ULTRA-WIDEBAND TRANSCEIVER WITH ERROR CORRECTION FOR
CORTICAL INTERFACES IN NANOMETER CMOS PROCESS

by

Yi Luo

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

Approved:

Chris Winstead, Ph.D.
Major Professor

Todd Moon, Ph.D.
Committee Member

Sanghamitra Roy, Ph.D.
Committee Member

Koushik Chakraborty, Ph.D.
Committee Member

Xiaojun Qi, Ph.D.
Committee Member

Mark R. McLellan, Ph.D.
Vice President for Research and
Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2017
ABSTRACT

Ultra-Wideband Transceiver with Error Correction for Cortical Interfaces in Nanometer CMOS Process

by

Yi Luo, Doctor of Philosophy
Utah State University, 2017

Major Professor: Chris Winstead, Ph.D.
Department: Electrical and Computer Engineering

This dissertation reports a high-speed wideband wireless transmission solution for the tight power constraints of cortical interface application. The proposed system deploys Impulse Radio Ultra-wideband (IR-UWB) technique to achieve very high-rate communication. However, impulse radio signals suffer from significant attenuation within the body, and power limitations force the use of very low-power receiver circuits which introduce additional noise and jitter. Moreover, the coils’ self-resonance has to be suppressed to minimize the pulse distortion and inter-symbol interference, adding significant attenuation. To compensate these losses, an Error correction code (ECC) layer is added for functioning reliably to the system. The performance evaluation is made by modeling a pair of physically fabricated coils, and the results show that the ECC is essential to obtain the system’s reliability.

Furthermore, the $g_m/I_D$ methodology, which is based on the complete exploration of all inversion regions that the transistors are biased, is studied and explored for optimizing the system at the circuit-level. Specific focuses are on the RF blocks: the low noise amplifier (LNA) and the injection-locked voltage controlled oscillator (IL-VCO). Through the
analytical deduction of the circuit’s features as the function of the $g_m/I_D$ for each transistor, it is possible to select the optimum operating region for the circuit to achieve the target specification. Other circuit blocks, including the phase shifter, frequency divider, mixer, etc. are also described and analyzed. The prototype is fabricated in a 65-nm CMOS (Complementary Metal-Oxide-Semiconductor) process.
PUBLIC ABSTRACT

Ultra-Wideband Transceiver with Error Correction for Cortical Interfaces in Nanometer CMOS Process

Yi Luo

The wireless data transmission in the application of cortical interfaces requires a minimum of 27.0 Mbps for high-resolution visual functions within a very tight power constraint (< 10 mW). Most of the published designs for this application deploy narrow band techniques and have the throughput from several Mbps to tens of Mbps up to date. This work proposed a wideband solution with the use of Error-correction codes that can realize a throughput more than 100 Mbps within a power of 10 mW.
To my wife, Yuzhe and our adorable son Evan
ACKNOWLEDGMENTS

Creating a Ph.D. thesis is not an individual experience; rather it takes place in a social context and includes several persons, whom I would like to thank sincerely.

The most important acknowledgment of gratitude I wish to express is to my noble and esteemed guide, Dr. Chris Winstead. Dr. Winstead is a great thinker and a man who never tires of his passion for looking for new insights and ways to comprehend the reality around us. This thesis covers a variety of areas including communication and information theory, analog/RF and digital circuitry and needs many CADs and tools. Dr. Winstead always has a clarity of thought targeting the research goals and has the immense knowledge and experience guiding me to finish the work. He supported me not only by providing a research assistantship over almost seven years but also academically and emotionally through the rough road to complete this thesis. Thanks to him I had the opportunity to fabricate the chip. And during the most difficult times, he gave me the moral support and the freedom I needed to move on. This dissertation would not have been possible without his expert guidance, key insights, patience, and motivation.

I also would like to express my gratitude to my committee members: Dr. Todd Moon, Dr. Sanghamitra Roy, Dr. Koushik Chakraborty and Dr. Xiaojun Qi, for their insightful comments, encouragement and their time and effort in service to my doctoral committee despite their already heavy loads of responsibility. In particular, Dr. Moon’s class gives me a solid foundation on the mathematical methods and information theory, while Dr. Roy’s class gives me basic background on digital VLSI. The discussions in and out of these courses were always helpful and insightful.

