flows through $L_{ser}$, and $C_L$ begins to discharge and hence the output voltage decreases. Therefore, the series inductor makes the capacitances discharge in series in time, where the fall time will be shorter than the parallel case where all the capacitances discharge together. However, as we can see, this bandwidth extension technique trades the increased bandwidth with the larger delay.

Based on this concept, more advanced techniques have been developed, such as shunt and double-series peaking and T-coil bandwidth enhancement technique [6]. These techniques prove to be more effective in bandwidth extension. For example, the T-coil wideband technique can potentially achieve an improvement of about 3 times.

![Figure 2.3: Inductive shunt and series peaking technique.](image-url)
As we can see, inductive shunt peaking and series peaking are very effective approach for bandwidth extension, and one additional good feature is that they don’t consume or need little extra power. However, the drawback for these approaches is the large chip area consumed by those inductors.

### 2.2.2 Active Inductive Peaking

Inductive peaking is very effective in bandwidth extension. However, using passive inductors leads to large chip area. To resolve this issue, active inductive peaking can be used to replace the passive inductive peaking [28, 43, 44]. As depicted in Fig. 2.4-(a), an inductor can be replaced by a combination of a transistor and a resistor. The output impedance of the circuit in Fig. 2.4-(a) can be expressed as

\[
Z_{\text{out}}(s) = \frac{sRC_{gs} + 1}{g_m + sC_{gs}}
\]  

(2.4)

If the resistor value (R) is larger than the reciprocal of the transconductance of the transistor (1/g_m), the magnitude of Z_{out} increases with frequency, and thus this circuit acts as an inductor with the value of C_{gs}/g_m(R − 1/g_m) [5]. Thereby same as passive inductor, this active inductor can be used for bandwidth extension. Fig. 2.4-(b) shows a practical implementation of this technique. Compared to the passive inductive peaking, a large amount of chip area is saved.

This structure, however, suffers from large voltage drop caused by the V_{gs} of the M_3 and M_4, causing low output swing and limiting the use of this structure, especially at low supply voltage. Therefore in typical implementation of the active inductor, the bias voltage V_{ind} of the transistor M_3 and M_4 is connected to a voltage higher than the supply voltage.
Figure 2.4: Active inductive peaking technique (a) replacement of the inductor by a transistor and a resistor; (b) an example.

2.2.3 Negative Impedance Compensation

The inductive peaking techniques compensate the decreasing impedance of the output capacitance at higher frequencies, negative impedance compensation technique provides an alternative approach targeting the output capacitance [29,45]. As illustrated in Fig. 2.5, in addition to the main amplifier, a negative capacitance stage is added, and its output impedance can be expressed as [45]

\[
Z_{out}(s) = -\frac{1}{sC} \frac{g_{m56} + s(C_{gs56} + 2C)}{g_{m56}}
\] (2.5)
This negative capacitance can compensate the output capacitance of the main amplifier, therefore the equivalent output parasitic capacitances reduces, which results in a larger bandwidth.

In [29, 45], negative resistance compensation is also used to achieve a high gain. With the negative resistance stage, the equivalent load resistance can be expressed as

\[ R_L = R/\left(-\frac{1}{g_{m34}}\right) = \frac{R}{1 - g_{m34}R} \]  

(2.6)

where the negative resistance \(-1/g_{m34}\) is introduced by the negative resistance stage. As we can see, as long as the \(1/g_{m34}\) is large than \(R\), the total effective load resistance is enlarged, which benefits for a high gain. Therefore the negative impedance compensation, including capacitance and resistance, ensures the wide bandwidth and the high gain simultaneously. Although straightforward in concept and implementation, negative impedance compensation technique consumes more power due to the extra negative capacitance and resistance compensation stages.
2.2.4 Negative Miller Capacitance

Negative impedance compensation technique reduces the equivalent parasitic capacitances at the output. Negative Miller capacitance technique, or neutralization technique, on the other hand, reduces the input equivalent parasitic capacitances \([6, 41, 46]\). This technique is implemented by cross-placing Miller capacitances between the inputs and the outputs (Fig. 2.6). With the Miller capacitances, the effect of the parasitic capacitances between the gate and drain of the input transistors \((C_{gd})\) is partially canceled by the anti-phase transmission through the feedback Miller capacitances. Therefore the input capacitance introduced by Miller effect of the \(C_{gd}\) is greatly reduced, which benefits for a much smaller total input parasitic capacitance.

On one hand, the smaller input capacitance can move the second pole to a much higher frequency, and mitigates its effect on the bandwidth. On the other hand, it can also reduce the capacitive loading to the preceding stage, which benefits for a large bandwidth for a cascaded system. In order to achieve a good parasitic capacitance cancelation, an accurate estimate of the gate-drain parasitic capacitances is required,
which is sometimes not easy due to the dependence of the parasitic capacitances on gate voltage.

### 2.2.5 $f_T$ Doubler

An alternative technique to reduce the input capacitance is the $f_T$ doubler [41,47]. As depicted in Fig. 2.7, due to the series of the two parasitic gate-source capacitance, the effective input capacitance of a $f_T$ doubler is now approximately half of the simple differential stage. While the input capacitance is reduced by approximately half, the gain of the $f_T$ doubler is same as the simple differential stage, which is the product of the transistor transconductance and the load resistor. Therefore by using the $f_T$ doubler, the bandwidth can be extend without sacrificing the gain. Note that owing to the existence of other parasitic capacitances introduced by the transistors and by the tail current sources, the input capacitance of the $f_T$ doubler is higher than the approximated value estimated above, however, it is still an effective wideband circuit topology, especially for output buffer.

Some drawbacks of this topology need to be addressed here. First is the doubled power consumption and the doubled parasitic capacitances at the output node due to the doubled circuits. Second is the doubled output swing due to the doubled tail current sources, which may limit the circuit usage by forcing the transistors into triode region [5].

### 2.2.6 Inverse Scaling

For a cascaded system with N identical stages, the -3-dB bandwidth can be expressed as [5,6]

$$\omega_{-3dB} = \omega_o \sqrt{2^{1/N} - 1} \quad (2.7)$$
As we can see, with the increase of the number of stage, the bandwidth of the cascaded system reduces. Although wideband techniques can be used in each stage to extend the bandwidth, there is an alternative bandwidth extension technique from the system level. Inverse scaling technique, introduced in [48], was used to extend the bandwidth for a cascaded system. Fig. 2.8 illustrates two cascaded stages to show the basic concept of the inverse scaling technique. The transistor size and bias current of the preceding stage are $k$ times larger than the subsequent stage, while the load resistor
of the preceding stage is $k$ times smaller than the subsequent stage. The up-scaling of
the resistor ensures the constant voltage swing for each stage with the down-scaling
of the bias current for the cascaded system.

By down-scaling, the preceding stage sees a smaller load capacitance introduced by
the subsequent stage, which extends the bandwidth of the preceding stage. Therefore
by using this scheme, the bandwidth of the whole cascaded system can be extended.
This technique, however, trades the bandwidth extension with larger power consump-
tion, larger chip area and larger capacitance at the input of cascaded system.

2.2.7 Capacitive/Resistive Degeneration

![Figure 2.9: Capacitive/resistive degeneration.]

The wideband techniques discussed above mainly focus on dealing with the in-
put/output capacitances: resonating with or reducing the parasitic capacitances.
Here we will talk about an alternate bandwidth extension approach dealing with
the transconductance: capacitive/resistive degeneration. The technique is shown in
Fig. 2.9. As demonstrated in [5], the equivalent transconductance can be expressed
as we can see, due to the capacitive/resistive degeneration, a zero is introduced in the transfer function of the equivalent transconductance. By increasing the transconductance at high frequencies, the gain roll-off caused by the parasitic capacitances at high frequencies can be compensated, and hence the bandwidth is extended. Note that the gain will eventually decrease due to the higher frequency pole in the transconductance transfer function.

If the zero $1/(R_sC_s)$ is at the same frequency as the output pole $1/RC_L$, the bandwidth of the amplifier can be improved by $1+g_mR_s/2$ [5]. However, same as other degenerated amplifiers, the capacitive/resistive degenerated amplifier suffers from a low dc gain, which is degraded by $1+g_mR_s/2$. Therefore, this technique trades the large bandwidth with the low gain.

In addition to the bandwidth extension, this technique also exhibits a smaller input capacitance compared to the topology without degeneration. With capacitive/resistive degeneration, the input capacitance is reduced by $1+g_mR_s/2$. This in turn contributes a lower capacitive load to the preceding stage and benefits for a larger bandwidth of the preceding stage.

With the advantages of large bandwidth and low input capacitance, the low gain, however, limits the usage of this wideband technique to some extent, since more stages, and hence more power consumption, are demanded to achieve the gain requirement.
2.2.8 Shunt-Series Amplifier

In addition to the output capacitance, as mentioned earlier, the main pole limiting the bandwidth also depends on the output resistance. Therefore, reducing the resistance is also a potential approach for bandwidth extension. As shown in Fig. 2.10, a negative feedback resistor $R_f$ is inserted between the drain and gate of the amplifier, leading to an extensively used wideband amplifier structure: shunt-series amplifier, or resistive-feedback amplifier.

With the feedback resistor $R_f$, the input and output resistances ($R_{in}$ and $R_{out}$) of the amplifier are reduced. According to Miller effect, the input resistance can be expressed as

$$R_{in} = \frac{R_f}{1 - A_v} \quad (2.9)$$

where $A_v$ is the voltage gain from gate to drain. The output resistance can be given by [6]

$$R_{out} = \frac{R_f + R_s}{1 + R_s/R_1} \quad (2.10)$$
Since the signal flowing through the feedback resistor $R_f$ is opposite to that generated by the transistor $M_1$, it is preferable to choose a large $R_f$ and minimize this signal flow. Therefore $R_f$ is normally much larger than the source resistor $R_s$, and the output resistance can be approximately written as

$$R_{\text{out}} \approx \frac{R_f}{1 + R_s/R_1}$$

(2.11)

The small input/output resistances in turn benefit for a large bandwidth.

However, as other degenerated amplifier, this structure suffers from a low dc gain. Assuming $R_1$ is much larger than $1/g_{m1}$, the gain from gate to drain ($A_v$) is only about $-R_d/R_1$. However, with this $A_v$ value, and if choosing $R_d$ equal to $R_s$, we can get the same values of the input and output resistances. This is a quite useful characteristic, since it enables a simultaneous input and output impedance matching over a wide frequency range.

### 2.2.9 Cherry-Hooper Amplifier

As shown in Fig. 2.11, there is another technique based on local resistive feedback, which was proposed by Cherry and Hooper [49]. Different from the shunt-series amplifier, the Cherry-Hooper amplifier deals with a two-stage amplifier. As shown in Fig. 2.11-(a), a simple two-stage cascaded amplifier has two main poles, locating at the outputs of the two stages, respectively. Assuming two identical stages, the poles are equal and can be written as

$$\omega_p = \frac{1}{R_dC_L}$$

(2.12)

In order to extend the bandwidth of a two-stage amplifier, a local feedback resistor $R_f$ is inserted between the output of the second stage and that of the first stage.
Figure 2.11: (a) Simple two-stage cascaded amplifier (b) Cherry-Hooper amplifier; (c) Differential Cherry-Hooper amplifier.

(Fig. 2.11-(b)). With this feedback resistor, similar as the shunt-series amplifier, the output resistances of the first and second stage become very small [5]. Assuming that the output capacitances of the first and second stages in Cherry-Hooper amplifier are approximately equal to those in the simple two-stage cascaded amplifier, with the
smaller output resistances, the poles at the outputs of the first and second stage of the Cherry-Hooper amplifier are pushed to much higher frequencies [5]:

$$\omega_p \approx \frac{2g_{m2}}{C_{L1} + C_{L2}}$$  \hfill (2.13)

where assuming two poles are at the same frequency. As we can see, with a large transconductance, the pole frequency of the Cherry-Hooper amplifier can be much higher than that of the simple two-stage amplifier. Therefore a larger bandwidth can be achieved by Cherry-Hooper topology.

Fig. 2.11-(c) illustrates a practical implementation of differential Cherry-Hooper amplifier. Note that since both of the current sources in the first and second stage flow through the load resistors, Cherry-Hooper amplifier may have a large output swing, posing difficulties for the use of this technique in low supply voltage.

2.2.10 Active Feedback

In [41,43,47,50], active feedback was presented as a bandwidth extension technique (Fig. 2.12). Each active feedback stage consists of two main gain stages, and one feedback transconductance stage to return part of the output to the input of the second main stage. By active negative feedback, the gain-bandwidth product (GBW) can be effectively increased. According to [41], GBW of the active feedback stage is beyond the technology $f_T$, and is given by

$$A_0\omega_{-3dB} = f_T \frac{f_T}{f_{-3dB}}$$  \hfill (2.14)

In practical implementation, this technique can be combined with other wideband techniques to achieve a larger bandwidth. In [41], the inductive peaking and negative Miller capacitance techniques are employed with active feedback technique for further
bandwidth extension. This technique, however, may cause more power consumption due to the extra gain stage and feedback transconductance stage.

![Active Feedback Technique](image)

**Figure 2.12: Active Feedback Technique.**

### 2.2.11 Distributed Circuits

![Distributed topology](image)

**Figure 2.13: Distributed topology.**

Last but not the least, we will discuss the distributed circuits. As an important
bandwidth extension technique, distributed topology has been extensively used in wideband circuits [51–53]. As shown in Fig. 2.13, several transistors are uniformly distributed along transmission lines represented by series lumped inductors, and the gates and drains of the transistors are connected to the input and output transmission line, respectively.

Since the output transmission line exhibits an equivalent impedance of $Z_0/2$ as the load to the each transistor, where $Z_0$ is the characteristic impedance of the transmission line, when the input signal feeds into the input transmission line, the first transistor generates an output voltage of $g_{m1}(Z_0/2)V_{in}$. As the input signal goes through the input transmission line and feeds into each transistor in succession, the output of each transistor sums together in the output transmission line. Assuming the wave velocities of the input and output transmission lines are equal, then the amplifier exhibits an output of $ng_{m1}(Z_0/2)V_{in}$, where $n$ is the number of tapped stage.

Intuitively speaking, the input and output capacitances of the amplifier are absorbed into the input and output transmission line, respectively. Therefore the parasitic capacitances of the transistor do not directly impact the poles of the amplifier, they only reduce the characteristic impedance of the transmission lines. As long as the operation frequency remains below the cutoff frequency of the transmission line, the characteristic impedance maintains $Z_0$, leading to a constant gain over a large bandwidth.

By using a large number of transistor stages, a high gain can be achieved with a large bandwidth, which is at the cost of the increasing delay between the input and output. Therefore the distributed amplifier trades the bandwidth with the delay. Note that due to the transmission line loss and supply voltage limitation, the number of stage is limited in practical implementation [5]. In addition, this topology suffers from large power consumption and large chip area.
2.2.12 Current Mode Logic Circuits

![Current mode logic circuit topology diagram](image)

Figure 2.14: Current mode logic circuit topology.

Before finishing all the reviews, we will discuss a high-speed circuit technique: current mode logic (CML) circuits, which have been widely used in high-speed circuits, such as buffers [10, 43, 54], latches [10, 55], multiplexers [56–58], demultiplexers [59–61], and frequency dividers [62, 63]. These circuits were dominated by static CMOS circuits, which have the advantages of large noise margin and zero static power consumption. However, the PMOS with slower carrier mobility, and the large rail-to-rail operation can potentially cause the speed limitation of the static CMOS circuits. In addition, static CMOS circuits are sensitive to supply and ground bounce, which may result in jitter performance degradation.

CML circuit can mitigate these issues significantly. As illustrated in Fig. 2.14, the basic CML circuit consists of a differential pair, a pair of load resistor, and a tail current source. In the typical operation, the input signals are large enough so that the differential pair completely switches on and off, and the tail current flows completely to one side or the other side. Therefore, the output voltage swing is determined by the tail current source $I_s$ and load resistor $R$:

$$V_{swing} = I_s R \quad (2.15)$$
which is smaller than the rail-to-rail voltage. This benefits a high-speed operation and a low supply voltage. In addition, there is no PMOS in the CML circuit, avoiding the speed limitation factor in static CMOS circuits and further enabling the high-speed operation. In practical implementation, CML circuits can be combined with wideband techniques discussed above to further extend the operation speed. For example, inductive peaking and negative Miller capacitance techniques are used in CML circuits for a higher operation speed [55]. Furthermore, due to the differential structure, CML circuits naturally have high common-mode rejection capability and are insensitive to supply and ground bounce [55,64].
Chapter 3

Ultra-Wideband Impulse-Radio Receiver

In this Chapter, we will present an energy-efficient, high-speed, reconfigurable IR-UWB receiver architecture, and then focus on the analysis and implementation of all the IR-UWB building blocks, including the wideband low-noise amplifier (LNA), distributed pulse correlator, timing control circuit, variable gain amplifier and analog-to-digital converter. Major attention will be put on the wideband low-noise amplifier and the distributed pulse correlator, since these two blocks not only place important roles in the receiver, but face large design challenges as well.

3.1 IR-UWB Transceiver Architectures

Although this work will mainly focus on the IR-UWB receiver architecture, since the pulse detection function in the receiver share the same concept of distributed and time-interleaved architecture as the pulse generation function in the transmitter [18, 21], and for the sake of the system integrity, we will briefly review the IR-UWB transmitter architectures. The review of the transmitter also benefits for the future
demonstration of IR-UWB system, including transmitter, receiver and antennas.