My sincere thanks also go to Dr. Reyhan Baktur for her help during the coil fabrication and test. Her advice and suggestions inspired my new ideas and helped me solve the technical problems in my research efficiently. I would like to thank Dr. Ryan Gerdes for allowing me participating part of his project in which I learned a lot of skills for testing.
My graduate studies would not have been the same without the support and encouragement provided by all my colleagues and friends. Many thanks go to Gopalakrishnan Sundararajan for his collaboration on this project. He optimized and layout the digital circuitry and maintained the CAD system to ensure the completion of the project on schedule. I learned a great deal from the discussions with him. I would also like to thank Tasnuva Tithi for her valuable investigation on the ADC design. Thank you to Taha Shahvirdi for his hands-on help on the coil test. Besides, I also appreciated the friendship and encouragement from Toribio David, Abiezer Tejeda, Eduardo Monzon, Magathi Jayaram and Peter Beshay during the early period of this research.

My deepest appreciation also goes to the ECE Department and all of the staff members. I am very thankful to Mary Lee Anderson for the invaluable support in helping me solve various administrative matters; Scott Kimber for the immediate technical assistance and the server system maintenance; Heidi Harper for the immediate help with the test equipment and tools for so many times. You have contributed greatly in allowing me to beat paper deadlines.

I would also like to thank my good friends, Yanyan Yang, Guanyu Ma and JiaWang Gao for their unreserved help and encouragement. The time undertaken with them in Los Angeles greatly helped me put myself together and get me back on my feet.

Most importantly, at a personal level, I would like to sincerely thank my family without whose support the completion of this Ph.D. would not have been possible. I wish to thank my beloved wife, Yuzhe Liang, for her unconditional support and sacrifice all these years. Even in the most difficult time, she never gave me up and continuously trusted, encouraged and helped me. A special appreciation gives to my parents and parents-in-law, they have given up many things for me to be at USU, and supported me whenever I needed it. Also thanks to my little boy Evan, for being my everlasting source of joy and motivation.