3.1.1 Transmitter

There are several implementation approaches proposed for IR-UWB transmitter (Fig. 5.5): carrier-based up-conversion [65, 66], direct pulse generation [67, 68], waveform synthesis based on high speed digital-to-analog converters (DAC) [69]. Carrier-based up-conversion defeats one of the major advantages of IR-UWB (no carrier) and results in large circuit complexity and power consumption. Direct pulse generation tends to have limited reconfigurability because the high-Q pulse shaping filter is typically implemented as an off-chip discrete passive filter. Waveform synthesis provides good time and amplitude resolution and is reconfigurable for different pulse shapes. However, the high sampling rate required presents a challenge to both the digital-to-analog converter (DAC) implementation and the input data stream generation.

Figure 3.1: Conventional IR-UWB transmitter architectures.
3.1.2 Non-Coherent Receiver

Several different IR-UWB receiver architectures have been proposed, including direct conversion [70, 71], analog correlation [2, 67, 72, 73] and non-coherent energy detection [74–76].

![Non-Coherent IR-UWB receiver architecture.](image)

A non-coherent receiver typically detects the received UWB signal energy by multiplying the signal with itself. As shown in Fig. 3.2, it includes a UWB LNA, an envelop detector, an integrator, a variable gain amplifier (VGA), an analog-to-digital converter (ADC) and a digital baseband. The received weak pulse signal is first amplified by the UWB LNA, which amplifies the pulse signal while introducing the noise as small as possible. The envelop detector performs a squared operation, which is often implemented using a multiplier or mixer [74–77]. The squared signal is further integrated to make the narrow output pulse from the envelop detector longer, which makes it easier to be detected by the following building blocks (VGA and ADC). After the integration, the energy of the pulses can be recovered. The VGA provides additional gain, and changes the gain according to the output of the integrator, making the non-coherent receiver achieve a large dynamic range. The ADC converts the energy information into corresponding digital signals for the baseband processing.

The non-coherent energy detection requires neither channel estimation nor timing synchronization, which leads to a simple receiver architecture. In addition, the data decision is typically made by comparing the recovered pulse energy with the noise, which further simplifies the ADC requirement and hence the receiver architecture.
For example, [76, 77] use a comparator as a simple ADC to convert an analog signal to a digital form. Such a simple architecture, however, makes it more susceptible to noise and interference, and its achievable data rate is much lower than the coherent receiver architectures. Therefore most of the non-coherent receivers are more suitable for low data-rate applications (hundreds of Kb/s to tens of Mb/s).

3.1.3 Direct Conversion Receiver

In a direct conversion receiver, the received UWB pulses, often after down-converted to baseband, are directly sampled by high speed ADCs. As shown in Fig. 3.3, the direct conversion receiver typically consists of a UWB LNA, a VGA, an anti-aliasing filter, multiple parallel ADCs and the digital baseband. The received weak pulse signal is first amplified by the UWB LNA. The VGA is also used to maintain a large dynamic range. The anti-aliasing filter is added to solve the aliasing problem. Following the filter is the multiple parallel ADCs. For example, 32 phase-interleaved and 32 time-interleaved 1-bit ADCs are used in [71] and [70], respectively. A timing generator is included in the ADCs and generates the multiple clock signals for the ADCs. A digital baseband performs most of the signal processing function, including correlation and filtering.

Owing to digital-domain implementation of most function, the direct conversion receiver is a very flexible receiver architecture, and can take full advantages of digital signal processing, such as robustness, flexibility, scalability. However, it poses
requirement of ADCs with high sampling rate. For example, for a low-band (3.1 to 4.8 GHz) IR-UWB system, the sampling rate of the ADC needs to be approximately 10 GSamples/s to meet the Nyquist criterion. Even for the IR-UWB system operating in the band of sub-960 MHz, a sampling rate of 2 GSamples/s is still demanded. For example, in [70, 71], 32 1-bit ADCs with 2GHz sampling rate are implemented for a sub-960-MHz IR-UWB system. Multiple high-sampling-rate ADCs inevitably increase the system complexity and power consumption. In addition, a large amount of baseband processing are also required in digital baseband [70, 71], which further increases the system complexity and power consumption. Due to these challenges, the direct conversion receiver typically targets low data-rate applications operating in sub-960 MHz band.

### 3.1.4 Analog Correlation Receiver

![Analog Correlation IR-UWB receiver architecture.](image)

In order to avoid the use of high speed ADC, an analog correlation receiver (Fig. 3.4) moves the correlation function from the digital domain to the analog frontend: the received UWB pulses are correlated with a template pulse either locally generated or from a transmitted reference. This architecture reduces the required sampling rate of the ADC from the Nyquist rate as in direct conversion receiver to the pulse rate, which is normally much lower than the Nyquist rate. As implemented
in [2, 67, 73, 78], the speed requirement of the ADC are greatly reduced, and is determined by the pulse rate. The reduced requirement on speed benefits low power consumption of the ADC and hence the whole receiver. Additionally, a low resolution (as low as 3-bit) is sufficient for the analog correlation receiver [79], which further simplifies the design complexity and reduces the power consumption of the ADC. The achievable data-rate of the analog correlation receiver is higher than that in the previous two architectures. Compared to the other two receiver architectures, analog correlation receiver can achieve higher data rates [2, 67, 78, 80].

In the typical analog correlation receiver (Fig. 3.4), the UWB LNA, integrator, VGA and ADC behave similar as the ones in the non-coherent receiver. The timing controller synchronizes the receiver timing with the timing of the input pulses. The template pulse generator provides the template pulse for the correlator.

As mentioned above, there are two methods to provide the template pulse. In the transmitted-reference-based approach [72], a reference pulse is transmitted along with a data pulse. In the receiver, the reference pulse is delayed and then correlated with the data pulse. This approach suffers large performance degradation due to the noisy template even when the channel estimation is obtained using the reference pulse. Moreover, the wideband analog delay line required is difficult to implement, since it needs to achieve true time delay while maintaining the wideband characteristics of the reference pulse.

In the locally-generated-pulse approach, a template pulse is generated by a dedicated local pulse generator. Different approaches have been presented to generate the local template pulse. For example, [2, 67] generate a second-order derivative of Gaussian pulse as the template pulse. By using a saturation-region transistor, a weak-inversion-region transistor and a passive network, the generator performs square, exponential and second-order derivation functions, respectively. In [73], a simple digital template generator is used to generate two Gaussian pulses as the template. Instead
of using Gaussian pulse or its derivatives, [78] presents an alternative approach. It uses a windowed sine waveform as the template pulse, where a window generator windows a sine signal generated by a phase-locked loop (PLL).

Although the analog correlation receiver relaxes the requirement on ADC, it still faces several design challenges, such as the wideband LNA, the analog correlator, the template pulse generator, and timing synchronization. But due to relatively simple architecture and low power consumption, less susceptibility to noise and interference, and higher achievable data rate, we will develop the high-speed, energy-efficient IR-UWB receiver based on the analog correlation receiver with locally generated template pulse, and address the design challenges mentioned above through a custom developed analog correlator.

3.2 Proposed IR-UWB Receiver Architecture

Let us first examine a conventional analog correlation IR-UWB receiver as shown in Fig. 3.4. Here the correlation is accomplished by multiplying the incoming UWB pulses with a local template pulse from a local template pulse generator, and then integrating the correlator output over the pulse duration time to generate the correlation output. The three wideband analog signals (incoming UWB pulses, local template pulse, and correlator output) present great challenges for low power operation and CMOS implementation. The timing jitter from the analog local template generator also poses a serious problem for accurate timing synchronization.

Fig. 3.5 shows the proposed analog correlation IR-UWB receiver based on a new correlator called distributed pulse correlator (DPC), which includes built-in local template pulse generation and integration [21]. Note that there is only a single wideband analog signal (incoming UWB pulses) in this architecture, and therefore it can significantly reduce the power consumption and circuit complexity of the receiver.
In addition to the DPC, the analog correlation receiver consists of a LNA to amplify the received UWB signal, a timing control block to generate the trigger to the DPC and synchronize it with the incoming UWB pulse, a VGA and an ADC to amplify and digitize the DPC output, respectively.

### 3.3 Wideband Low-Noise Amplifier (LNA)

#### 3.3.1 Topologies of Wideband LNA

Recently, many possible solutions have been presented for wideband LNAs [81–84]. Distributed amplifiers [81] (Fig. 3.6-(a)) can provide very large bandwidth because of their unique gain-bandwidth trade-off. However, large power consumption and chip area make them unsuitable for typical low-power, low-cost UWB applications. Common-gate amplifiers [82,85] (Fig. 3.6-(b)) exhibit excellent wideband input matching, but suffers from a relatively large noise figure (NF). Narrow-band LNAs like an inductively degenerated common-source amplifier can also be converted into a wideband one by adding a wideband input matching network [83] (Fig. 3.6-(c)). However, the insertion loss of the passive input matching degrades the NF rapidly with frequency. Resistive-feedback amplifiers [84,86–88] (Fig. 3.6-(d)) have very good wideband input matching characteristic. However, low NF and low power consumption can be hardly achieved simultaneously across a large frequency range.
Figure 3.6: Possible wideband LNA topologies: (a) Distributed amplifier; (b) Common-gate amplifier; (c) Passive input matching; (d) Resistive feedback amplifier.

3.3.2 A Wideband LNA Design Example

As we can see from the discussion above, compared to other topologies, resistive-feedback amplifiers have the advantages of simple structure, small chip area, relatively low power consumption and relatively low NF. If the trade-off between the low
power consumption and low NF can be solved, this topology can be more attractive. In [89], noise cancelation technique is used to relax this trade-off in resistive-feedback amplifiers. In this work, the bandwidth of a resistive-feedback amplifier with noise cancelation is further extended by using the inductive shunt peaking and series peaking techniques. Therefore, good input matching, low power consumption, and low NF can be achieved simultaneously over a wide bandwidth.

**Circuit Analysis and Design**

![Schematic of the wideband resistive-feedback noise-canceling LNA.](image)

The wideband resistive feedback noise canceling LNA is shown in Fig. 3.7. The first stage is based on resistive feedback LNA. The $L_{\text{load}}$ and $L_{\text{g}}$ are added for inductive shunt peaking and series peaking, respectively, which will be discussed later. The second stage ($M_{3}$ and $M_{4}$, where $M_{3}$ acts as a source follower and $M_{4}$ acts as common-source amplifier) is added for wideband output matching and partially noise cancelation by connecting gate of $M_{4}$ with the gate of $M_{1}$.

Fig. 3.8 shows the wideband noise path and signal path of the LNA. Assume the
Figure 3.8: Wideband noise cancelation.

The main noise contribution of the LNA is from $M_1$, and its noise current is modeled by $i_n^2$. The noise current flows into the drain of $M_2$ (node X) through $M_2$ and part of it flows into input node (node Y) through $R_f$ and $C_f$. The noise voltage at node X and Y are fully correlated and have the same phases. These two noise voltages at node X and Y will appear at output node through $M_3$ and $M_4$, respectively. The noise voltage at the source of $M_3$ keeps the same phase as node X, while the noise voltage at the drain of $M_4$ has the opposite phase from node Y. Therefore, the two noise voltages at output node are still correlated but with opposite phases. The noise current of $M_1$ can be partially or completely canceled at the output node. It can be proved [89] that the noise current of $M_1$ can be completely canceled if

$$\frac{g_{m4}}{g_{m3}} = 1 + \frac{R_f}{R_s} \quad (3.1)$$

where $g_{m4}$ and $g_{m3}$ are the transconductance of $M_4$ and $M_3$, respectively, and assuming source follower $M_3$ has a voltage gain of one.

While the noise can be partially or completely canceled, the wanted signal can
be added at the output node. The signal voltages at node Y and X have opposite phases. Therefore, after \( M_3 \) and \( M_4 \), the signal voltages at output node have same phases and will not be canceled. Assuming source follower \( M_3 \) has unit voltage gain, then the overall voltage gain of the LNA is

\[
A_v = \frac{V_{out}}{V_{in}} = \frac{V_X A_{M3} + V_Y A_{M4}}{V_Y} = \frac{1 - g_{M1} R_f}{1 + \frac{R_f}{R_{load}}} - \frac{g_{m4}}{g_{m3}} \tag{3.2}
\]

In the case of complete noise cancellation, the overall voltage gain of the LNA is

\[
A_{v,nc} = \frac{1 - g_{M1} R_f}{1 + \frac{R_f}{R_{load}}} - 1 - \frac{R_f}{R_s} \tag{3.3}
\]

Assuming complete noise cancellation and neglecting noise contribution of the cascade device \( M_2 \) and series inductor \( L_g \), the noise factor after first stage is mainly determined by \( R_s, R_f \) and \( R_{load} \). Here the noise contribution by \( L_{load} \) can be included in that by \( R_{load} \). The noise factors of first stage can be derived [82]

\[
F_1 = 1 + \frac{R_f (1 + g_{M1} R_s)^2}{R_s (1 - g_{M1} R_f)^2} + \frac{(R_s + R_f)^2}{R_s R_{load} (1 - g_{M1} R_f)} \tag{3.5}
\]

where the second and third term are the noise contribution by \( R_f \) and \( R_{load} \), respectively. The noise factors at the output node can be derived as

\[
F_t = 1 + \frac{R_s (R_s + R_f)^2}{R_f (1 + g_{M1} R_s)^2} \tag{3.6}
\]

where the second term is the noise contribution by \( R_f \). The noise contribution by \( M_3 \) and \( M_4 \) are ignored here, since they are very small compared to that by \( R_f \). As is shown in (6), in addition to the noise contribution from \( M_1 \), the noise contribution
from $R_{\text{load}}$ can also be canceled at the output node, which can further lower the total noise factor.

Fig. 3.9 shows the simulated results of the LNA noise figure with and without noise cancelation. As shown in the figure, when the noise cancelation is enabled, the noise figure across the bandwidth that we are interested is lower than the case without noise cancelation.

![Simulated results of noise figure with and without noise cancelation.](image)

**Figure 3.9:** Simulated results of noise figure with and without noise cancelation.

The inductor $L_{\text{load}}$ of the LNA is used for shunt peaking purpose at high frequency. The inductive load can provide a resonant peaking at the output when the amplifier starts to roll off at high frequencies and equalize the power gain of the LNA to a constant value across the bandwidth by compensating the decreasing impedance of capacitance with the increase of frequency [90]. The series inductor $L_g$ is used to further boost the gain at high frequencies and extend the bandwidth by resonating with the parasitic gate to source capacitance of $M_1$ and parasitic capacitance formed by bottom plate of $C_f$ to substrate [42, 91]. Fig. 3.10 shows the simulated power gain with different value of $L_g$. As is shown in the Fig. 3.10, a higher gain and larger bandwidth can be achieved with an optimized $L_g$. In addition to the gain
boost and bandwidth extension, the resonance between $L_g$ and the parasitic gate to source capacitance of $M_1$ and parasitic capacitance formed by bottom plate of $C_f$ to substrate can benefit the input matching, high gain and low noise. As is shown in [82,83], resistive feedback LNA suffers from the tradeoff between input matching, gain and noise figure. The use of the $L_g$ to resonate with the parasitic capacitance at input node relaxes of the tradeoff and can achieve the good input matching, high gain and low noise simultaneously.

**Measurement results**

The wideband resistive feedback LNA is designed and implemented in a standard 0.18$\mu$m digital CMOS process. The microphotograph of this LNA is shown in Fig. 3.11. The chip occupies an area of 0.78$mm$ $\times$ 0.68$mm$. The LNA consumes 11.1 mW at a voltage supply of 1.8 V.

Fig. 3.12 shows the measured power gain of the LNA. The measured power gain achieves a maximum of 12.5 dB and the 3-dB bandwidth is 0.7 - 6.5 GHz. The LNA
remains 1-dB flatness from 0.7 to 4.5 GHz.

The measured input, output return loss and the measured reverse isolation are presented in Fig. 3.13. The input return loss is better than -11 dB in the 3-dB bandwidth and remains the performance up to 12 GHz. The output return loss is
Figure 3.13: Measured return loss and reverse isolation of the wideband LNA.

below -8 dB up to 12 GHz and remains a value of less than -10 dB across the most of the band. As is shown in Fig. 3.13, the measured reverse isolation is below -30 dB across the band of interest.

Figure 3.14: Measured and simulated noise figure of the wideband LNA.

The measured and simulated noise figures are illustrated in Fig. 3.14. The measured noise figure is from 3.5 to 4.2 dB across the 3-dB bandwidth, which is slightly
Figure 3.15: Measured IIP3 of the wideband LNA at 3GHz.

Figure 3.16: Measured IIP3 of the wideband LNA at different frequencies.
higher than the simulated noise figure. This is mainly due to the lack of very accurate noise model of deep sub-micron CMOS. The measured and simulated noise figures show the good effect of the wideband noise cancelation on the LNA.