Yi Luo
## CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
</tr>
<tr>
<td>PUBLIC ABSTRACT</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
</tr>
<tr>
<td>1.1 Motivation</td>
</tr>
<tr>
<td>1.2 Two Approaches: Narrowband and Wideband Modulations</td>
</tr>
<tr>
<td>1.2.1 Narrowband Modulation</td>
</tr>
<tr>
<td>1.2.2 Wideband Modulation</td>
</tr>
<tr>
<td>1.3 Challenges on Wideband Modulation</td>
</tr>
<tr>
<td>1.4 Contributions and Outline</td>
</tr>
<tr>
<td>2 SYSTEM ARCHITECTURE AND ANALYSIS</td>
</tr>
<tr>
<td>2.1 The State-of-Art for IR-UWB Receivers</td>
</tr>
<tr>
<td>2.1.1 Coherent Pulse-Template Correlation Receiver</td>
</tr>
<tr>
<td>2.1.2 Non-Coherent Self-Correlation Receiver</td>
</tr>
<tr>
<td>2.1.3 Super-regenerative Receiver</td>
</tr>
<tr>
<td>2.1.4 Injection-locked Receiver</td>
</tr>
<tr>
<td>2.1.5 Comparison of the Four Architectures</td>
</tr>
<tr>
<td>2.2 Proposed System Architecture and Link Analysis</td>
</tr>
<tr>
<td>2.2.1 Proposed Architecture</td>
</tr>
<tr>
<td>2.2.2 Link Transfer Function</td>
</tr>
<tr>
<td>2.2.3 Link Impulse Response</td>
</tr>
<tr>
<td>2.3 Coil Fabrication and Measurement</td>
</tr>
<tr>
<td>2.3.1 Coil Fabrication</td>
</tr>
<tr>
<td>2.3.2 UWB Transmitter</td>
</tr>
<tr>
<td>2.3.3 Coil Measurement</td>
</tr>
<tr>
<td>2.4 Receiver Performance Evaluation</td>
</tr>
<tr>
<td>3 DESIGN METHODOLOGY FOR NANOMETER CMOS PROCESS</td>
</tr>
<tr>
<td>3.1 Introduction</td>
</tr>
<tr>
<td>3.2 MOSFET ( g_m/I_D ) Model</td>
</tr>
<tr>
<td>3.2.1 ( g_m/I_D ) Versus Normalized Current ( i = I_D/(W/L) )</td>
</tr>
<tr>
<td>3.2.2 ( f_T ) Versus ( g_m/I_D )</td>
</tr>
<tr>
<td>3.2.3 ( g_{ds}/I_D ) Versus ( g_m/I_D )</td>
</tr>
<tr>
<td>3.2.4 ( g_{mb}/g_m ) Versus ( g_m/I_D )</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>NARROW BAND AND WIDEBAND RECEIVER COMPARISON</td>
<td>9</td>
</tr>
<tr>
<td>2.1</td>
<td>LRC VALUES AND COUPLING COEFFICIENT OF THE COIL</td>
<td>20</td>
</tr>
<tr>
<td>4.1</td>
<td>ADVANTAGE (√) AND DISADVANTAGE (×) OF LNA STAGES</td>
<td>59</td>
</tr>
<tr>
<td>4.2</td>
<td>MATLAB CALCULATIONS AND SPECTRERF SIMULATIONS COMPARISON FOR THE CG-LNA</td>
<td>68</td>
</tr>
<tr>
<td>4.3</td>
<td>MATLAB CALCULATIONS AND SPECTRERF SIMULATIONS COMPARISON FOR THE IL-VCO</td>
<td>84</td>
</tr>
<tr>
<td>5.1</td>
<td>AVERAGE POWER FOR EACH CIRCUIT BLOCK OF THE RECEIVER</td>
<td>113</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Cortical implant system for stimulating the visual cortex [1].</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Receiver Data Rate vs. Power Consumption</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>FCC spectrum mask for UWB indoor and outdoor communications [2].</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>Time domain and frequency domain mapping of carrier-based UWB signals with different pulse width. Carrier frequency $f_c = 4$ GHz.</td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>Time domain and frequency domain mapping of carrier-less UWB signals with different pulse width.</td>
<td>13</td>
</tr>
<tr>
<td>2.4</td>
<td>IR-UWB receiver architecture overview.</td>
<td>14</td>
</tr>
<tr>
<td>2.5</td>
<td>Receiver Data Rate vs. Power Consumption</td>
<td>18</td>
</tr>
<tr>
<td>2.6</td>
<td>Data Rate vs. Energy/bit.</td>
<td>19</td>
</tr>
<tr>
<td>2.7</td>
<td>Proposed UWB transmission system: Transmitter, Inductive Coil Link, and IR-UWB Receiver (modified from [3]).</td>
<td>20</td>
</tr>
<tr>
<td>2.8</td>
<td>Time responses for carrier-base UWB pulse (input pulse width = 2 ns).</td>
<td>22</td>
</tr>
<tr>
<td>2.9</td>
<td>Time responses for carrier-less UWB pulse (input pulse width = 2 ns).</td>
<td>23</td>
</tr>
<tr>
<td>2.10</td>
<td>Bode plot of $H_{12}(s)$.</td>
<td>24</td>
</tr>
<tr>
<td>2.11</td>
<td>The UWB pulse generator schematic.</td>
<td>25</td>
</tr>
<tr>
<td>2.12</td>
<td>Photo of (a) UWB pulse generator; (b) Tx coil; (c) Rx coil.</td>
<td>25</td>
</tr>
<tr>
<td>2.13</td>
<td>Measured S-parameters of the fabricated coil.</td>
<td>26</td>
</tr>
<tr>
<td>2.14</td>
<td>Experimental setup for measuring time response of the inductive coil link between the beef tissue.</td>
<td>28</td>
</tr>
<tr>
<td>2.15</td>
<td>Time response measurement of the coils.</td>
<td>29</td>
</tr>
<tr>
<td>2.16</td>
<td>Correlation coefficients of 300 measured and simulated samples.</td>
<td>30</td>
</tr>
<tr>
<td>2.17</td>
<td>Measured power spectral density for both pulses from Tx and Rx.</td>
<td>30</td>
</tr>
</tbody>
</table>
2.18 (a) System performance when the jitter RMS is 35 ps, 40 ps, 45 ps; NF of LNA is constant at 5 dB; (b) System performance when the LNA NF = 4 dB, 5 dB, 6 dB; jitter RMS is constant at 40 ps; (c) System performance when jitter RMS is 60 ps, 80 ps; LNA NF=5 dB; (d) Showing the quantities of performance improved by the ECC. The BER with ECC is unobservable when $V_{\text{rxb}} = 24.7$ mV and clock jitter < 55 ps.

3.1 $I_d$ versus $V_{ds}$ for different transistor length $L$. $V_{gs}$ and $W/L$ are constant at 0.6 V and 10, respectively.

3.2 $I_D$ versus $V_{GS}$ in logarithmic scale for an NMOS transistor with $W/L = 10 \mu m/70 \text{nm}$.

3.3 Comparison of square-law based method and $g_m/I_D$-based method [4].

3.4 $g_m/I_D$ versus $i$ for an NMOS transistor.

3.5 (a) NMOS transistor; (b) PMOS transistor; (c) MOSFET capacitances.

3.6 Layout of an NMOS transistor with $W = 6 \mu m \ (n_f = 4, \ W_n = 1.5 \mu m)$, $L = 70 \text{nm}$.

3.7 $g_m/I_D$ versus $i$ for different transistor width $W$.

3.8 $g_m/I_D$ versus $i$ for different transistor length $L$.

3.9 $g_m/I_D$ versus $i$ for different drain-source voltage.