The measured IIP3 with two-tone test \[?] is shown in Fig. 3.15 and Fig. 3.16. The measurement is performed at 1GHz, 3GHz and 5GHz. As is shown in Fig. 3.15 and Fig. 3.16. The measured IIP3 is -5.8 dBm, -5.9 dBm and -5 dBm at 1 GHz, 3 GHz and 5 GHz, respectively.

$$\text{Table 3.1: Performance summary of wideband LNA}$$

<table>
<thead>
<tr>
<th>Reference</th>
<th>CMOS Tech.</th>
<th>Bandwidth (GHz)</th>
<th>S21_{max} (dB)</th>
<th>S11_{max} (dB)</th>
<th>NF (dB)</th>
<th>IIP3 (dBm)</th>
<th>Supply (V)</th>
<th>Power (mW)</th>
<th>FOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>[84]</td>
<td>0.13-(\mu)m</td>
<td>5.9</td>
<td>16</td>
<td>-9</td>
<td>4.7-5.7</td>
<td>-</td>
<td>2</td>
<td>38</td>
<td>3.2</td>
</tr>
<tr>
<td>[86]</td>
<td>0.18-(\mu)m</td>
<td>2.4-6.6</td>
<td>9.8</td>
<td>-9</td>
<td>2.3-5.2</td>
<td>-70@4GHz</td>
<td>1.8</td>
<td>12.6</td>
<td>2.8</td>
</tr>
<tr>
<td>[87]</td>
<td>0.13-(\mu)m</td>
<td>1.7</td>
<td>17</td>
<td>-10</td>
<td>2.4@3GHz</td>
<td>-4.1</td>
<td>1.4</td>
<td>25</td>
<td>16.3</td>
</tr>
<tr>
<td>[88]</td>
<td>90-nm</td>
<td>0.4-1</td>
<td>16</td>
<td>-10</td>
<td>3.5-5.3</td>
<td><a href="mailto:-170@0.9GHz">-170@0.9GHz</a></td>
<td>1.2</td>
<td>16.8</td>
<td>1.1</td>
</tr>
<tr>
<td>[89]</td>
<td>0.25-(\mu)m</td>
<td>0.002-1.6</td>
<td>13.7</td>
<td>-8</td>
<td>1.9-2.4</td>
<td><a href="mailto:000@0.9GHz">000@0.9GHz</a></td>
<td>2.5</td>
<td>35</td>
<td>1.9</td>
</tr>
<tr>
<td>[85]</td>
<td>0.13-(\mu)m</td>
<td>0.8-2.1</td>
<td>14.5</td>
<td>-8.5</td>
<td>2.6</td>
<td>160@0.9,2GHz</td>
<td>1.5</td>
<td>11.6</td>
<td>3.9</td>
</tr>
<tr>
<td>This Work</td>
<td>0.18-(\mu)m</td>
<td>0.7-6.5</td>
<td>12.5</td>
<td>-11</td>
<td>3.5-4.2</td>
<td>-505GHz</td>
<td>1.8</td>
<td>11.1</td>
<td>7.5</td>
</tr>
</tbody>
</table>

\[84,87\]: differential outputs

Performance is summarized in Table 3.1. Comparisons with the previously reported LNAs are also listed in Table 3.1. Compared to the previously published LNAs, especially the wideband resistive feedback LNAs, this work achieves good input matching, low noise figure and low dc current consumption across a wide range of frequencies simultaneously.

A figure of merit (FOM) is used here to compare the performance of different LNAs with similar function. The FOM here evaluates the maximum power gain, 3-dB bandwidth, excess noise factor and the power consumption of the LNA and it is defined as

$$FOM = \frac{|S21|BW[GHz]}{|F - 1|Power[mW]}$$  \hspace{1cm} (3.7)
Based on the calculated FOM in Table I, this work has a good performance and is better than all the other works except [87]. But note that [87] used better technology, which typically has the benefits of higher gain and lower supply voltage.

3.4 Distributed Pulse Correlator

3.4.1 Rationale

The analog correlator performs the UWB pulse detection by multiplying the input UWB pulse with the local template pulse. The unique characteristics of the UWB pulse pose several challenges for the correlator design. First, IR-UWB pulses feature large bandwidth and sub-nanosecond time duration. This implies that the required sampling rate of the correlator is much higher than the pulse rate. For example, as we mentioned above, the correlator’s sampling rate would be approximately 10 GSamples/s for an IR-UWB system operating in the low-band (3.1 to 4.8 GHz), while the pulse rate ranges typically from hundreds of Kpulse/s to hundreds of Mpulse/s. Such a high sampling rate significantly increases the circuit power consumption. Besides, the large bandwidth of the UWB pulses also pose a stringent requirement on bandwidth of the correlator, especially at the interface with UWB pulse and template pulse as these two pulses with large bandwidth are directly fed to the correlator. Secondly, IR-UWB works in a low duty-cycled fashion, which implies that in order to save power consumption, the correlator is preferable to also work with low duty-cycle, only consuming power when there is an input pulse. Thirdly, the correlator also requires a high speed correlation, which arises from the correlation operation of short input pulses and local template pulses.

The distributed pulse correlator (DPC) shown in this work is custom developed for the pulse detection of UWB impulse radios. Analysis and implementation of the
DPC will be presented below. Here we will discuss two important techniques used in the DPC development. To achieve a high sampling rate while maintaining a low power consumption, time interleaving, a circuit and architecture technique proven effective in Nyquist-rate data converters [92], is well suited, as demonstrated previously in distributed waveform generator (DWG) for UWB pulse generation [18–20]. Time interleaving achieves a high sampling rate while keeping each time-interleaved block working at a much lower pulse rate, therefore the large power consumption overhead associated with the high sampling rate is greatly reduced. To overcome the bandwidth limitation of time interleaving structure, distributed circuit technique [93] can be applied, which absorbs the parasitics of each time-interleaved block and thereby extends the bandwidth.

3.4.2 Distributed Pulse Correlator Architecture

As shown in Fig. 3.17-(a), a DPC time-interleaves multiple analog multipliers in parallel. Input UWB pulses $V_a$ are distributed to these multipliers by the power distribution block. Each multiplier is triggered by an impulse generated by the impulse generator at a specific time $t_k$. The impulse samples the UWB pulse, and the samples are then multiplied by a tunable gain $c_k$, as illustrated in Fig. 3.17-(b). This is equivalent to correlating the UWB pulse with a sequence of narrow impulses with tunable amplitude. The combination of these narrow impulses essentially forms the built-in local template pulse. All multiplier outputs are then summed and integrated to generate the overall correlation result [21].

As mentioned earlier, such a time-interleaved correlator architecture allows the detection of low pulse-rate UWB pulses with very high sampling rate. The high sampling rate is achieved by the small delay between neighboring taps set by the trigger distribution block. Since these multipliers only consume power when they
Figure 3.17: (a) Generic architecture of a distributed pulse correlator (DPC); (b) Waveform illustration of the DPC.
are triggered, this architecture is especially energy efficient for low duty-cycle IR-UWB pulse detection. Further, the template pulse generators are embedded in the analog multipliers, therefore large analog bandwidth is only needed at the power distributed block, which, as mentioned earlier, can be achieved by employing the distributed circuit technique. Additionally, the timing of the built-in local template pulse generation in a DPC is entirely determined by the digital trigger distribution network, and hence the timing jitter is minimized. Lastly, the transversal structure of a DPC enables reconfigurable local template pulses when changing the tunable gains \((c_k)\) and tap delay, which makes the receiver more waveform and frequency agile, and can handle different channel environments, PVT variations and different regulatory requirements. The transversal structure of the DPC also enables the incorporation of linear equalization to address the issues of multipath and ISI.

3.4.3 Built-In Template Pulse Generation

![Figure 3.18: A DPC can be considered as a transversal finite impulse response filter.](image)

As is shown in Fig. 3.18, a DPC can be treated as a transversal finite impulse
response (FIR) filter. A function $\delta(t - k\tau)$ is used to represent the triggered impulse sequence with a tap delay $\tau$. The UWB pulse $V_a$ is sampled by the impulse sequence, and then multiplied with a tunable coefficient $c$ in each tap. The impulse sequence with the tunable individual coefficient essentially forms the built-in local template pulse in the DPC. All the taps are then summed to generate the output signal $y(t)$. The impulse response of this FIR filter can be expressed as

$$h(t) = \sum_{k=1}^{N} c_k \delta(t - k\tau) \quad (3.8)$$

As we can see, the tap coefficients $c_k$, the tap delay $\tau$ and the impulse signals $\delta$, which mainly determine the local template pulses, can be controlled to change the impulse response of the DPC. The corresponding frequency response is

$$H(j\omega) = \sum_{k=1}^{N} c_k \exp(-j k \omega \tau) \quad (3.9)$$

Hence the frequency response of the FIR filter output can be described as

$$Y(j\omega) = V_a(j\omega)H(j\omega) = \sum_{k=1}^{N} c_k V_a(j\omega) \exp(-j k \omega \tau) \quad (3.10)$$

and the time-domain output can be derived as

$$y(t) = \sum_{k=1}^{N} c_k V_a(t - k\tau) \quad (3.11)$$

by tuning the tap coefficients $c_k$ and the tap delay $\tau$, the transversal structure of the DPC enables reconfigurable local template pulses, which makes the receiver more frequency and waveform agile.
3.4.4 Performance Analysis

The template pulse places an very important role in the analog correlation receiver. Here we will use signal-to-noise ratio (SNR) of the DPC as a measure to examine the performance of the receiver with the variation of the template, such as the shape mismatch and the timing mismatch. Using the fourth-derivative of the Gaussian pulse \cite{94} as the input UWB pulse and based on the definition of the SNR \cite{95}, let us first examine the performance relative to the template coefficients. As shown in Fig. 3.19-(a), where assuming a good timing (well-synchronized with the input pulse and exactly 100-ps tap delay), we can construct a well-matching template pulse using five Gaussian pulses with proper tap coefficients. Each Gaussian pulse is expressed as

\[ p(t) = \pm A \exp\left(-2\pi t^2/\alpha^2\right) \]  \hspace{1cm} (3.12)

where \( A \) is the pulse amplitude which is controlled by template tap coefficient \( c_k \), and \( \alpha^2 \) is the shape factor related to the pulse width \cite{94}. As we can see from Fig. 3.19-(b), the constructed template pulse shows a very good SNR compared to the idea template pulse with the same shape as the input pulse. When the template coefficients change, the template pulse offsets from the well-matching shape, degrading the SNR. As illustrated in Fig. 3.19-(b), when changing the normalized coefficients from the minimum to the maximum, the SNRs reach optimal points during the change, which are the coefficients of the previous well-matching shape. Similar results can be got when changing the pulse width. In Fig. 3.19-(c), with the change of the shape factors of the Gaussian pulses, the widths change correspondingly, and the optimal SNRs can be got when the impulses match the corresponding taps of the input pulse. Therefore in order to maximize the SNR performance, the tuning capability for tap coefficient and pulse width is desirable.

With a good template pulse shape, the timing mismatch also affects the SNR...
Figure 3.19: Simulation results: (a) comparison of input pulse and template pulse; normalized SNR with regard to (b) template tap coefficient; (c) shape factor (related to pulse width); (d) timing drift or asynchronization; (e) individual tap timing error.
performance. The timing mismatch may be introduced by two main sources. One is the imperfect synchronization or timing drift. Without perfect synchronization, the whole template may offset the input pulse in time domain. In addition, due to the jitter in timing control block (providing Trigger to the DPC), the whole template pulse may also drift from the original place even though a good synchronization is achieved at the beginning, leading to the same effect as the imperfect synchronization. Fig. 3.19-(d) illustrates the SNR performance with regard to the timing error caused by the imperfect synchronization or timing drift, where assuming no jitter is introduced from the trigger distribution block in the DPC. As we can see, the SNR degrades with the increasing timing error. Therefore to keep a good system performance, we need to achieve a good synchronization and minimize the timing drift, which requires proper acquisition and tracking schemes.

The other reason for timing mismatch is due to the jitter of the trigger distribution block in the DPC, which may cause each Gaussian pulse offset from its ideal place. Fig. 3.19-(e) shows the SNR performance with the change of the impulse position. As expected, the SNR degrades with the increasing offset of the impulse position. Thereby a delay-locked loop (DLL)-based trigger distribution is needed to maintain an accurate timing control.

3.5 IR-UWB Receiver Prototype Implementation

3.5.1 Ultra-Wideband LNA

To build an ultra-wideband LNA for the proposed IR-UWB receiver, several modifications need to be applied to the previous wideband LNA. First, a bandpass frequency characteristic is needed for UWB LNA, which means the higher-corner cutoff frequency of the previous wideband LNA needs to be extended to around 10 GHz,
and the lower-corner cutoff frequency needs to be moved to around 3 GHz from DC. These bandwidth requirements imply the changes on the bandwidth extension part of the LNA, and input matching network as well. Secondly, due to the merit of differential signaling, such as less sensitive to environmental noise, the receiver internally uses differential signaling instead of single-ended signaling. However, because of the difficulty of generation of differential UWB pulses, most of the UWB system uses single-ended UWB signal to communicate between the transmitter and the receiver. Therefore, in order to maximize the merit of differential signaling, it’s better to convert the single-ended signal to differential signal in the first stage of the receiver, i.e., UWB LNA. Hence, single-ended to differential conversion is needed in the UWB LNA. Based on these design considerations, a UWB LNA for the IR-UWB receiver is presented to address these issues.

Circuit Analysis and Design

![Schematic of the UWB LNA in the receiver prototype.](image-url)

Fig. 3.20 illustrates the UWB LNA for the IR-UWB receiver. A high-pass filter ($L_1$ and $C_1$), together with the source degenerated inductor $L_s$ and the parasitic capacitance of $M_1$, acts as the input matching network to improve the UWB input
matching of the resistive-feedback structure [76,96]. In addition, the high-pass filter also sets the lower-corner cutoff frequency of the UWB LNA. Compared to the multi-section reactive input matching network [83], this simple high-pass input matching network has less passive components, which benefits for smaller noise figure and smaller chip area. On the other hand, the source degeneration inductor $L_s$ of the input matching network can improve the non-linearity distortion. Combined with the resistive feedback, this shunt-series dual feedback also reduces the dependence of the amplifier on device parameters [6].

The low-Q inductor $L_{11}$ is utilized here for shunt peaking at high frequency. The gain roll-off at high frequency can be compensated by the resonant peaking provided by the inductive load, which can also equalize the power gain of the LNA to a constant value across the bandwidth of interest by compensating the decreasing impedance of capacitance at higher frequency [97].

![Figure 3.21: Power gain versus frequency with different $L_d$.](image)

In order to further boost the bandwidth and improve the gain flatness, the inductor $L_d$ is added for series peaking. This interstage passive network isolates the total
capacitance at the output of the first stage from the parasitic capacitance of the common-source transistor $M_1$ and absorbs these capacitance, which benefits for the bandwidth enhancement [98]. In addition, adding the series peaking $L_d$ can ease the design of the shunt peaking part, which has the tradeoff between the power gain and the bandwidth caused by the opposite effect from the $L_{i1}$. Therefore, with a smaller capacitance shown at the output node, a larger value of $L_{i1}$, which leads to a high power gain, can be chosen to resonate at the same high frequency. An optimized $L_d$ can be chosen through the simulation. Fig. 3.21 shows the simulated power gain with different value of $L_d$. As shown in Fig. 3.21, an optimized $L_d$ can be found for the high gain and large bandwidth simultaneously.

A quasi-differential stage is cascaded to provide single-ended to differential conversion. Source followers are utilized to achieve a good wideband output matching with the input of the following DPC, which is implemented by on-chip transmission line with 50-Ω characteristic impedance.

**Measurement results**

The UWB LNA was designed using Agilent Advanced Design System (ADS), and fabricated in a standard 0.18 μm digital CMOS technology. On-wafer measurements are performed for the power gain, input and output return loss, reverse isolation, noise figure, IIP3 and group delay. The LNA consumes 12.2-mW dc power at a voltage supply of 1.8 V.

The measured power gain of the LNA is shown in Fig. 3.22. The measured power gain achieves a maximum of 10.1 dB, and the 3-dB bandwidth ranges from 2.3 to 9.8 GHz. Compared to the simulated power gain and the bandwidth, which are also shown in Fig. 3.22, the measured power gain is lower and the bandwidth is smaller. The discrepancy between the measurement and simulation results may be caused by the inaccuracy of the transistor and inductor model, which may result in extra
parasitic effects. With the inaccuracy taken into account, a modified simulation is performed. As is shown in Fig. 3.22, the measurement result is close to the modified simulation result.

The measured input, output return loss and the measured reverse isolation are illustrated in Fig. 3.23. The input return loss shows a good input matching of below
-10 dB from 2.8 to 7.1 GHz, and it is lower than -6.5 dB across the whole bandwidth of interest. The output return loss implies a good output matching better than -11 dB up to 13.5 GHz. As is shown in Fig. 3.23, the measured reverse isolation is lower than -26.5 dB over the 3-dB bandwidth.

Figure 3.24: Measured and simulated noise figure of the UWB LNA.

The measured and simulated noise figures are presented in Fig. 3.24. The noise figure ranges from 3.9 to 6.9 dB over the 3-dB bandwidth, and a maximum noise figure of 5.2 dB is achieved across the frequency from 2.7 to 9.8 GHz. The slight discrepancy between the measured and simulated noise figure may result from the inaccuracy of the noise model of the deep sub-micron CMOS.

Two-tone test is carried out for the measurement of third-order intermodulation distortion. The measurement is performed at 3 GHz, 6 GHz (which is shown in Fig. 3.25.) and 9 GHz. The measured IIP3 is -5.5 dBm, -4.1 dBm and -4.9 dBm at 3 GHz, 6 GHz and 9 GHz, respectively.