3.10 $f_T$ versus $g_m/I_D$ for NMOS and PMOS transistors.

3.11 $g_m/I_D$ and $f_T$ versus overdrive voltage $V_{ov}$.

3.12 The product of $g_m/I_D$ and $f_T$ peaks in moderate inversion [4].

3.13 $g_{ds}/I_D$ versus $g_m/I_D$ for different $V_{ds}$.

3.14 $g_{ds}/I_D$ versus $g_m/I_D$ for different transistor length $L$.

3.15 $g_{mb}/g_m$ versus $g_m/I_D$ for NMOS and PMOS transistor.

3.16 Normalized capacitance $C_{ij}$ versus $g_m/I_D$ for NMOS transistor, $ij = \text{gs, gd, gb, bd, bs}$.

3.17 Normalized capacitance $C_{ij}$ versus $g_m/I_D$ for PMOS transistor, $ij = \text{gs, gd, gb, bd, bs}$.

3.18 Test circuits to acquire $g_m/I_D$, $g_{ds}/I_D$, $g_{mb}/g_m$ and $C_{ij}$. 
3.19 AC analysis for (a) single-ended inductor and (b) differential inductor.

3.20 (a) Serial inductor modeling and (b) parallel inductor modeling.

3.21 Layout of a differential inductor with its parameters: coil width \( w \), number of turns, and diameter \( d \).

3.22 Series resistance, parallel resistance and quality factor \( Q \) versus \( L \) at \( f_0 = 1.6 \) GHz.

4.1 Block diagram of the UWB transceiver prototype.

4.2 Basic LNA stage: (a) CS-LNA; (b) CG-LNA.

4.3 Gain-boosting basic idea: (a) conventional CG-LNA; (b) gain-boosted CG-LNA.

4.4 Complete gain-boosted CG-LNA.

4.5 Effective transconductance calculation of capacitive cross-coupled CG-LNA.

4.6 Equivalent half circuit model of the CG-LNA.

4.7 Small signal model for gain-boosted CG-LNA.

4.8 Power gain calculation.

4.9 SpectreRF simulation results of \( S_{11}, S_{22}, \) power gain, and noise figure (NF) for the three CG-LNA designs.

4.10 Voltage gain when \( V_{\text{tune,LNA}}=0 \) V and \( V_{\text{tune,LNA}}=1.2 \) V.

4.11 Injection locking phenomenon: (a) free-running oscillator; (b) injection-locked oscillator.

4.12 Injection-locking phenomenon: (a) first-harmonic injection locking; (b) super-harmonic injection locking; (c) sub-harmonic injection locking.

4.13 Injection-locked VCO.

4.14 Cross-coupled NMOS small signal analysis.

4.15 Complete VCO Core small signal model.

4.16 Small signal model for cascode structure.

4.17 Complete small signal model of IL-VCO.
4.18 IL-VCO free-running status: transient signal and coarse tunable frequencies from $b_3b_2b_1b_0 = 0000$ to $b_3b_2b_1b_0 = 1111$. .............................................. 85

4.19 IL-VCO running status with injected signal. .............................................. 86

4.20 IL-VCO locking range 1.52-1.66 GHz. .............................................. 86

4.21 Capacitor bank. .............................................. 88

4.22 Quadrature generation by divide-by-2 flip-flops. .............................................. 89

4.23 Basic diagram for quadrature VCO. .............................................. 89

4.24 $RC - CR$ polyphase filter. .............................................. 90

4.25 Layout sketch for the polyphase filter. .............................................. 92

4.26 Phase shifter working principle. .............................................. 93

4.27 Complete phase shifter circuit. .............................................. 95

4.28 Phase shift in one quadrature: $SI = 0$, $SQ = 0$ (totally $2^7 = 128$). .............................................. 96

4.29 Regenerative (Miller) divider [5]. .............................................. 96

4.30 Conventional divide-by-16 CML frequency divider. .............................................. 97

4.31 Conventional $LC$-Type frequency divider [6]. .............................................. 98

4.32 Modified divide-by-16 frequency divider. .............................................. 99

4.33 Modified resettable CML circuit. .............................................. 100

4.34 Timing diagram of the divide-by-16 frequency divider. .............................................. 101

4.35 Zoomed-in timing diagrams: (a) minimum requirement for $T_{ck}$; (b) maximum requirement for $T_{ck}$. .............................................. 102

4.36 Small signal model to calculate $R_L$ and $C_L$ for the CML circuit. .............................................. 104

4.37 Buffer: a complementary self-biased differential amplifier (CSDA) [7]. .............................................. 105

4.38 Output clock from the frequency divider. .............................................. 106

4.39 Average power for the frequency divider. .............................................. 107

4.40 Complete mixer circuit. .............................................. 107

4.41 Trellis diagram for the encoder with constraint length $K = 3$, generator polynomial $G(x) = [1 + x^2 \ 1 + x + x^2]$. .............................................. 109

4.42 Block diagram of the Viterbi decoder. .............................................. 110
INTRODUCTION

Inspired by the wireless data transmission in the application of cortical interfaces, this dissertation asserts that, the Impulse Radio Ultra-wideband (IR-UWB) signaling with Error Correction technique is a viable solution for the short distance, high data rate, low cost, highly integrated, low power radio communication. Existing narrowband transceivers are mostly order(s) of magnitude away from the transmission speed target for cortical interfaces. IR-UWB system is widely considered as a high speed, low power solution for short distance application, however, the power of existing designs is still outside the power constraint for cortical interfaces. The goal of this dissertation is to identify and specify a candidate architecture for the low power ($<10$ mW), high speed ($>100$ Mbps), short distance ($<10$ mm) applications especially for cortical interfaces. The system has been fabricated in a nanometer 65-nm CMOS (Complementary Metal-Oxide-Semiconductor) process.

1.1 Motivation

There have been continuous demands for high-speed, very low-power communication systems for bio-implantable electronic devices. One such area that is of significant interest is cortical interfaces which are used to stimulate the brain’s cortex directly. Visual cortical prosthetics are a class of cortical interfaces that restore functional vision in those suffering from partial or total blindness [8].

An implant system for stimulating the visual cortex is illustrated in Fig. 1.1 [1]. In this system, image data is captured by a camera integrated into eyeglass frames, then transmitted through an inductive transcutaneous link to an array of microelectrode stimulators implanted in the visual cortex. Direct cortical stimulation results into the perception of an image.

The system requires a high rate of data transmission to interface with a large array
Fig. 1.1: Cortical implant system for stimulating the visual cortex [1].

of stimulating sites. Weiland and Humayun showed that at least 1000 pixels are required to restore important visual functions such as face recognition and reading [9]. In such a visual implant, for example, the resolution of $32 \times 32$ (1024) pixels requires 10 bits for addressing purpose. If 8 bits are used for pulse amplitude control, 4 bits for polarity, parity-checking, and sequencing, this results in a total word length of 22 bits. Assuming that each of the channels is to be driven at the frequency, say, 200 Hz (for physiological reasons [10]), the data rate required is higher than $(4 \text{ words per biphasic pulse}) \times 22 \text{ (bit/word)} \times 1024 \text{ (channels)} \times 200 \text{ (pulses/channel/s)}$, or 18.0 Mbps (Megabit per second) [10]. If 300 Hz of refresh rate is required (for high-resolution visual functions), the minimum data rate increases to 27.0 Mbps. Thus, transceivers with data rate on the order of tens of Mbps or higher could be expected in high demand in the future.

To guarantee safe long-term use of bio-implantable devices, severe power constraints must be satisfied. The power density must be no greater than 0.8 mW/mm$^2$ to avoid tissue damage through heating effects [11]. For example, Lazzi showed that a chip with a power of 12.4 mW and dimension of $4 \times 4 \times 0.5 \text{ mm}$, would induce a temperature increase on the surface of the chip of approximately 0.8 °C when positioned in the center of a human
eyeball [12]. Generally, 10 mW is considered as the safe upper limit of the overall system since the temperature rises of more than 2 °C would damage the surrounding neurons [8]. It has proven difficult to design wireless receiver systems which meet this power constraint while simultaneously delivering high date rate in the order of tens or even hundreds of Mbps.

1.2 Two Approaches: Narrowband and Wideband Modulations

1.2.1 Narrowband Modulation

There are two modulation approaches for cortical interface application: narrowband and wideband. In the available literature, most of the designs are optimized for narrowband modulation, delivering high-amplitude signals to the implanted receiver by utilizing the harmonic resonance in coil interfaces functioned in the near-field domain [13]. In such solutions, high-quality factor ($Q$) is required to increase the selectivity of the coil. Thus a great number of narrowband techniques deploying high-$Q$ coils have been reported.

In 2004 Ghovanloo et al. demonstrated a Frequency-Shift Keying (FSK) system operating at 2.5 Mbps with 0.38 mW power consumption [14]. In 2007 Harrison described a bi-directional Amplitude-Shift Keying (ASK) system in which the forward link operated at 6.5 kbps (kilobits per second) and a reverse link operated at 330 kbps, with a total power consumption of 13.5 mW [15]. Harrison’s system was subsequently improved upon and used to implement a 100-site stimulation system [16].