In Fig. 3.26, the measured group delay is reported. The group delay ranges from 70 to 162 ps over the 3-dB bandwidth. A relatively small group delay variation of ±20 ps is achieved across the frequency from 1.8 to 8.3 GHz.
3.5.2 Distributed Pulse Correlator

The schematic of a DPC as implemented in the prototype receiver is shown in Fig. 3.27. 8-tap analog multipliers are time-interleaved in parallel. The trigger signal is distributed through a DLL, where the delay per stage was tuned by current and
Figure 3.27: Schematic of the DPC in the receiver prototype.

was designed at nominally 100 ps, which corresponds to a sampling rate of 10 GSample/s. The input differential UWB pulses are distributed to each multiplier through a differential on-chip transmission line in order to achieve the large bandwidth for \( v_a \), similar to a distributed amplifier. Differential correlation output currents from each pulsed multipliers are integrated on capacitors \( C_1 \) and \( C_2 \). After the correlation is done and with additional hold time \( \tau_{\text{hold}} \) for the next stage circuit, switches \( S_1 \) and \( S_2 \) reset the voltage levels on \( C_1 \) and \( C_2 \), which also helps reducing the effects of DC offset and 1/f noise on the integration capacitors. The hold time delay is controlled by the same control voltage as the trigger distribution block to ensure an accurate delay.

The on-chip transmission lines in the DPC were implemented using a pair of multi-layer coplanar waveguides (MCPW) [99]. This distributed circuit structure extends the bandwidth of the time-interleaved DPC by absorbing the input capacitance of each pulsed multiplier, and diminishes the amplitude attenuation and the rise/fall time slow-down of the input pulses for each pulsed multiplier. As illustrated
Figure 3.28: Simulation results of the transmission line: (a) attenuation with different number of taps; (b) delay from each tap to the succeeding taps.
in Fig. 3.28-(a), the MCPW-based transmission lines exhibit low loss and large bandwidth, showing a better performance than LC artificial transmission lines in conventional CMOS distributed circuits. Fig. 3.28-(b) shows the simulation result of the transmission line delay from each tap to the succeeding taps. A very linear delay distribution can be achieved. The linear delay is ensured by both large bandwidth and the equally distributed taps. Compared to the delays from the trigger distribution block, the delays from this transmission line is very small. However, note that due to the accumulation of the delay, the latter taps suffer from more delay than the former taps, and those delays have the similar effect as the jitter introduced by trigger distribution, may degrading the SNR.

3.5.3 Pulsed Multiplier

Analysis of the pulsed multiplier

In a DPC, the most critical component is the multiplier, which also carries out the sampling operation. Therefore, conventional analog multipliers needs to be customized for the pulsed operation of sampling. Gilbert cell is a popular topology in the implementation of four-quadrant analog multipliers [100], and the multiplication function in the correlator of IR-UWB receivers [101]. As shown in Fig. 3.29-(a), in a pulsed Gilbert cell, the input differential signal (RF) is converted to current by the transconductance differential pair $M_5$-$M_6$, and then sampled by the two upper differential pairs $M_1$-$M_4$. The multiplication with $c_k$ is performed by controlling the dc current through the tail current source. This circuit, however, is ill-suited for pulsed operation with low duty cycle due to the constant dc bias current, and suffers from strong sampling signal pass-through.

In order to achieve the truly pulsed operation, a new pulsed multiplier was developed, as shown in Fig. 3.29-(b). Instead of generating differential currents, either
Figure 3.29: Schematic of (a) a Gilbert cell, (b) the pulsed multiplier in a DPC (simplified schematic), and (c) a switched transconductor mixer.

$M_5$ or $M_6$ is turned on, and the result current impulse samples the differential input signal by the transconductance differential pair $M_1$-$M_2$ or $M_3$-$M_4$. The multiplication with $c_k$ is performed by controlling the dc bias on $M_1$-$M_4$ and hence the amplitude of the sampling current impulse. It is worth noting that the pulsed multiplier in this work is similar in topology but different in operation from a switched transconductor mixer [102], which are shown in Fig. 3.29-(c). Our multiplier works in a pulsed fashion and does not need continuous complementary LO signals, i.e., it is only activated
during the sampling impulse, and remains off most of the time.

Compared to the pulsed Gilbert cell, the pulsed multiplier has several merits to make it a suitable topology for the pulsed operation with low duty cycle. First, no constant dc bias current is needed for the pulsed multiplier. It only consumes power when there is sample signal ($Sample^+$ and $Sample^-$), which leads to a very low power topology for pulsed operation.

Secondly, since there is no tail current source as in the pulsed Gilbert cell, the pulsed multiplier only requires a peak switch voltage of $V_{GS,M_{5,6}}$. However, for the pulsed Gilbert cell, due to the cascode structure, a much higher gate voltage is required to switch on/off the current impulse at the upper differential pair (see Table 3.2). The low peak switch voltage benefits the low supply voltage (which is 1.3 V in this case), and further low power consumption.

Thirdly, compared to the pulsed Gilbert cell, the pulsed multiplier is also less susceptible to the thermal noise arising from the impulse switches, which is because the noise current generated by the pulse switches are common-mode output currents and can be rejected at the output. In addition to the pulse switches, the transconductance transistors also contribute thermal noise, which can be modeled as a drain current proportional to $g_m$ [102]. Note that in order to cancel out the thermal noise from the transconductance transistors at the output, a good match between those transistors is required, which can be assured by a good circuit and layout design.

With the pulse switches moving to the tail, the transconductance of the pulsed multiplier is not degraded and equals to the $g_m$ of the transconductance differential pairs, where assuming the same transistor sizing and bias of the differential pairs. Besides, same as pulsed Gilbert cell, the pulsed multiplier is also four-quadrant, which is a required function for correlator.

The comparative analysis are summarized in Table 3.2. As we can see, compared to the pulsed Gilbert cell, the pulsed multiplier is more suitable for the pulsed application
with low duty cycle.

Table 3.2: Comparisons between pulsed multiplier and Gilbert cell

<table>
<thead>
<tr>
<th></th>
<th>Pulsed multiplier</th>
<th>Pulsed Gilbert cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power consumption</td>
<td>Lower</td>
<td>Higher</td>
</tr>
<tr>
<td>Peak switch voltage</td>
<td>$V_{GS,M_{5,6}}$</td>
<td>$V_{DS, CurSrc} + V_{DS,M_{5,6}} + V_{GS,M_{1,2,3,4}}$</td>
</tr>
<tr>
<td>Thermal noise of the pulse switches</td>
<td>Common mode</td>
<td>Differential</td>
</tr>
<tr>
<td>Thermal noise of the transconductance devices</td>
<td>$\propto g_m$</td>
<td>$\propto g_m$</td>
</tr>
<tr>
<td>Transconductance</td>
<td>$g_m$</td>
<td>$g_m$</td>
</tr>
<tr>
<td>Number of quadrant</td>
<td>Four</td>
<td>Four</td>
</tr>
</tbody>
</table>

Implementation of the Pulsed Multiplier

![Schematic of the pulsed multiplier with built-in impulse generator.](image)

Figure 3.30: Schematic of the pulsed multiplier with built-in impulse generator.

Fig. 3.30 shows the schematic of the pulsed multiplier used in the DPC. The pulsed multiplier on one hand samples the UWB pulse, and on the other hand, it also multiplies the sampled signal by a tunable coefficient. According to the polarity of the UWB pulse to be detected, the trigger signal is switched between two paths. Within either signal path of the impulse edge generation, the trigger signal is further
split to generate a rising edge and a delayed falling edge. These two edges combine on either $M_5$-$M_6$ or $M_7$-$M_8$ to generate a narrow current impulse. This current impulse turns on and sets the transconductance of the differential pair $M_1$-$M_2$ or $M_3$-$M_4$. The amplitude of current impulse can be manually tuned by changing the reference bias current. In future implementation, the control signals can be generated by the baseband processor and stored in 8 multi-bit registers, which further control the bias currents. These control signals can also be calibrated in baseband processor according to the detection results. The register-based approach can be implemented using digital circuits, and has very low power consumption. Within this short duration of current impulse, differential pair $M_1$-$M_2$ or $M_3$-$M_4$ multiplies part of the input UWB pulse with the transconductance. So the local template pulse shape is determined by the various transconductance of each pulsed multiplier.

This pulsed multiplier topology also benefits the integrator design. Since $M_5$-$M_8$ remain off most of the time, the leakage problem on the integration capacitor is greatly mitigated compared to the conventional multiplier-based correlators. Thus small size of integration capacitor can be used to improve the conversion gain.

### 3.5.4 Timing Control Circuit

The synchronization between Trigger and the incoming amplified signal $V_a$ is performed by a DLL-based timing control block. To fully cover the minimum time slots for certain pulse rate, the DLL needs to have the corresponding tuning ability. For a 250-MHz pulse rate, the minimum time between two consecutive pulses is 4 ns. Therefore, in the prototype receiver, a 40-tap DLL with a timing resolution of 100 ps is used. It has a measured root-mean-squared (rms) jitter of 10.3 ps. A 6-bit controlled MUX is used to select one of the DLL outputs as the right timing for Trigger.
During the synchronization phase, a training sequence is sent by the transmitter. The control signals of the MUX change incrementally to sweep the whole pulse period, and the proper timing can be set by detecting the acquisition according to the output of the ADC. In this prototype, the MUX is manually controlled, which can be carried out by the baseband processor in future implementation. In addition, due to the timing drift discussed earlier, a tracking scheme is also preferred in future implementation, which can detect the small degradation of the output signal, and change the control signals around the previous optimal timing.

3.5.5 Variable Gain Amplifier

A VGA is needed to amplify the output of the DPC to improve the dynamic range of the receiver. As is shown in Fig. 3.32, a two-stage cascaded VGA with exponential gain control characteristic is utilized. The variable gain ranges from -10 to 32 dB with a maximum bandwidth of 400 MHz, and the overall variable gain is equally distributed over these two stages. The gain controller is tuned by external gain control signal $V_{ref}$ to control the current bias of the gain cell. The gain cell generates the exponential gain characteristic according to the change of the current.
bias [103]. A common mode feedback is also included in each gain cell to stabilize the output common mode level for this fully differential structure.

### 3.5.6 Analog-to-Digital Converter

A 3-bit ADC converts the VGA outputs to the corresponding digital signal for baseband processing. Thanks to the analog correlation receiver architecture, the requirement of the ADC sampling frequency is released, where a pulse rate of up to 250 MHz is needed. Flash architecture is chosen due to their simple structure.
Figure 3.33: Block Diagram of the ADC in the receiver prototype.

and fast operation, and it is well suited for low bit resolution ADC [104]. Fig.3.33 illustrates the block diagram of the 3-bit flash ADC in the receiver prototype. The ADC uses the reference ladder to subdivide the ADC reference voltages into $2^3$ sets of differential reference voltages and feeds them into the corresponding preamplifiers. The clocked, regenerative comparators [105] follow the preamplifiers and generate the thermometer code, which is then processed by the digital back-end and converted into the binary code.
3.5.7 Measurement Results

The DPC-based IR-UWB receiver prototype was designed using Agilent Advanced Design System (ADS), and fabricated in a 0.18-μm standard digital CMOS technology with low-resistivity substrate. The chip micrograph is shown in Fig. 3.34. The die size including pad frame is 3 mm × 2.4 mm. The active area is 1.23 mm². Note that a stand-alone LNA core circuit was also implemented to characterize its pulse and frequency response.

![Chip micrograph of the prototype IR-UWB receiver.](image)

The prototype receiver is characterized on-wafer in time domain using a test setup shown in Fig. 3.35. Signal source provides continuous-wave (CW) sinusoidal signals and split them into three. Two are used as the input trigger for the impulse generator and sampling oscilloscope, respectively, and one is used as the clock signal for the receiver. Input Gaussian-shape pulse $V_{in}$ is generated by the impulse generator, which is modulated by baseband data stream from an arbitrary waveform generator (AWG), which is synchronized with the CW signal source by its 10-MHz internal clock.

The 70-ps pulse duration $V_{in}$ is converted into a differential UWB pulse after
Figure 3.35: Test setup for the prototype receiver characterization.

Figure 3.36: Measured pulse response of the LNA.
Figure 3.37: Measured and simulated normalized coefficient with different bias current.

passing the LNA with about 400-ps pulse duration and 60-mV voltage swing, as shown in Fig. 3.36.

To verify the functionality and reconfigurability of DPC, each tap in the DPC is first used to correlate with different part of the UWB pulse independently by sweeping the trigger time delay. Correct output polarity and amplitude are observed. As shown in Fig. 3.37, the coefficient shows a good tuning linearity with the change of reference bias current in both measurement and simulation. Then multiple taps are used to correlate with the input UWB pulses, the DPC output with correct timing and local template is shown in Fig. 3.38-(a). The DPC output pulse show three phases during the operation: correlation, hold and reset. The hold phase is designed at about 1 ns for the following VGA and ADC, and the reset phase take about 2 ns. The relatively long reset time is caused by the parasitic capacitance of the integration capacitors, which can be further reduced in the improved design. Using the same local template, the DPC outputs (before the VGA) with different timing offset between the input UWB pulses and trigger signals are shown in Fig. 3.38-(b). With no timing offset,
the input UWB pulse is fully detected. With 100-ps timing offset, since the polarity of each tap of template pulse is just opposite to the input pulse, a negative DPC output with smaller amplitude than the fully detected case is observed. With 500-ps timing offset, there is almost no correlation waveform observed. So accurate timing and synchronization are critical for the correct detection using DPC. Fig. 3.38-(c) shows the correlation results for different local templates without timing offset: when tap 1, tap 1,5 or tap 1,4,5 are used, which only cover part of the UWB pulse, the correlation result has smaller amplitude compared to the fully detected case using tap 1~5 as shown in Fig. 3.38-b. These results verify that the built-in template pulse is reconfigurable.

![Figure 3.38](image_url)

(a) (b) (c)

Figure 3.38: Measured DPC outputs with (a) correct timing and templates; (b) different timing offset; (c) different templates.

Fig. 3.39 shows the complete receiver test results with maximum achievable pulse rate of 250 MHz and with the modulated data stream from an AWG. The DPC achieves an energy efficiency of 40 pJ/pulse for a 250-MHz pulse rate. Note that this figure of merit will remain constant even when the pulse rates is changed, which
Figure 3.39: Measurement results of the receiver with modulated data.

requires a different DLL frequency, since all its components are inherently duty-cycled. The LNA, VGA, ADC and DLL consume 12.2, 3.7, 2.9 and 2.4 mA, respectively, from a 1.8-V power supply. Therefore, an energy efficiency of 190 pJ/pulse for 250-MHz pulse rate is achieved. The DPC-based receiver performance is summarized in Table 3.3, which demonstrates the validity of the time-interleaved, distributed technique for an energy efficient, high sampling rate, large bandwidth pulse detection and the whole receiver. The comparison with other receivers is shown in Table 3.4.

<table>
<thead>
<tr>
<th>Technology</th>
<th>0.18-μm</th>
<th>LNA gain</th>
<th>10.1 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply voltage</td>
<td>1.8 V</td>
<td>LNA NF</td>
<td>3.9 dB</td>
</tr>
<tr>
<td>Modulation</td>
<td>OOK</td>
<td>DPC sample rate</td>
<td>10-GS/s</td>
</tr>
<tr>
<td>Pulse rate</td>
<td>250 Mpulse/s</td>
<td>DPC energy efficiency</td>
<td>40 pJ/pulse</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>3 ~ 10 GHz</td>
<td>VGA gain range</td>
<td>-10 ~ 32 dB</td>
</tr>
<tr>
<td>RX energy efficiency</td>
<td>190 pJ/pulse</td>
<td>DLL resolution</td>
<td>100 ps</td>
</tr>
</tbody>
</table>

A figure of merit (FOM) is used here to compare the performance of different receivers with similar function. The FOM here evaluates the energy efficiency and
Table 3.4: Performance comparisons with other IR-UWB receiver

<table>
<thead>
<tr>
<th>Reference</th>
<th>CMOS Technology</th>
<th>Bandwidth (GHz)</th>
<th>Pulse/ Data Rate (MHz)</th>
<th>Energy Efficiency (pJ/Pulse)</th>
<th>FOM (pJ/Pulse/GHz)</th>
<th>Coherent (Y or N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[74]</td>
<td>0.18 μm</td>
<td>3 ~ 5</td>
<td>50</td>
<td>576</td>
<td>288</td>
<td>N</td>
</tr>
<tr>
<td>[75]</td>
<td>0.18 μm</td>
<td>3 ~ 5</td>
<td>16.7</td>
<td>2500</td>
<td>1250</td>
<td>N</td>
</tr>
<tr>
<td>[76]</td>
<td>0.18 μm</td>
<td>3 ~ 5</td>
<td>100</td>
<td>144</td>
<td>72</td>
<td>N</td>
</tr>
<tr>
<td>[77]</td>
<td>90 nm</td>
<td>3.6 ~ 4.3</td>
<td>1</td>
<td>1100</td>
<td>1571</td>
<td>N</td>
</tr>
<tr>
<td>[67]</td>
<td>0.18 μm</td>
<td>3 ~ 5</td>
<td>200</td>
<td>405</td>
<td>202</td>
<td>Y</td>
</tr>
<tr>
<td>[73]</td>
<td>0.18 μm</td>
<td>&lt; 1</td>
<td>1</td>
<td>1000</td>
<td>1000</td>
<td>Y</td>
</tr>
<tr>
<td>[2]</td>
<td>0.18 μm</td>
<td>3 ~ 5</td>
<td>165</td>
<td>600</td>
<td>300</td>
<td>Y</td>
</tr>
<tr>
<td>[106]</td>
<td>0.13 μm</td>
<td>6 ~ 10</td>
<td>2000</td>
<td>72.9</td>
<td>18.5</td>
<td>Y</td>
</tr>
<tr>
<td>[78]</td>
<td>0.13 μm</td>
<td>0 ~ 0.96</td>
<td>39</td>
<td>108</td>
<td>112</td>
<td>Y</td>
</tr>
<tr>
<td>This work</td>
<td>0.18 μm</td>
<td>3 ~ 10</td>
<td>250</td>
<td>190</td>
<td>27</td>
<td>Y</td>
</tr>
</tbody>
</table>

bandwidth and it is defined as

\[
FOM = \frac{\text{Energy Efficiency (pJ/Pulse)}}{\text{Bandwidth (GHz)}}
\]  

(3.13)

As we can see, this work has a good performance and is better than all the other works except [106]. But note that [106] used better technology than ours.