More recent designs placed a greater emphasis on high-speed transmission. In 2007 Coulombe et al. described an ASK system achieving 1.5 Mbps at 0.9 mW [17]. In 2008 Mandal et al. described a Load-Shift Keying (LSK) system operating at 4 Mbps with 2.5 mW power consumption [18]. Zhou et al. demonstrated a Phase-Shift Keying (PSK) system at 2 Mbps while consuming 6.2 mW [19], and Luo et al. demonstrated a Binary Phase-Shift Keying (BPSK) design operating at 20 kbps and consuming 3 mW [20]. In 2012, Nabovati et al. described a high-speed BPSK demodulator capable of transmitting one bit per carrier cycle. The design achieved 16 Mbps with a carrier frequency of 16 MHz, while consuming
only 27 µW in simulation [21]. In 2015, Zgaren et al. presented a FSK receiver for 8 Mbps while consuming only 0.64 mW [22].

In 2011, a speed/power of 10.2 Mbps/3 mW was demonstrated by Inanlou et al. using an impulse-based wideband solution known as Pulse Harmonic Modulation (PHM) [23,24]. In this technique, two pulses with specific amplitude and time delays are transmitted every bit period through a high-Q coil link whose resonance frequency centered at 67.5 MHz. The first pulse generates a decaying oscillation at the harmonic frequency of the coil, while the second pulse produces a smaller oscillation with some delay that opposes in phase to the first oscillation, thus the inter-symbol interference (ISI) is minimized, reaching a high data rate without reducing the coils’ quality factor.

In 2013, an improved specification of the PHM receiver was proposed by Kiani et al. with an additional automatic gain control (AGC) mechanism, and it reached a speed/power of 20 Mbps/0.24 mW [25]. To the authors’ knowledge, this system has the highest throughput and is most efficient cortical interface receiver reported to date within the narrowband modulation category. Like most cortical interface designs, the system is a narrowband solution that benefits from resonance in the coil interface. It is unlikely that this method – or any narrowband methods – can be directly applied at much higher carrier frequencies because the coils’ resonant frequency can only be controlled within a limited range.

### 1.2.2 Wideband Modulation

Because the narrowband approach is limited in transmission speed, researchers have begun to consider wideband modulations. Kiani et al. proposed a Pulse Delay Modulation (PDM) technique that can offer a wide bandwidth for simultaneous data and power transmission across inductive links [26,27]. To transmit each bit, a pattern of narrow pulses are generated at the same frequency of the power carrier across the transmitter (Tx) data coil with specific time delays to initiate decaying ringing across the tuned receiver (Rx) data coil [27]. However, the data rate is limited by the low frequency of the power carrier. It was reported to be 13.56 Mbps while consuming 2.2 mW for the receiver side in [27].
Since the FCC (Federal Communications Commission) deregulated the use of Ultra-Wideband (UWB) in 2002 [28], UWB soon became quite popular for short range ultra-low-power applications such as contact-less chip testing [29] and body area networks (BAN) [30, 31]. The IR-UWB modulation was proposed in 2007 by Charles [32, 33] as a promising candidate for biomedical devices, due to its inherent high transmission speed (up to 1 Gbps [34]), low power consumption and low complexity. However, additional Clock Data Recovery (CDR) circuit is necessary because of the blind synchronization at the receiver. There is comparatively little work has been done on realizing UWB solutions for cortical interfaces. In 2010, Jow et al. demonstrated a UWB solution for back telemetry applications (transmission from internal to external), but only the design methodology for the link coils is presented [35]. In 2012, the authors described a system model for an implantable UWB receiver for cortical stimulators [3].

1.3 Challenges on Wideband Modulation

There are three challenges that may limit the application of UWB modulation for forwarding transmission (from external to internal) in cortical interfaces. These challenges are itemized in the following paragraphs. The contributions of this dissertation are to address these challenges, to demonstrate feasible solutions for realizing the target system, and to quantify trade-offs for optimizing the UWB forward link.

Coil implementation challenge

In an IR-UWB system, the data is modulated as a series of short pulses in time (less than 2 ns), whose central frequency is usually hundreds of MHz or on the order of GHz, with bandwidth greater than 500 MHz. To transmit such high-frequency signals, the self-resonance frequency (SRF) of both Tx and Rx coils need to be kept quite large in order to allow the high-frequency components of such narrow pulses to pass through the inductive link, minimizing the pulse distortion and ISI. Obtaining high SRF in the hundreds of MHz or even GHz range for implanted coils is quite challenging because the on-board [36, 37] or handmade [38] coils’ inductance, dimensions, and separations are much larger than the
on-chip coils used in chip-to-chip communication, while their parasitic resistance is much lower [29]. Also, the high conductivity of the tissue increases the parasitic capacitance around the coils, simultaneously decreases the SRF (since SRF = \( \frac{1}{2\pi\sqrt{LC}} \), where \( L \) and \( C \) are the inductance and capacitance of the coil, respectively). Jow and Ghovanloo reported a vertical and a figure-8 data coils by using a thin multistrand Litz wire (diameter=100\( \mu \)m) and the standard FR4 PCB (1.5 mm thick), which lead the coil’s SRF up to 98.8 MHz and 256 MHz [39], well below the UWB range.