Figure 3.40: Measurement setup for the UWB impulse radio test.

The UWB impulse radio is further demonstrated using a DWG-based UWB transmitter [20] and two planar UWB monopole antennas [107], as shown in Fig. 3.40.
Fig. 3.41 shows the measurement results with different pulse rate and different transmitted waveforms. The first two in Fig. 3.41 are measured with the same pulse rate of 250 MHz while with different transmitted waveforms, and the third one is measured with a pulse rate of 150 MHz. Note that the DLL-based timing control block reserves an external control pad to tune the delay of each stage, which makes the timing control block work as a controlled delay line. At different pulse rate, by tuning the timing control block, synchronization between the received pulses and local template timing can be achieved. Note that the UWB antennas and LNA change the pulse waveforms. By setting the polarity and coefficient of each DPC tap according to the received pulse waveforms, the correct pulse detection can be obtained. The UWB impulse radio is also demonstrated with the modulated data (Fig. 3.42). Due to the limitation of the AWG, simple modulated data stream is utilized in this measurement. As shown in Fig. 3.42, the transmitted data can be successfully recovered at the receiver.
Figure 3.41: Measurement results of the transceiver with different pulse rate and different transmitted waveforms.

Figure 3.42: Measurement results of the transceiver with modulated data.
Chapter 4

Intra-Chip Free-Space Optical Interconnect Transceiver

4.1 Optical Transceiver Architectures

In this section, we will first review the conventional embedded-clock transceiver architecture, which is widely used in optical systems. After that, two relatively new transceiver architectures emerging in electrical serial-link communication systems will be reviewed: forwarded-clock and shared-clock architectures. We will compare these three transceiver architectures and discuss their advantages and disadvantages for the FSOI system.

4.1.1 Embedded-Clock Transceiver

Fig. 4.1 illustrates the conventional optical communication system [5, 108–110]. The transmitter composes of a laser driver (LD), an n:1 serializer (SER) and a phase-locked loop (PLL). The SER serializes the n-bit low-speed parallel outputs of the baseband and feeds the high-speed serialized result to the laser driver. The PLL
provides the required accurate clock frequencies for the SER. According to the data from the SER, the laser driver delivers the corresponding modulation current to the laser.

![Conventional optical communication system with embedded clock.](image)

Figure 4.1: Conventional optical communication system with embedded clock.

The laser converts the signals from the electrical form to the optical form. Those optical signals then transmit through the optical channel (such as optical fiber, optical waveguide, or free space) and are detected by the receiver at the far end. Before processed by the receiver, the optical signals are converted back to the electrical form by the photodiode (PD).

At the receiver side, a transimpedance amplifier (TIA) first converts the output current of the photodiode to a voltage with low noise, sufficient gain and enough bandwidth. A limiting amplifier (LA) then further amplifies the received signal. The amplified signal may exhibit substantial noise, therefore a decision circuit (DC) is used to detect the signal and generate a clean output. A deserializer (DES) performs the opposite operation of the SER. It deserializes the high-speed serial signal to n-bit low-speed parallel signals. Owing to the cost limitation, conventional optical transceiver typically has only one high-speed serial data link, and the clock information is embedded in this high-speed data transmission. Thus a clock recovery circuit (CRC) is needed to recover the frequency and phase information of clock from the received data and provide an optimal sampling phase for the decision circuit [111].
The clock recovery circuit and the decision circuit normally compose the clock and data recovery (CDR) circuit. The output frequencies of the CRC are also served as the accurate clock frequencies for the DES.

Due to the cost issue, the conventional optical system with embedded clock is a good choice for long-distance optical communication system since only one high-speed optical serial link is required. However, for a short-distance energy-efficient optical communication system (e.g., from several milimeter to several centimeter for the FSOI system), this transceiver architecture may not be a good option. First, the clock recovery circuit in the receiver is normally implemented based on PLL, which has high complexity and consumes large power. Secondly, different from those conventional optical systems, where a variable data rate is required [108,109], an FSOI system has a fixed data rate (e.g., 10 Gb/s in the FSOI system). Therefore frequency tracking and recovery as in the conventional optical system are not necessary, and the phase recovery is sufficient.

### 4.1.2 Forwarded-Clock Transceiver

Forwarded clock systems were presented in [112,113] for a short-distance electrical serial-link communication. In this system architecture, in addition to the typical data communication link, a clock communication link is also implemented. The transmitter generates the required clock and feeds this clock into the receiver. Therefore the clock frequency information is known in the receiver, and the clock frequency recovery is not required.

An optical system based on the forwarded clock is illustrated in Fig. 4.2. Compared to the clock frequency and phase recovery in optical system based on the embedded clock, this structure avoids the frequency tracking and recovery due to the presence of forwarded clock. Therefore simple phase recovery is sufficient, and the design complexity and power consumption on clock recovery are greatly reduced.
However, this additional clock transmission link brings other system hardware overhead. First, an additional optical communication link is needed, including a laser, a PD, and an optical channel. This inevitably brings extra hardware cost. Secondly, a driver is needed in the transmitter to drive the laser, and correspondingly, an amplifier is needed in the receiver to convert the PD output current to voltage form and amplify the voltage signal to satisfy the input level requirement of the phase recovery circuit. These two circuits further increase the hardware cost and power consumption. Therefore this forwarded-clock optical system is also ill-suited for FSOI system.

### 4.1.3 Shared-Clock Transceiver

Recently, electrical serial-link communication systems with shared clock were presented in [114,115]. In this system architecture, a reference clock feeds the transmitter and receiver simultaneously. This architecture is well-suited for the systems with multiple transceivers, such as the multi-channel or multi-core (as FSOI) systems. In these kinds of system, a global clock distribution network is normally available to provide the clock to each channel or core, therefore this global clock can be naturally used as the reference clock for the transceiver in each channel or each core, and no extra
hardware cost is required.

Fig. 4.3-(a) illustrates an optical system based on shared clock. Note that two PLLs are shown in Fig. 4.3-(a), i.e., one in the transmitter and one in the receiver. For the transceiver implementation in each processor core, however, since these two PLLs are identical (having same reference clock and generating same high-frequency clock), the PLL can be shared by the transmitter and the receiver (Fig. 4.3-(b)). Therefore only one PLL is needed in the transceiver in each processor core.

Compared to the embedded-clock optical system, where the PLL in the transmitter and the CRC in the receiver are different and can not be shared, the shared-clock optical system, due to the shared PLL, has a simpler transceiver architecture, lower design complexity and smaller chip area. Besides, since this architecture ensures same frequency for the transmitter and the receiver, no frequency recovery is needed in the receiver, and only phase recovery is required. In addition, due to the shared clock, no optical clock transmission link is needed between the transmitter and the receiver as the forwarded-clock optical system, therefore there is only one high-speed optical transmission link. Lastly, since the shared reference clock is in the electrical form, therefore the system needs no extra electro-optic (E/O) or opto-electronic (O/E) conversion, which saves great amount of hardware overhead and power consumption. Therefore, this system architecture overcomes the major drawbacks of the embedded-clock system and forwarded-clock system, and we will develop the FSOI system based on this shared-clock architecture.

4.2 Proposed FSOI transceiver Architecture

According to the review of different optical system architectures, we will develop our FSOI system based on shared clock architecture, which is well-suited for short-distance energy-efficient multi-core optical communication. As shown in Fig. 4.4,
Figure 4.3: (a) Optical communication system with shared clock; (b) Implementation of a shared-clock optical transceiver.

the transmitter and the receiver have a shared clock source. Two identical PLLs generate the same 10-GHz frequency for the transmitter and the receiver. Note that, as mentioned earlier, the PLLs can be shared by the transmitter and the receiver when implementing the transceiver in each core, and thus only one PLL is needed.
in the transceiver in each processor core. The 64:1 serializer converts the 64-bit low-speed parallel data into one high-speed 10-Gb/s data, while the 1:64 deserializer performs the opposite operation. According to the data from the serializer, the laser driver delivers the corresponding modulation current to the laser. In the receiver, the transimpedance amplifier (TIA) first converts the output current of the photodiode to a voltage, which is then further amplified by the limiting amplifier (LA). The amplified signal is detected by the decision circuit (DC) and a clean output is generated. The phase selector is used to generate the optimal phase to trigger the decision circuit and sample the data.

![Diagram of FOSI system with shared clock](image)

Figure 4.4: An FOSI system with shared clock.

To further simplify the FSOI system architecture and lower the power consumption, an injection-locked oscillator (ILO) can be used to replace the PLL as the clock generation block. Compared to a PLL, an ILO has the advantages of low design complexity, low power consumption, high stability and fast locking. A shared-clock FSOI system prototype with ILO-based clock generation is shown in Fig. 4.5. Note that serializer and deserializer are removed from this prototype to focus on the modification of the clock generation block. The transmitter includes a laser driver, an injection-locking oscillator (ILO), and a 7-bit pseudorandom binary sequence (PRBS) generator. With a 10-GHz reference clock, the ILO is locked to 10 GHz and triggers
the PRBS generator. The PRBS generator generates the 10-Gb/s data pattern for test purpose and feeds those data to the laser driver. The laser driver delivers corresponding modulation current to the VCSEL.

![Diagram of an FOSI system with ILO-Based clock generation and shared clock.](image)

Figure 4.5: An FOSI system with ILO-Based clock generation and shared clock.

The receiver consists of a passive transimpedance amplifier ($R_{TIA}$), a limiting amplifier, a decision circuit, an ILO, a phase selector. Instead of using the active TIA, a simple resistor-based passive TIA is used to provide the conversion from the photocurrent to the voltage. Compared to the active TIA, passive TIA consumes much less power. However, the resistor needs to be carefully chosen, since this simple resistor-based current to voltage conversion brings the design tradeoff between bandwidth, gain and noise [5,116]. The LA is ac coupled to the photodiode and the passive TIA, and further amplifies the received signal. The decision circuit detects the noisy signal and generates a clean output. The 10-GHz ILO generates the multi-phase 10-GHz clock for the decision circuit. The phase selector controlled by an external signal is used to choose the optimal phase to trigger the decision circuit and sample the data.
4.3 FSOI Transmitter Prototype Implementation

4.3.1 Laser Driver

The laser driver provides the modulation current to the VCSEL. As shown in Fig. 4.6, the laser driver consists of a pre-amplifier, a main driver, and a pre-emphasis block (which will be discussed shortly).

![Diagram of Laser Driver]

Figure 4.6: Laser Driver.

The pre-amplifier acts as the buffer of the preceding stage, and it provides the driving ability for the main driver and the pre-emphasis circuit. As illustrated in Fig. 4.7-(a), it is based on the CML topology for a high-speed operation.

As shown in Fig. 4.7-(b), the main driver generates the modulation current $I_{VCSEL}$ for the VCSEL, which is the difference between $I_{bias}$ and $I_{mod}$. The $I_{mod}$ generator is based on the transconductance amplifier (TCA). With high output impedance [117], a TCA can tolerate voltage variations due to the series resistance of VCSEL [118], therefore it is well-suited for current modulation. To make the current source $I_{bias}$ less susceptible to channel length modulation, instead of using conventional cascode current source suffering from large voltage headroom, an improved current source [119] with less sensitivity to the drain-source voltage is used here. Therefore high output impedance can be achieved with little voltage headroom penalty.
Figure 4.7: Laser driver circuit schematics (a) pre-amplifier; (b) main amplifier; (c) pre-emphasis.

Due to the multi-transverse mode structure of VCSEL and associated lateral carrier diffusion, VCSEL shows non-ideal transient behaviors, such as the slow rise/fall
time [26, 28, 120]. In order to compensate for the slow rise/fall time (especially the slower fall time) of the output optical signal, the pre-emphasis technique is used. It can speed up the rise/fall time and improve the eye diagram. The example waveforms of pre-emphasis is illustrated in Fig. 4.6. In addition to the modulation current, small current pulses appear on the rise and fall edges, which can speed up the rise and fall time. The implementation of pre-emphasis is shown in Fig. 4.7-(c), it includes a delay cell controlling the width of the current pulses, and two current pulses generation circuits (for rise/fall time pre-emphasis respectively). The current pulses are generated by AND/OR operation of the original signals $V_1$ and their delayed versions $V_2$.

Due to the process, voltage and temperature (PVT) variations, the real modulation current may deviate the designed value, therefore a tuning ability for the modulation current is preferred. In the driver implementation, the modulation current, pre-emphasis current, and the delay of the pre-emphasis are designed to be tunable to make the laser driver reconfigurable.

Figure 4.8: Simulation results of the laser driver with (a) pre-emphasis off; (b) pre-emphasis on.

Fig. 4.8 illustrates the simulated eye-diagram results of the laser driver at 10 Gb/s.
The results with pre-emphasis on/off are simulated respectively to show the impact of pre-emphasis. As shown in Fig. 4.8, with pre-emphasis on, the additional pre-emphasis current enhance and speed up the rising/falling edge, leading to a larger eye-opening for the eye-diagram. The laser driver consumes 24.5-mW power from a 1.5-V power supply, and occupies an area of 0.0033 mm².

4.3.2 Serializer

The outputs of the digital baseband are fed to the 64:1 serializer, which serializes the parallel low-speed input into high-speed serial output. As illustrated in Fig. 4.9-(a), the serializer is based on the binary tree architecture, and consists of a set of 2:1 multiplexer circuits. Each stage shares one clock and is controlled and synchronized by this clock.

Fig. 4.9-(b) shows the topology of the 2:1 multiplexer, which includes two latch-based D-flip-flops (DFF), one latch, and one selector. The flip-flops are used to align the incoming data with the clock. The extra latch in one of the two data path can avoid the simultaneous transition of two data stream, which eliminates the glitch issue [5]. There is a DFF-based retimer at the end of the serializer, it can reduce the data jitter arising from the intersymbol interference.

To reduce the power consumption, while the high-speed 2:1 stage is based on conventional CML topology (Fig. 4.10-(a)), the 64:2 stages with slower operation speed are implemented using static CMOS circuits which have no static power consumption (Fig. 4.10-(c)). The selector in the 2:1 multiplexer is also based CML topology for high-speed operation (Fig. 4.10-(b)). A CMOS-to-CML conversion circuit is added between the 64:2 and 2:1 stages. The retimer is also based on the CML topology due to its high-speed operation requirement. To further lower the power consumption, the clock chain in the serializer is shared with the divider chain in the phase-locked loop (PLL). In the simulation, the serializer consumes 75.3-mW power from a 1.5-V
Figure 4.9: (a) 1:64 serializer; (b) 2:1 multiplexer cell.
Figure 4.10: (a) Conventional CML-based latch; (b) CML-based selector; (c) CMOS digital latch.

power supply.

4.3.3 Clock Generator

The clock generator plays important roles in both transmitter and receiver. In the transmitter, the clock generator provides the clock frequencies for the serializer. In the receiver, in addition to the provision of the clock frequencies for the deserializer, the clock generator also provides the multi-phase clocks for the decision circuit. The
multi-phase selector chooses the optimal phase to trigger the decision circuit.

In our prototypes, two kinds of clock generators were used. One is based on conventional phase-locked loop, the other is an improved version, which is based on the injection-locked oscillator and provides very low power consumption. We will briefly discuss these two implementations.

**Phase-Locked Loop**

![Phase-Locked Loop](image)

Figure 4.11: Phase-Locked Loop.

The PLL-based clock generator is based on the conventional charge-pump PLL. As shown in Fig. 4.11, it consists of a phase-frequency detector (PFD), a charge-pump (CP), a second-order low-pass filter (LPF), a three-stage differential voltage-controlled oscillator (VCO), and a divider chain (/N). The PFD is based on conventional 3-state topology including two DFFs and one NAND gate. The DFFs are implemented using true single phase clocking (TSPC) circuits, which benefits for a high-speed operation and low power consumption [121]. The charge-pump has similar structure as [122]. It is optimized for less susceptible to charge-sharing and mismatch issues. The VCO is based on three-stage ring oscillator structure, the circuit developed in [123] is used here. It has a large linear tuning range and a constant output voltage swing over a large frequency range. The divider chain is shared with the serializer and deserializer to save power consumption. As illustrated in Fig. 4.9-(a) and Fig. 4.20-(a), the divider
chain is implemented using both CML and CMOS circuits for low power consumption.

**Injection-Locked Oscillator**

![Diagram of ILO topology, delay cell, and ring oscillator core](image)

Figure 4.12: (a) ILO topology; (b) Delay cell; (c) Ring oscillator core.