**Very low Signal-to-Noise Ratio (SNR)**

One possible solution to the bandwidth limitation is reducing the quality factor(Q) value of the coil by adding series or parallel resistors so that the transfer function is flat across the entire UWB frequency range. With this approach, the amplitude of the received signal will then be reduced significantly because of the coils’ low-Q, as well as the receiver’s selectivity. As a result, the SNR at the Rx side is significantly reduced, possibly leading to unacceptable performance.

More than that, the high-frequency signals are highly absorbable *in vivo* and would be “blocked” by the human skull and conductive tissues [40–45], further decreasing the system’s SNR. The experiment has shown that the path loss (power attenuation of an electromagnetic wave) for an IR-UWB signal in the band of 3.5-4.5 GHz, varies between –20 dB and –30 dB for an in-body propagation distance of approximately 10 mm [45].

**High power consumption**

Typical IR-UWB receivers usually contain low-noise amplifier (LNA), multiplier, analog-to-digital converter (ADC) and synchronization circuitry, which would consume tens or hundreds mW [46–61]. While these receiver circuits may be considered “low power” in many application domains, they are still much higher than the 10 mW constraint specified for bio-implantable electronics.

Data for UWB radios as well as narrowband radios for bio-implantable devices are included in the plot Fig. 1.2, and listed in the Table 1.1. The plot shows receiver data
throughput plotted against power. The shaded region is our final target of the full receiver for cortical interfaces: data rate $> 100$ Mbps, while power $< 10$ mW. An interesting observation from this plot is that almost all the available researches function at up to order of magnitude less than the target specification. The narrowband solutions stay in the low power region as well as their low-speed characterization. The wideband solutions have high data throughput at the cost of high power consumption. Interesting exceptions are the designs in [56, 59, 62] which display admirable speed/power efficiency. The low-complexity receiver in [62] consists of only an LNA, a down conversion mixer and a buffer, leading to a very low power of $5$ mW when operating at 100 Mbps. The design in [59] does not include LNA and the one in [56] has a disadvantage of large frequency offset. All of these three designs, however, do not include a data synchronization mechanism.

![Figure 1.2: Receiver Data Rate vs. Power Consumption.](image)

1.4 Contributions and Outline

The goal of this research is the exploration, design, implementation, and demonstration
of a highly integrated, low power, impulse UWB transceiver for high rate cortical interfaces data transmission. This research focuses on the complete transceiver and will attempt to quantify the trade-offs between system performance and implementation. The maximum power target for the receiver is set at 10 mW at the speed of 100 Mbps. A 65-nm CMOS process is chosen for implementation as it allows for the potential of full integration of the digital logic with the analog processing blocks in a very low power cost.

In order to achieve a single-chip solution with low power consumption, the following approach is proposed:

- Investigation of ultra-wideband communication and determination of feasibility.
- Exploration, design, and fabrication of inductive, low-quality factor coil data link.
- Identification, modeling, and evaluation of the proposed system architecture.
- Identification of low power design techniques ($g_m/I_D$ methodology) for the circuits implemented in nanometer 65-nm CMOS process.

The remainder of this dissertation is divided into three parts. Chapter. 2 presents the state-of-art of the UWB system, followed by the proposed system model and performance evaluation. Chapter. 3 details the circuit design $g_m/I_D$ methodology and constructs a database for circuit optimization. Chapter. 4 discusses the circuit-level design and power/performance optimization by using the $g_m/I_D$ methodology. Chapter. 5 summarizes the main contributions of this dissertation.
Table 1.1: NARROW BAND AND WIDEBAND RECEIVER COMPARISON