As mentioned above, in the transceiver architecture which is further improved for low power, an injection-locked oscillator (ILO) is used to replace the conventional PLL as the clock generator. The ILO is an open loop system compared to the closed
loop PLL, therefore it enjoys low power consumption, low design complexity, and it also features high stability and fast locking.

As shown in Fig. 4.12-(a), the ILO is based on a 3-stage ring oscillator. To enhance the injection and enlarge the locking range, instead of one port single phase injection, the ILO features multi-port multi-phase injection by using multiple delay cells to provide proper phase to each stage. Each delay cell is also based on CML topology to enhance the operation speed (Fig. 4.12-(b)). The Fig. 4.12-(c) shows the ring oscillator core. Similar as [124], the injection signal is generated by the lower differential pair and is injected as the differential current. The tuned capacitor can adjust the free-running frequency of the ring oscillator, leading to a large working range of the ILO. The simulated results show that at 10 GHz the ILO has a locked phase noise of -107.5 dBc/Hz at 1-MHz frequency offset. The ILO consumes 13.8-mW power from a 1.5-V power supply, which is much lower than the conventional PLL-based clock generator (65.5 mW). The ILO only occupies an area of 0.011 mm$^2$.

4.4 FSOI Receiver Prototype Implementation

4.4.1 Transimpedance Amplifier

Active Transimpedance Amplifier

As the first stage of the receiver, the transimpedance amplifier (TIA) converts the photocurrent to the voltage with sufficient gain and bandwidth. Due to the large parasitic capacitance of the photodetector, the bandwidth of a TIA is normally dominated by the input pole, which is determined by the parasitic capacitance of the photodetector and the input impedance of the TIA. In order to extend the bandwidth of the TIA to accommodate a high-speed operation, either the photodetector parasitic capacitance or the TIA input impedance needs to be reduced. Here we focus on the
latter approach.

The regulated cascode (RGC) common-gate input stage [125–127] were proposed as the input stage of the TIA to provide a low input impedance. As illustrated in Fig. 4.13, with the local feedback, the equivalent transconductance of the common-gate amplifier increases. Therefore the input impedance of the common-gate amplifier, which is inversely proportional to the equivalent transconductance, effectively lowers, and the input pole is pushed to a higher frequency and has little impact on the bandwidth of the TIA. In another word, the TIA with RGC input stage can tolerate a large photodetector parasitic capacitance.

Figure 4.13: Transimpedance amplifier.

In addition, the RGC input stage also improves the isolation of the main amplifier and its resistor from the photodetector parasitic capacitance. The improved isolation eliminates the usage of a dummy input capacitor, which enables an easy implementation of single-ended to differential conversion [126].

The main amplifier is based on CML topology with bandwidth extension. In order to achieve the minimum chip area, the active inductive peaking technique, instead
of conventional inductive peaking technique using passive inductors, is used for the bandwidth enhancement. However, as mentioned earlier, the active inductive peaking technique suffers from the large voltage headroom, therefore a voltage higher than the supply voltage is used to bias the active inductors.

As illustrated in Fig. 4.13, the TIA also employs a tunable offset compensation block at the input of the TIA to compensate the dc offset arising from the dc current and dark current. The transistors’ size of the offset compensation block are optimized to minimize the impact of the parasitic capacitance on the TIA input.

![Simulated transimpedance gain of the active transimpedance amplifier.](image)

The simulated transimpedance gain of the TIA is shown in Fig. 4.14. The TIA has a transimpedance gain of 52 dBOhm, a 3-dB bandwidth of 8.0 GHz, which is good for the 10-GHz data-rate high-speed communication. The TIA consumes 7.9-mW power from a 1.5-V power supply, and occupies an area of 0.0016 \( \text{mm}^2 \).

**Passive Transimpedance Amplifier**

In the receiver with ILO-based clock generation, a simple resistor-based passive TIA is used. Compared to the active TIA, the passive TIA benefits for a simpler
structure and small power consumption, which further lowers the power consumption of the whole system. This passive implementation, however, suffers from direct trade-off between gain and bandwidth. The transimpedance gain is equal to the resistor value, and the bandwidth is inversely proportional to the product of the resistor and the parasitic capacitance of photodetector. Therefore a passive TIA with high gain tends to have a small bandwidth. Besides, since the integrated input referred current noise of a passive TIA is also inversely proportional to the resistor [116], a large-bandwidth passive TIA also suffers from bad noise performance. Based on these considerations, the passive TIA needs careful design to choose an optimal resistor value for a balanced performance of the gain, bandwidth and noise.

4.4.2 Limiting Amplifier

![Diagram of Limiting Amplifier](image)

Figure 4.15: Limiting amplifier topology.

The limiting amplifier further amplifies the output signal of the TIA to satisfy the input sensitivity requirement of the data recovery circuit (i.e., decision circuit). Fig. 4.15 illustrates the topology of the limiting amplifier. It consists of one input stage, four gain stages and one dc offset compensation block. The dc offset compensation circuit is added to compensate the dc offset caused by the device mismatch and the low frequency noise source, it can reduce the low-frequency gain and therefore minimize the dc offset contribution to the output [45, 48]. Thus, the limiting amplifier shows a band-pass transfer characteristic. The offset compensation is based on
R-C filter, and feeds the output back to the input stage, which sums the TIA output and the feedback signal (Fig. 4.16-a). To minimize the chip area, the R and C are implemented by NMOS transistors.

![Figure 4.16: Limiting amplifier: (a) input stage; (b) gain stage.](image)

The gain stage is illustrated in Fig. 4.16-b. Same as the active TIA, active inductive peaking technique, instead of passive inductive peaking technique, is used here.
for the bandwidth enhancement with minimum chip area. As mentioned earlier, the bandwidth of a cascaded system reduces with the increase of the cascaded stages, where assuming each stage has the same the bandwidth. Therefore, as a cascaded system, in order to achieve a large system bandwidth, the limiting amplifier is required to have a larger bandwidth than the TIA. To further boost the bandwidth, we employ active feedback technique mentioned in Chapter 2 to assist the active inductive peaking technique [43,47].

The simulated voltage gain of 4-stage limiting amplifier is shown in Fig. 4.17. The limiting amplifier has a voltage gain of 34 dB, 3-dB bandwidth of 9.1 GHz, and lower-corner cut-off frequency of 0.1 MHz. The limiting amplifier consumes 21-mW power from a 1.5-V power supply, and occupies an area of 0.013 $mm^2$.

![Figure 4.17: Simulated voltage gain of the limiting amplifier.](image)

4.4.3 Decision Circuit

Due to the substantial noise existing in the received data, a decision circuit is placed between the limiting amplifier and the deserializer to sample the data and provide a clean output data for the deserializer. As illustrated in Fig. 4.18-(a), the
The decision circuit is based on a D-flip-flop (DFF) consisting of two latches.

Figure 4.18: (a) DFF-based decision circuit; (b) Modified CML-latch used in the decision circuit (where the tail transistor in conventional CML-latch is removed).

The latches are based on the CML topology for high-speed operation [5]. Since the signal coming out of the limiting amplifier has lower common-mode voltage, the CML circuits in the decision circuit need to take this into account. Different from the conventional CML circuits having a tail transistor to act as a current source (Fig. 4.10-(a)), the CML circuits here remove this tail transistor (Fig. 4.18-(b)). Thereby, the minimum input common-mode voltage for this new CML circuits is only the gate-source voltage of the lower transistors, while for the conventional CML circuits, they need to count the drain-source voltage of the tail transistor and the gate-source voltage of the transistor in the differential pair (Fig. 4.10-(a)).

Although the signal coming into the decision circuit is already amplified by limiting amplifier, sensitivity is still a design consideration in decision circuit design. High sensitivity enables the decision circuit recover small amplitude input signal, which benefits a larger communication range between the transmitter and the receiver. Fig. 4.19-(a) shows the time-domain simulation results of the decision circuit at 10-Gb/s data rate. The input signal with 100-mW swing can be successfully recovered and a output signal with larger than 600-mW swing is achieved, which is
sufficient for the processing of the following stage. As shown in Fig. 4.19-(b), in order to generate a 0.6-V output swing, only 60-mV differential input voltage is needed. The decision circuit consumes 13.8-mW power from a 1.5-V power supply, and occupies an area of 0.0009 $mm^2$.

Figure 4.19: Simulated results of the decision circuit: (a) Time-domain waveforms; (b) Input sensitivity of the decision circuit.
4.4.4 Phase Selector

The three-stage VCO (in the phase-locked loop or injection-locked oscillator) generates three differential outputs with equal-spaced phases. The phase selector, consisting of three CML-based selectors (Fig. 4.10-(b)), interpolates three additional phases and generates six differential phases with equal-space. According to the external control signals, one of them will be selected as the optimal phase to trigger the decision circuit and sample the data. In the prototype, the optimal phase selection is a manual process by sweeping all six phases. It is an open-loop process, however it can be implemented as an automatic close-loop process with the help of baseband processor in future design. The phase selector consumes 5.1-mW power from a 1.5-V power supply, and occupies an area of 0.001 mm\(^2\).

4.4.5 Deserializer

The deserializer performs the opposite operation of the serializer. It deserializes the high-speed serial data into low-speed parallel data. As illustrated in Fig. 4.20-(a), it is also based on binary tree structure, and each cell uses a 1:2 demultiplexer. Each stage shares one clock, and is controlled and synchronized by this clock. Each 1:2 demultiplexer consists of two flip-flops and one latch (Fig. 4.20-(b)). The two flip-flops operate at opposite edge of the clock to separate one data stream into two data streams. The extra latch in one path aligns the two outputs for the next stage.

Same as the serializer, in order to save the power consumption, the low-speed 2:64 stages are implemented using static CMOS circuits (Fig. 4.10-(c)), and the high-speed 1:2 stage is implemented in conventional CML topology (Fig. 4.10-(a)). A CML to CMOS conversion circuit is added between the two kinds of circuit. The clock chain is also shared with the divider chain in the PLL to save power consumption. In the simulation, the deserializer consumes 49.9-mW power from a 1.5-V power supply.
Figure 4.20: (a) 1:64 deserializer; (b) 1:2 demultiplexer cell.
To verify the function of both serializer and deserializer, the simulation is done by connecting the deserializer outputs to the serializer inputs and using a 10-Gb/s random data pattern as the input of the deserializer. Fig. 4.21 shows the waveforms of the critical high-speed nodes. Note that as mentioned above, the first stage of the serializer/deserializer is based on CML topology, therefore it has smaller swings. From the second stage, the serializer/deserializer is based on CMOS circuits, therefore rail-to-rail signals are observed.

4.5 Measurement and simulation Results

The FSOI transceiver prototypes were designed using Cadence Spectre circuit simulator, and fabricated in a 0.13-μm standard digital CMOS technology. The chip micrograph is shown in Fig. 5.15-(a), and the layout of the transceiver is shown in Fig. 5.15-(b). Several transceiver prototypes were implemented in this chip. The total die size is 5 mm × 5 mm. The active area of the PLL-based transmitter (including 64:1 serializer, laser driver and PLL) is 0.55 mm², and the active area of the ILO-based transmitter (including laser driver and ILO) is 0.025 mm². For the receiver side, the active area of PLL-based receiver (including active TIA, limiting amplifier, decision circuit, phase selector, PLL and 1:64 deserializer) is 0.44 mm², and the active area of ILO-based receiver (including passive TIA, limiting amplifier, decision circuit, phase selector and ILO) is 0.053 mm².

Fig. 4.23 shows the simulated results of the ILO-based transceiver at the data rate of 10 Gb/s. The waveforms of the important nodes are shown, including PRBS generator output, ILO output, laser driver output, PD output, passive TIA output, limiting amplifier output and decision circuit output. As shown in Fig. 4.23, the laser driver employs the pre-emphasis and improves the rising/falling edge. On the receive side, the receiver chain can well amplify the small PD output current, and
Figure 4.21: Simulated results of the serializer and deserializer.
Figure 4.22: (a) Chip photo of the optical transceivers; (b) Layout of optical transceivers.

successfully recover the 10-Gb/s data. The laser driver, limiting amplifier, decision circuit, phase selector, and ILO consume 24.5 mW, 20.8 mW, 6.7 mW, 5.1 mW and 13.8 mW from a 1.5-V power supply, respectively, which shows the whole transceiver has an energy-efficiency of 7.1 pJ/b at 10 Gb/s.
Figure 4.23: Simulated results of (a) transmitter; (b) Receiver.
After normal fabrication in a 0.13-μm standard digital CMOS technology, the optical transceiver chip will be post-processed and be made into a 3-D chip with the microprocessor fabricated in another chip. The post-processing for the 3-D chip is still going on, and an intermediate version is available for measurement. Due to the limitation of the available chip, only stand-alone laser driver and receiver front-end (TIA and limiting amplifier) can be measured. Here frequency-domain measurements were performed to characterize the voltage gain of the building blocks mentioned above. Note that for the receiver, transimpedance gain is our focus, however, voltage gain is used for initial characterization due to the test limitation. Similarly, for the laser driver, time-domain performance is our focus, which can characterize the pre-emphasis performance, however, due to the lack of data pattern generator, same as the receiver front-end, frequency-domain voltage gain test is used for initial characterization.

Fig. 4.24 shows the measurement and simulation results of the voltage gain of the laser driver and receiver front-end (including the passive TIA version and the active TIA version). The measured bandwidth of the laser driver, passive TIA and limiting amplifier, active TIA and limiting amplifier are 3.5 GHz, 3.3 GHz, 3.6 GHz, respectively. As shown in Fig. 4.24, the measurement results have considerable bandwidth degradation compared to the simulation results. The reason is mainly due to the back metal added during the post-process in this intermediate version chip. As shown in Fig. 4.25-(a), there are mainly on-chip transmission lines, through-silicon via (TSV) and back metal from the internal circuit input/output to actual external input/output pad. Due to the limitation on layout floor plan and post-processing for 3-D chip, the circuit input/output first go through long M5 routings (about 2 mm on average) to the on-chip pads, then go through the vias and TSV to the back side of the wafer, then go through the long back metal (about 1 mm on average) to the actual input/output pads which are used for test purpose. The back metal is titanium-tungsten-titanium, and there is a 1-μm silicon-dioxide layer between the
Figure 4.24: Measurement and simulation results of: (a) Laser driver; (b) Receiver front-end (passive TIA and limiting amplifier); (c) Receiver front-end (active TIA and limiting amplifier);
back metal and the silicon substrate. The 1-μm thin layer is the major bandwidth limitation factor.

Here we use a low-pass filter with 8-GHz bandwidth and 50 Ω impedance to characterize the effect of the transmission line, TSV and back metal. As shown in Fig. 4.25-(b)), by carefully designing the transmission lines, we can minimize the bandwidth degradation introduced by the transmission lines. However, there is no control on TSV and back metal, since they are determined by the process. As we can see, the TSV and the back metal introduce larger degradation on bandwidth compared to the transmission line, especially the back metal which becomes the dominant bandwidth limitation factor. As we mentioned above, this is mainly caused by the thin dielectric layer between back metal and substrate. With all these factors included, the whole bandwidth is degraded to 3.4 GHz from the 8-GHz bandwidth of the stand-alone low-pass filter, which well matches the measurement results.

4.5.1 Discussions

Receiver with latched sampler

In order to further simplify the transceiver architecture, latched sampler can be used in the receiver. Owing to the superior sensitivity and higher immunity to power supply noise, latch-based sampler has been emerging in serial-link transceivers as the data sampler [10,128]. An FSOI system front end with latched sampler is shown in Fig. 4.26, where the conventional limiting amplifier and decision circuit are replaced by the latched sampler. The latched sampler will make the system architecture much simpler and lower power consumption. In addition, due to the embedded deskew capability of ILO [129,130], it is possible to use this phase-tuning capability to perform the phase recovery and therefore eliminate the dedicated phase selector block, which can further improve the system energy-efficiency.
Figure 4.25: (a) Back metal illustration (not to scale, dielectric not shown); (b) Simulation results of effect on bandwidth of on-chip transmission lines, TSV and back metal.
A receiver employing latched samplers will be investigated and implemented in next Chapter, where latched samplers are used in the receiver for the on-chip interconnect system for multi-core processors to effectively simplify the receiver architecture and reduce the overall power consumption.

![An FSOI system with latched sampler.](image)

**Figure 4.26:** An FSOI system with latched sampler.

**Multi-channel system**

As mentioned earlier, in order to achieve a high data rate, we can use advanced CMOS technologies or advanced bandwidth enhancement techniques. An alternative way is to use multi-channel to transmit the data [26, 28, 131, 132]. It is a very attractive approach, because compared to single-channel communication, multi-channel communication can provide the benefit of much higher data rate with the same CMOS technology and same bandwidth extension techniques. However, by using multiple channels, the power dissipation and chip area inevitably increase by multiple times. Therefore, in order to resolve this issue, energy-efficient architecture for multi-channel FSOI system is desirable.

A single-channel FSOI with ILO-based clock generation and shared clock is an energy-efficient architecture. The multi-channel FSOI system is preferable to maintain these two features. In order to further reduce the power dissipation for a multi-channel FSOI system, some modifications on the system architecture are needed. A 4-channel FSOI system with energy efficient architecture is illustrated in Fig. 4.27.
First, only one ILO-based clock generation block will be used in the transmitter side, and it will be shared by all the transmitters. Compared to the duplication of a single-channel with its ILO-based clock generation block by four times, a 4-channel FSOI system with shared ILO-based clock generation not only eliminates the power dissipation on three ILO-based clock generation blocks, but also ensure a synchronized operation for all four channels. Second, the embedded deskew capability of ILO can be used to perform the phase recovery and therefore eliminate the dedicated phase selector block, which can further improve the system energy-efficiency. Different from transmitter side, however, due to different channel responses, 4 ILOs are required to recover the phase of each channel.

A multi-channel system will be investigated and implemented in next Chapter, where an on-chip interconnect system for multi-core processors will employ multiple communication channel to effectively increase the throughput.
Chapter 5

On-Chip Interconnects for Multi-Core Processors

5.1 Transmission-Line-Based On-Chip Interconnects

As mentioned above, transmission line has been proposed for on-chip interconnects to achieve wide bandwidth and low latency. Different transceiver implementations have been proposed for high-speed communication. In [35], the digital data was modulated with high frequency carrier, which is similar as the conventional narrowband RF communication. As shown in Fig. 5.1-(a), the baseband digital data is up-converted to high frequency by a mixer in transmitter. In the receiver, another mixer is used to down-convert the received high-frequency data back to baseband. By up-converting and transmitting the data in high frequency, such a topology enables the data transmission in LC region of the interconnect where the inductive component of the wire dominates instead of the resistive component. This benefits from the wave nature of the interconnect and makes near speed-of-light data transmission possible [35]. However, the additional up/down conversion and local oscillator (LO) suffer from significant power consumption and more design complexity. In addition,
due to the narrowband operation, the data rate is limited. For example, in [35], a 7.5 GHz LO is used as the carrier, while the data rate is only 1 Gb/s over a 20-mm transmission line.

![Diagram](image)

(a)

(b)

(c)

Figure 5.1: Different transmission-line-based on-chip interconnect transceiver implementation: (a) Carrier-based; (b) Carrier-based with multiple access technique (FDMA is used for illustration); (c) Direct transmission.

In [133–135], baseband digital data is also modulated to high frequency to transmit so that the wire can be treated as transmission line and the latency performance can
be improved. As shown in Fig. 5.1-(b), multiple access techniques, such as frequency-
division multiple access (FDMA) and code-division multiple access (CDMA), are used
to increase the throughput of the system. However, this large throughput is at the
expense of increased power consumption and system complexity. For example, as
shown in Fig. 5.1-(b), multiple LOs at different frequencies are required for FDMA-
based interconnect system.

In [37], instead of using high frequency carrier to modulate the data, the data
is directly transmitted through the transmission line. As shown in Fig. 5.1-(c), this
topology features very simple system architecture. The wideband feature of the direct
transmission potentially enables communication with higher data rate. For example,
in [37], an 8-Gb/s data rate was achieved across a 5-mm long transmission line.

As discussed above, direct transmission has the simplest architecture while offers
high data rate. This topology, however, requires a wideband operation of the trans-
mitter and the receiver, posing design difficulty on these front-end circuits, especially
with the increasing data rate. Besides, this topology also needs to address the prob-
lem of how to energy-efficiently generate the high speed data from the baseband,
and similarly, how to process the high speed data in the receiver. This essentially
requires a well-designed serializer and deserializer. However, this issue was not ad-
dressed in [37]. In addition, high throughput is highly demanded. As mentioned in
Chapter 4, multi-channel is an effective approach to increase the throughput, which
was also not covered in [37]. To increase the flexibility of the interconnect system,
both point-to-point and broadcasting communications are also highly desired. In this
thesis, we will investigate a high-speed on-chip interconnect system addressing the
above considerations. We will optimize the system performance through both system
architecture design and circuits design.
5.2 Proposed Interconnect System

The presented interconnect system has been described in Chapter I. There are several important questions that need to be answered for this new interconnect system: 1) Does it really work in real silicon? 2) Can it support the high data rate to satisfy the bandwidth density requirement? 3) Can it tolerate cross-talk between the adjacent transmission lines when running at such high speed?

To demonstrate the feasibility and performance of this interconnect system, a chip prototype was designed and implemented using a 130-nm SiGe BiCMOS technology, which is the fastest technology available to us. Similar or better performance can be achieved using state-of-the-art CMOS technologies. As shown in Fig. 5.2, the prototype consists of two 20-mm, parallel-running, differential transmission lines (only one transmission line is shown in Fig. 5.2), and eight communication nodes distributed along the two transmission lines. Node 1 to 6 in Fig. 1.6 are equally distributed on one of the transmission lines. Another two communication nodes are placed on the other transmission line (not shown), which are used to characterize the cross-talk performance. Based on the transistor and transmission line performance in this technology, a target data rate of 25 Gb/s was chosen.

Figure 5.2: Interconnect prototype architecture.
5.3 Transmission Line Implementation

![Patterned ground shields](image)

Figure 5.3: Parallel-running differential coplanar waveguides with a Ground-Signal-Signal-Ground (GSSG) configuration.

The transmission lines are differential coplanar waveguides (CPW) in a Ground-Signal-Signal-Ground (GSSG) configuration (Fig. 5.3), which allows the fully differential transceiver design and reduce crosstalk between adjacent transmission lines. The transmission lines utilize patterned ground shields [136] to achieve a larger characteristic impedance and a smaller attenuation. Based on electromagnetic simulations using Sonnet, the optimized differential CPW has 6-μm signal line width, 17-μm signal-to-signal line spacing, 20-μm signal-to-ground line spacing, and 75-μm total width. The simulation shows it has an attenuation of 0.39 dB/mm at 25 GHz, and a 3-dB bandwidth of 7.9 GHz at 20-mm distance (Fig. 5.4). Therefore, additional circuit techniques are needed to compensate the transmission line loss and achieve a higher data rate.

5.4 Transmitter Prototype Implementation

The transmitter for the interconnect system is shown in Fig. 5.5, which consists of a front-end driver and a 2:1 serializer (SER). As shown in Fig. 5.6-(a), the 2:1
Figure 5.4: Simulated transmission line attenuation with different lengths.

Figure 5.5: Architecture of the transmitter for the on-chip interconnect system.

Serializer is based on tree structure [5]. It converts the lower speed parallel data into the higher speed serial data. The selector, depending on the clock signal, switches on/off the corresponding switch and selects either \( V_1 \) or \( V_2 \) to the output (Fig. 5.6-(b)). The BiCMOS latch, similar as the CMOS latch, is made of the sensing stage to sense the input, and a regenerative-pair-based storing stage to further amplify the difference from the sensing stage (Fig. 5.6-(c)). But different from the CMOS latch, emitter followers are needed in the storing stage to satisfy the DC operation point requirement of bipolar transistors.
Figure 5.6: (a) Topology of the transmitter multiplexer; (b) Schematic of the selector; (c) Schematic of the latch.
The selector and latches are based on current-mode logic (CML) topology in order to achieve high-speed operation. The differential pairs in the signal path use bipolar transistors, while the clocks are processed by MOSFETs. The BiCMOS topology benefits from both high gain of bipolar transistors and good switching characteristics of MOSFETs [137]. Tail transistors, which serve as current source in typical CML circuits, are removed in the selector and latches to reduce the input common-mode voltages by eliminating the voltage drop introduced by the tail transistors. By doing this, it becomes possible to lower the supply voltage to 2 V, helping to reduce the overall power consumption. However, without the tail current source, the sizing of the MOSFETs and the swing of the clocks need to be well controlled to satisfy the current requirement of these circuits.

The schematic of driver is shown in Fig. 5.7-(a), where the equalization technique is applied to partly compensate the frequency dependent loss of the transmission line. The high-pass-filter-based pre-emphasis generates an overshoot at each rising/falling edge of the transmitted data (Fig. 5.7-(b)), which effectively accelerates the rising/falling time, thus improves the eye-opening of the eye-diagram. The output stage is based on an emitter follower and a common-emitter amplifier, improving the driving capability and the current efficiency.

5.5 Receiver Prototype Implementation

Fig. 5.8 illustrates the architecture of the receiver. A pre-amplifier, which is based on a differential pair, is used in the receiver front-end to first amplify the weak incoming signals (Fig. 5.9).

As shown in Fig. 5.8, instead of using more amplifiers to do further amplification and D-flip-flop-based decision circuit to sample the data, latched samplers, featuring superior sensitivity and higher immunity to power supply noise and recently being
Figure 5.7: (a) Schematic of the transmitter driver; (b) Illustration of pre-emphasis scheme.

Figure 5.8: Architecture of the receiver for the on-chip interconnect system.
used in serial-link transceivers as the data sampler [10,128], are used here to sense the input signals with small amplitude and generate the large-swing outputs. The latched samplers share the same topology as the latch in the serializer (Fig. 5.6-(c)), but are optimized for higher sensitivity. Fig. 5.10 illustrates the simulated input sensitivity of the latched sampler at the data rate of 25 Gb/s. In order to generate a 0.5-V output swing, which is normally sufficient to ensure a successful detection by the following stage, only 18-mV differential input voltage is needed. Such a latched sampler with high sensitivity reduces the requirements on the front-end amplifiers, hence simplifying the receiver architecture and saving chip area and power consumption.

To further simplify the system architecture, two latched samplers are used here and they are triggered by half-rate complementary clocks [10]. By doing this, in addition to sample the data, these high-speed latched samplers can also perform the function of the 1:2 deserialization, which typically requires the highest operation speed and largest power consumption in a deserializer. The combination of data sampling and 1:2 deserialization enables a receiver with simple architecture and low power consumption.
5.6 Isolation

Since the interconnect system supports both point-to-point and broadcasting communication, isolation of the non-communicating nodes from the transmission line is very important. Isolation switches, which is similar as the T/R switch in wireless transceiver [38, 39], are typically inserted between the transmitter/receiver and the transmission line [33,34]. The insertion loss when the switch is on and isolation when the switch is off are two important characteristics for an isolation switch. Fig. 5.11 shows the simulated insertion loss and isolation of a typical isolation switch. As we can see, the insertion loss and isolation degrade with the increasing frequency. What is worse is that there is strong design trade-off between the insertion loss and isolation. To achieve a small insertion loss, a switch with larger size is desirable, however, this costs large silicon area and most importantly introduces more parasitic capacitance which degrades the isolation performance. Therefore, in order to achieve reasonable isolation, those switches inevitably introduce extra loss to the link when they are switched on, especially at higher frequency. A pair of such switches, one on
transmitter side and one on receiver side, is normally required for isolation purpose, therefore considerable loss will be introduced to the transmission link.

![Simulated insertion loss and isolation](image)

**Figure 5.11:** Simulated insertion loss and isolation of an isolation switch.

To eliminate the isolation switches and thus the insertion loss, it’s preferable not to use the dedicated switches. Therefore the transmitter and receiver are designed to be internally switched on/off, which introduces no additional link loss. In the driver, by turning off the bias current and the isolation switch, the driver switches off and isolates the corresponding transmitter from the transmission line. The simulated isolation performance is shown in Fig. 5.12, where a minimum isolation of -47 dB is obtained at 25 GHz.

Same as driver, there is no dedicated isolation switch inserted between pre-amplifier and transmission line, therefore the receiver avoids suffering from the insertion loss caused by conventional isolation switch. By turning off the bias current, the pre-amplifier switches off and isolates the corresponding receiver from the transmission line. The simulated isolation performance is shown in Fig. 5.13, where a minimum isolation of -29 dB is obtained at 25 GHz.
5.7 Input/Output Impedance

As mentioned above, the communication nodes are directly connected to the transmission line without using the isolation switches, therefore the impedances of the communication nodes are very important for the point-to-point and broadcasting
communication in the interconnect system. For the point-to-point communication, if the two communication nodes are far away, since the transmission line already introduces loss, the additional degradation on signal caused by impedance mismatch of the middle nodes is unwanted. Similarly, for broadcasting communication, this kind of degradation is also unwanted.

According to Fig. 5.14, where $Z_0$ is the characteristic impedance of the transmission line and $Z$ is the impedance of the communication node, if $Z$ is equal to $Z_0$, then the equivalent impedance $Z_e$ looking from the right side of the node B is $Z_0/2$, which causes impedance mismatch with the transmission line at the left side of the node B. Similar situation happens at node A if $Z$ is equal to $Z_0$. Therefore the whole interconnect system will have impedance mismatch at each communication node, leading to reflection and thus inter-symbol interference (ISI) in random data [5] and degradation on voltage swing of each communication node. On the other hand, if $Z$ is much larger than $Z_0$, then the equivalent impedance $Z_e$ will remain close to $Z_0$, which greatly mitigates the impedance mismatch issue for the whole interconnect system. Therefore, large impedance is desirable for the communication nodes. The simulation shows that if a communication node is added onto the transmission line, an impedance of 300 $\Omega$ is required for the communication node to ensure a variation of voltage swing within 5 % compared to the case without that communication node.

![Figure 5.14: Transmission line with multiple communication nodes.](image-url)
For the receiver, due to the feature of supporting both point-to-point and broadcast communication, the system can have only one receiver turning on or have multiple receivers turning on. Therefore, the input impedance with the receiver’s first stage (pre-amplifier) turning on and turning off are both important. The simulations show the pre-amplifier has a minimum input impedance of 375 Ω and 1171 Ω at 25 GHz with the pre-amplifier turning on and off, respectively, which satisfies the impedance requirement.

On the other hand, for the transmitter, there is always only one transmitter turning on each time in this interconnect system, therefore only the output impedance of the driver when turning off is required to satisfy the impedance requirement mentioned above. The simulations show the driver has a minimum output impedance of 552 Ω at 25 GHz with the driver turning off, which satisfies the impedance requirement. Note that the output impedance requirement of the driver when turning on has different design consideration. Since it needs to drive the transmission line, a low output impedance is preferable.

5.8 Measurement Results

The interconnect prototype was designed using Cadence Spectre circuit simulator, and fabricated in IBM 130-nm SiGe BiCMOS (8HP) technology. As shown in Fig. 5.15, the prototype consists of two parallel-running transmission lines. As mentioned above, they are implemented using differential coplanar waveguides with a Ground-Signal-Signal-Ground (GSSG) configuration, which allows good noise rejection and less crosstalk between adjacent transmission lines. They are placed in a meandering fashion to maximize the total length to 20 mm in the 4 mm × 4 mm chip. Two transmitters (node 1 and 4) and four receivers (node 2, 3, 5 and 6) are placed at different locations on the same transmission line for transmission test, and
another two (node 7 and 8) on the other transmission line to measure crosstalk. The whole chip including the test circuits is 4 mm × 4 mm. The active area of the transmitter and receiver only occupy 0.023 mm² and 0.025 mm², respectively. Note that stand-alone transmitter driver and receiver pre-amplifier circuits were also fabricated to characterize their performance.

Fig. 5.16 shows the measured frequency response of the stand-alone transmitter driver and receiver pre-amplifier. Note that due to the limitation of measurement instruments, the frequency response was measured up to 20 GHz. Within the 20 GHz frequency range, the driver and pre-amplifier both show good match with the simulation results and have gain reduction less than 2 dB, showing the 3-dB bandwidth larger than 20 GHz.

To characterize the interconnect system, 2⁷-1 PRBS generators based on linear feedback shift registers [138] were designed and implemented on the same chip. Two PRBS generators are connected to the transmitter as the inputs of the 2:1 serializer for
Figure 5.16: Measured and simulated voltage gain: (a) Transmitter driver; (b) Receiver pre-amplifier.

The testing purpose. The test setup of the whole interconnect system characterization is shown in Fig. 5.17. Two signal sources, synchronized using 10-MHz internal clock,
are used in the measurement. One provides the clock signal for the transmitter and receiver, the other is used as the trigger for the sampling oscilloscope. The oscilloscope is used to observe the results of the receivers. Note that the time domain waveforms observed on the oscilloscope, especially the ones with higher data rate, are worsened by the frequency dependent loss of the measurement components, e.g. cables and connectors.

![Figure 5.17: Test setup for the prototype chip characterization.](image)

To characterize the interconnect system, we first measure the point-to-point communication, we will check the pre-amplifier output and verify the effect of the pre-emphasis. We then measure the latched sampler outputs to verify the data transmission and recovery at 25 Gb/s (including 2:1 serialization and 1:2 deserialization). After that, we then verify the broadcasting communication. Finally we will use two transmission lines to characterize the crosstalk of this multi-channel interconnect system.

Fig. 5.18 shows the pre-amplifier output of Rx5 at 25.4 Gb/s with Tx1 as the transmission source, where two 12.7-Gb/s PRBS data streams generate the 25.4-Gb/s combined one. The communication distance from Tx1 to Rx5 is 16 mm. As we can see, the measured results well match the simulated results. Note that, as
we mentioned earlier, the time domain waveforms observed on the oscilloscope are worsened by the frequency dependent loss of the measurement components.

![Simulated Rx5 Pre-Amplifier Output @ 25.4Gb/s](image1)

![Measured Rx5 Pre-Amplifier Output @ 25.4Gb/s](image2)

Figure 5.18: Measured and simulated outputs of Rx5 pre-amplifier at the data rate of 25.4 Gb/s.

Fig. 5.19 shows the latch outputs of Rx6, a 20-mm transmission distance from Tx1. The transmitted data rate is 25.4 Gb/s. Thus the 1:2 deserialization was also performed at 25.4 Gb/s, and two 12.7-Gb/s data streams were generated by the 1:2 deserializer. The simulated Tx1 inputs are also shown in Fig. 5.19. As we can see, the measured Rx6 latch outputs well match the simulated Tx1 inputs, which demonstrates the successful data transmission and recovery, including the 2:1 serialization and 1:2 deserialization.

The measured eye-diagram of the Rx6 latch outputs at different data rate (25.4 Gb/s and 15.24 Gb/s) can be found in Fig. 5.20. As we can see, on one hand, the latched samplers triggered by half-rate complementary clocks perform the 1:2 deserialization. On the other hand, they also act as the decision circuit in the optical receiver, which can sample the data and generate clean outputs. As mentioned above, these features make the receiver architecture much simpler and greatly reduce the
Figure 5.19: Measured outputs of Rx6 latches with simulated inputs of Tx1 serializer at the data rate of 25.4 Gb/s.

Figure 5.20: Measured eye diagrams at the Rx6 outputs.
power consumption. The bit error rate (BER) is also measured at the highest data rate (25.4 Gb/s). A BER less than $10^{-12}$ can be achieved.

Fig. 5.21 and Fig. 5.22 show the measured eye diagrams with driver’s pre-emphasis on or off at the amplifier output of Rx 5 and Rx 2, respectively. The eye diagrams are measured at data rates of 10.16 Gb/s, 15.24 Gb/s, 20.32 Gb/s and 25.4 Gb/s. As shown in Fig. 5.21 and Fig. 5.22, due to the shorter communication distance and thus the less transmission line attenuation, Rx2 has better eye-diagram than Rx5 at the same data rate. However, as we can see, with the pre-emphasis on, the eye-diagram of both Rx2 and Rx5 show significant improvement, especially the Rx5’s eye-diagram at higher data rate.

As mentioned above, the interconnect system supports both point-to-point and broadcasting communications. Fig. 5.23 shows the measured broadcasting communication of Tx1. For the receiver we measured, we first measured its output with other receivers turning off. Then we turned on more and more receivers, and still measured that receiver’s output. For example, as shown in Fig. 5.23-(a), for Rx2, we first measured Rx2’s output, then we gradually turned on Rx3, Rx5, and Rx6, and recorded the Rx2’s output with only Rx2 turning on, with Rx2 and Rx3 turning on, with Rx2, Rx3 and Rx5 turning on, and with Rx2, Rx3, Rx5 and Rx6 turning on. Similar measurement approach was applied in Fig. 5.23-(b),(c) and (d), where Rx3, Rx5 and Rx6 outputs are observed, respectively. As we can see, with different numbers of receiver turning on as the communication nodes, the normalized eye width and height of receivers’ outputs show variations of less than only 10%. Note that since Rx6 is placed at the end of the transmission line where the termination resistor might be off from the characteristic impedance of the transmission line, the eye width and height of Rx6 with only Rx6 turning on is much smaller than the cases when more receivers turning on. This issue can be mitigated by adding dummy transmission line between the Rx6 and the termination resistor.
Figure 5.21: Measured eye-diagrams at amplifier output of Rx5 with pre-emphasis control. Pre-emphasis off: (a) 10.16 Gb/s; (c) 15.24 Gb/s; (e) 20.32 Gb/s; (g) 25.4 Gb/s. Pre-emphasis on: (b) 10.16 Gb/s; (d) 15.24 Gb/s; (f) 20.32 Gb/s; (h) 25.4 Gb/s.
Figure 5.22: Measured eye-diagrams at amplifier output of Rx2 with pre-emphasis control. Pre-emphasis off: (a) 10.16 Gb/s; (c) 15.24 Gb/s; (e) 20.32 Gb/s; (g) 25.4 Gb/s. Pre-emphasis on: (b) 10.16 Gb/s; (d) 15.24 Gb/s; (f) 20.32 Gb/s; (h) 25.4 Gb/s.
Figure 5.23: Measured broadcasting communication of Tx1 with different number of receivers turning on: (a) Rx2 output; (b) Rx3 output; (c) Rx5 output; (d) Rx6 output.

After characterizing the single transmission line communication, we then characterize the multi-channel system performance. As we mentioned above, the second transmission line is used for this purpose. For a multi-channel system, crosstalk is very important since it is very likely that two or more channels works simultaneously. Therefore, the data transmission might be disturbed by the neighboring channels. In
this prototype, the crosstalk performance is measured using Rx7 and Rx8 at the second transmission line. Crosstalk is characterized with regard to the Rx5 at the first transmission line. Fig. 5.24 illustrates the corresponding crosstalk-to-signal power ratio. With the increase of the transmission length and with the increase of the data rate, the crosstalk performance degrades. At worst case where data rate is 25 Gb/s and the transmission length is 20 mm, a crosstalk-to-signal (Rx8-to-Rx5) power ratio of -32 dB is achieved, which shows that our GSSG-configured transmission line features good crosstalk performance.

![Graph](image)

Figure 5.24: Measured crosstalk performance for the multi-channel interconnect system.

This interconnect system prototype has another feature which can be applied to future implementation where higher data rate or longer transmission line are targeted. In this feature, there is a switch controlling the termination resistor. When the switch is on, the transmission line is terminated. In this case, the equivalent impedance for the driver connecting to this terminal will be half $Z_0$, where assuming the termination resistor equals to the characteristic impedance of the transmission line ($Z_0$). Note that, in all the previous measurement, we turn on the switch to make the transmission
line well terminated.

On the other hand, when the switch is off, the transmission line is equivalent to connect to a very large resistor. Therefore, the equivalent impedance for the driver connecting to this terminal will be approximately $Z_0$. The larger impedance in this case will benefit the data transmission for the driver at the terminal. For example, when Tx1 transmits the data and the termination switch turns off, the equivalent load of the Tx1 will increase, leading to a larger voltage swing. As shown in Fig. 5.25-(a) and -(b), by turning off the termination switch, the output at Rx5 has larger eye-opening. Therefore, this feature can be used to increase the transmission distance of the terminal node. However, this feature only applies for the node at the terminal, for the other nodes, the transmission needs to be well terminated. For example, as shown in Fig. 5.25-(c) and -(d), when turning off the termination switch, the data transmitted from Tx4 is totally distorted.

The transceiver consumes 21.15 mA from a 2-V power supply at 25.4 Gb/s, corresponding to an energy efficiency of only 1.67 pJ/b. The energy efficiency is expected to be further improved at a higher data rate. Table 5.1 summaries the performance of the chip prototype. A comparison with other works is summarized in Table 5.2.

<table>
<thead>
<tr>
<th>Technology</th>
<th>130-nm SiGe BiCMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply voltage</td>
<td>2 V</td>
</tr>
<tr>
<td>Transmission line configuration</td>
<td>two 20-mm-length parallel transmission lines</td>
</tr>
<tr>
<td>Supported communication</td>
<td>point-to-point/broadcasting</td>
</tr>
<tr>
<td>Maximum data rate</td>
<td>25.4 Gb/s</td>
</tr>
<tr>
<td>Bit error rate (BER)</td>
<td>$&lt;10^{-12}$ at 25.4 Gb/s</td>
</tr>
<tr>
<td>Power consumption @ maximum data rate</td>
<td>Transmitter (w/serializer): 24.8 mW</td>
</tr>
<tr>
<td></td>
<td>Receiver (w/deserializer): 17.5 mW</td>
</tr>
<tr>
<td>Energy efficiency</td>
<td>1.67 pJ/b @ 25.4 Gb/s</td>
</tr>
<tr>
<td>Crosstalk-to-signal power ratio</td>
<td>$&lt;-32$ dB</td>
</tr>
<tr>
<td>Simulated latency</td>
<td>7.86 ps/mm</td>
</tr>
</tbody>
</table>
Figure 5.25: Measured results of (a) Tx1 to Rx5 with termination switch on @ 20.32 Gb/s; (b) Tx1 to Rx5 with termination switch off @ 20.32 Gb/s. (c) Tx4 to Rx5 with termination switch on @ 12.7 Gb/s; (d) Tx4 to Rx5 with termination switch off @ 12.7 Gb/s.

Table 5.2: Performance comparisons with other on-chip interconnect systems

<table>
<thead>
<tr>
<th>Reference</th>
<th>Technology</th>
<th>T-line length (mm)</th>
<th>Maximum data rate (Gb/s)</th>
<th>Energy Efficiency (pJ/b)</th>
<th>Latency (ps/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[35]</td>
<td>0.18 μm CMOS</td>
<td>20</td>
<td>2</td>
<td>8.05</td>
<td>15</td>
</tr>
<tr>
<td>[37]</td>
<td>90 nm CMOS</td>
<td>5</td>
<td>8</td>
<td>0.15</td>
<td>30 (Including TRx)</td>
</tr>
<tr>
<td>This work</td>
<td>130 nm SiGe BiCMOS</td>
<td>20</td>
<td>25.4</td>
<td>1.67</td>
<td>7.9</td>
</tr>
</tbody>
</table>
Chapter 6

Discussions and Conclusions

In this thesis, we investigate and present new energy-efficient, wideband transceiver architectures and circuits for high-speed communications and interconnects: ultra-wideband impulse-radio receiver, optical transceiver for intra-chip free-space optical interconnect, and on-chip interconnect system for multi-core processors. Here we will conclude our work and discuss the options for further improvement on energy-efficiency and data rate for future implementation.

6.1 Discussions

In the IR-UWB receiver, we have demonstrated the communication at a pulse rate of 250 Mpulse/s. The pulse rate is mainly limited by the timing control circuit which supports limited output frequency range, and by the relatively long reset time after the integration. Therefore for future implementation with higher pulse rate requirement, we can improve the delay-lock loop in the timing control circuit to support higher frequency, and optimize the integration capacitors to reduce the reset time after the integration. The higher pulse rate can also potentially improve the system energy efficiency. Since the DPC is inherently duty-cycled, the energy efficiency of the DPC
remains constant even when the pulse rate is changed. However, the other blocks, such as LNA and VGA, mainly consume static power. Therefore, a higher pulse rate can help to improve their energy efficiency, and thus the system energy efficiency.

For the optical interconnect transceiver, as mentioned previously, to further improve the energy-efficiency, the latched sampler can be used in the receiver to simplify the amplifier design and to avoid using a dedicated decision circuit. In addition, multi-channel communication can be applied to effectively increase the system throughput. In the electrical interconnect solution implemented later, we have taken these into consideration and addressed the related design issues.

As we can see, latched sampler-based receiver has simpler architecture and less power consumption. If combined with the proper sampling clock schemes, as demonstrated in the prototype, the latch samplers can also perform deserialization function (e.g., 1:2 deserialization in this prototype). This idea can be further extended for
improved energy-efficiency. For example, a receiver with 8 latched samplers are implemented in [34], which can be used for both data sampling and 1:8 deserialization (Fig. 6.1). However, with the number of the latched sampler increases, there will be other design challenges, such as the design difficulty and complexity coming from the multi-phase sampling clock generator, which may degrade the energy-efficiency. Therefore, the number of the latched sampler need to be carefully chosen for the optimum performance.

In the prototype measurement, we observed very large bandwidth of the driver and the pre-amplifier, which can potentially enable them to be used for even higher data rate. However, due to the limitation of the data pattern generator and the limitation of the test equipment, the measurement data rate is limited to 25 Gb/s.
Therefore, we try to explore the possibility of higher data rate communication in the simulation, where a 20 Gb/s data pattern generator is designed. As shown in Fig. 6.2, the on-chip interconnect system can well perform at the 40-Gb/s data rate. Due to the fact that most of the current consumption is from static current, therefore with such a high data rate, the energy-efficiency can be improved to 0.98 pJ/b.

6.2 Conclusions

In this thesis, we first present an energy efficient, reconfigurable, distributed analog correlation IR-UWB receiver. We mainly focus on two essential building blocks in the receiver: low-noise amplifier and correlator. We investigate the design trade-off in the wideband low-noise amplifier, and use an ultra-wideband low-noise amplifier as a design example to explore the topic. The amplifier is based on resistive-feedback topology and employs noise cancelation technique. It is implemented in a 0.18-μm standard digital CMOS technology, and consumes 11.1 mW from a 1.8-V voltage supply. In the measurement, it achieves a 3-dB bandwidth of 0.7 - 6.5 GHz, power gain of 12.5 dB, and noise figure of 3.5 - 4.2 dB within the 3-dB bandwidth. We then design a low-noise amplifier with extended bandwidth of 2.3 - 9.8 GHz for the receiver. The correlator in the receiver is based on the time-interleaved, digitally-assisted distributed circuit technique called distributed pulse correlator (DPC). The time-interleaved technique maintains a high sampling rate while lowers the power consumption for the DPC, and the distributed technique extends the bandwidth of such a time-interleaved structure. Based on the multiplier analysis, a pulsed multiplier is presented to provide a better performance than the conventional Gilbert cell-based multiplier in the low voltage pulsed application. A wideband on-chip transmission line serves as the power distribution block to distribute the input UWB pulses to each pulsed multiplier without the distortion of the pulse. Additionally, thanks to
the built-in local template pulse generator and integrator of the DPC, the power consumption and circuit complexity of a DPC-based analog correlation receiver are significantly reduced. A DPC-based receiver prototype was implemented in a 0.18-μm standard digital CMOS technology. The energy efficiency of 40 pJ/pulse for the DPC and 190 pJ/pulse for the whole receiver at 250-MHz pulse rate in the measurement are achieved. The UWB impulse radio including DWG-based transmitter, DPC-based receiver and UWB antennas is also successfully demonstrated.

For the interconnect for future on-chip high-speed communication in multi-core or multi-module system, we present the transceivers for both optical solution and electrical solution. The optical solution is a 3-D intra-chip free-space optical interconnect. It integrates electric and optic component together by a 3-D fashion. We present a shared-clock-based optical transceiver. It simplifies the clock generation design in the transceiver, where a simple phase recovery block is used instead of the conventional complex clock recovery block. It helps a low power and low design complexity transceiver architecture compared to the conventional embedded-clock and forwarded-clock transceiver architecture. It also features a simple electric-optic interface. To further simplify the architecture and save the current consumption, an injection-locked oscillator can be used to replace the conventional phase-locked loop as the clock generation block.

The 10-Gb/s optical transceivers were implemented in a 0.13-μm standard digital CMOS technology. One transceiver is designed to work with a dedicated microprocessor. It includes 64:1 serializer, laser driver, PLL-based clock generator, active transimpedance amplifier, limiting amplifier, decision circuit, phase selector and 1:64 deserializer. The 64:1 serializer/1:64 deserializer are divided into low speed block and high speed block, where the low speed block is implemented using static CMOS circuit to save current consumption, while the high speed block is implemented using CML topology for high speed operation. The laser driver employs efficient pre-emphasis to
speed up the rising/falling edge, leading to an improved eye-diagram of high data-rate. The active transimpedance amplifier optimized for low input impedance, which benefits a larger bandwidth. Limiting amplifier, as a cascaded stage, is important for the bandwidth of the whole system. Active feedback and active inductive peaking techniques are used to extend the limiting amplifier bandwidth. In the simulation, the transceiver shows that a 10-Gb/s data rate with 41.5-pJ/b energy-efficiency communication can be achieved.

Another transceiver is also designed to further optimize the power consumption, where an injection-locked oscillator is used to replace the PLL to serve as the clock generation block. The ILO features low power consumption, low design complexity, high stability and fast locking. Besides, a resistor is used to serve as a passive transimpedance and to save power consumption. The resistor value is optimized for balanced performance on transimpedance gain and bandwidth. In this transceiver, no serializer and deserializer are implemented. In the simulation, the transceiver shows that a 10-Gb/s data rate with 7.1-pJ/b energy-efficiency communication can be achieved.

For the electrical solution of the on-chip interconnect, we presents a transmission-line-based on-chip interconnect system, which serves as a new alternative solution for high-speed communications in multi-core chips. It offers large throughput, low latency and high energy efficiency. It is highly reconfigurable, i.e., supporting both point-to-point and broadcasting links. In this thesis, we demonstrate this interconnect system using a chip prototype. The chip prototype was implemented in 130-nm BiCMOS technology, and the design can be applied to CMOS implementations. The transmission lines employs Ground-Signal-Signal-Ground configuration and patterned ground shields to achieve low latency, large bandwidth, less crosstalk and high bandwidth density. The transmitter uses an efficient and low power pre-emphasis technique to compensate for the transmission line’s frequency-dependent loss. The BiCMOS and
CML topologies are used to achieve high-speed operation. The modified latch without tail transistor benefits a lower supply voltage, which is good for power consumption. The receiver adopts the latched samplers for high sensitivity at high date-rate, it can detect the small-amplitude input signal caused by high transmission line loss and generate the large output swing signal. By using the high-sensitivity latched sampler, the front-end amplifier design is simplified and a dedicated decision circuit is avoided. The latched samplers are triggered by half-rate complementary clocks to perform the deserialization function, which further simplify the architecture and save the current consumption. In order to eliminate the dedicated isolation switches and the related insertion loss, both the transmitter and receiver are designed to be internally switched in/out from the transmission lines. The optimized transmitter and receiver can achieve good isolation without introducing insertion loss, benefitting long distance data transmission. In order to support broadcasting communications, the impedance of the transmitter and receiver are also well designed and optimized. In the measurement, the interconnect system successfully demonstrates both point-to-point and broadcasting communications. It achieves a date rate of 25.4 Gb/s per transmission line with an energy efficiency of 1.67 pJ/b, latency of 7.9 ps/mm, and BER of $<10^{-12}$.

In summary, wideband bandwidth benefits high data rate communications. The prototypes demonstrate that by careful analyzing the system and well designing the transceiver architectures and circuits, the energy-efficient, wideband and high-speed communication and interconnect systems can be achieved in mainstream low-cost CMOS/BiCMOS technologies.
Bibliography