<table>
<thead>
<tr>
<th>Reference</th>
<th>Data Rate (Mbps)</th>
<th>Power (mW)</th>
<th>Comm. Scheme</th>
<th>Data Synch *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donnell et al. (2006) [46]</td>
<td>10</td>
<td>1.8</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Zheng et al. (2006) [49]</td>
<td>200</td>
<td>81</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Barras et al. (2009) [48]</td>
<td>5</td>
<td>42.4</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Zhou et al. (2009) [50]</td>
<td>2000</td>
<td>145.8</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Zhang et al. (2009) [51]</td>
<td>100</td>
<td>156</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Michael et al. (2009) [57]</td>
<td>10</td>
<td>11.2</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Nick et al. (2010) [52]</td>
<td>39</td>
<td>6.2</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Denis et al. (2010) [55]</td>
<td>16</td>
<td>11</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Pelissier et al. (2010) [59]</td>
<td>112</td>
<td>5.4</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Zhuo et al. (2011) [53]</td>
<td>33</td>
<td>16.3</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Marco et al. (2011) [54]</td>
<td>1</td>
<td>4.1</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Xia et al. (2011) [56]</td>
<td>100</td>
<td>13.2</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Prakash et al. (2011) [58]</td>
<td>10</td>
<td>10.8</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Hu et al. (2011) [60]</td>
<td>500</td>
<td>45</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Gambini et al. (2012) [63]</td>
<td>1</td>
<td>0.29</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Hu et al. (2013) [64]</td>
<td>250</td>
<td>47.5</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Wang et al. (2014) [65]</td>
<td>1</td>
<td>0.1</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Rezaie et al. (2016) [69]</td>
<td>100</td>
<td>5</td>
<td>UWB</td>
<td>×</td>
</tr>
<tr>
<td>Brenna et al. (2016) [66]</td>
<td>20</td>
<td>0.965</td>
<td>UWB</td>
<td>✓</td>
</tr>
<tr>
<td>Ghovanloo et al. (2004) [14]</td>
<td>2.5</td>
<td>0.38</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Ghovanloo et al. (2007) [67]</td>
<td>2.5</td>
<td>8.25</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Harrison et al. (2007) [15]</td>
<td>0.3</td>
<td>13.5</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Sawan et al. (2007) [17]</td>
<td>1.5</td>
<td>0.9</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Mandal et al. (2008) [18]</td>
<td>4</td>
<td>2.5</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Zhou et al. (2008) [19]</td>
<td>2</td>
<td>6.2</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Luo et al. (2008) [20]</td>
<td>0.02</td>
<td>3</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Chen et al. (2010) [68]</td>
<td>2</td>
<td>5.7</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Inanlou et al. (2011) [23,24]</td>
<td>10.2</td>
<td>3</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Nabovati et al. (2012) [21]</td>
<td>16</td>
<td>0.027</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Chou et al. (2013) [69]</td>
<td>4</td>
<td>0.27</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Kiani et al. (2013) [25]</td>
<td>20</td>
<td>0.24</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Wilkerson et al. (2013) [70]</td>
<td>1</td>
<td>0.082</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Tan et al. (2014) [71]</td>
<td>0.1</td>
<td>0.78</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Karimi et al. (2014) [72]</td>
<td>10</td>
<td>0.078</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Zgaren et al. (2015) [22]</td>
<td>8</td>
<td>0.64</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Ba et al. (2015) [73]</td>
<td>4.5</td>
<td>1.59</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
<tr>
<td>Hsieh et al. (2016) [74]</td>
<td>0.12</td>
<td>0.35</td>
<td>Narrowband</td>
<td>✓</td>
</tr>
</tbody>
</table>

* Data with synchronization (✓) and without synchronization (×).
2.1 The State-of-Art for IR-UWB Receivers

Every wireless communication system designer would be pleased to have more data modulation bandwidth because more bandwidth offers a higher data throughput by Shannon's Equation [123]:

\[ C = B \log(1 + \frac{S}{N}) \]  

where \( C \) is the maximum channel capacity in bits/s; \( B \) is the channel bandwidth in Hz; \( S \) is the signal power in watts, \( N \) is the noise power also in watts. This equation tells us that there are three things that we can do to improve the capacity of a channel. We can increase the bandwidth, increase the signal power or decrease the noise power. The ratio \( S/N \) is known as signal-to-noise ratio (SNR) of the channel. The capacity of a channel grows linearly with increasing the bandwidth \( B \), but only logarithmically with signal power \( S \). Thus, from Shannon’s equation, we can see that the UWB system have a great potential for high-capacity wireless communications.

UWB characterizes transmission systems with spectral occupancy over 500 MHz or a fractional bandwidth of more than 20%. The power spectral emission mask of the UWB systems by FCC is illustrated in Fig. 2.1 [2]. The regulation allows spectrum sharing with low emission limit (-41.3 dBm/MHz Equivalent Isotropically Radiated Power (EIRP)) where the transmitted signal does not cause harmful interference to other band devices.

UWB signaling can be broadly grouped into two categories: carrier-based UWB and carrier-less UWB. Carrier-based UWB signal consists of narrow width pulses modulated by an RF carrier. In this case, the center frequency of the modulated signal is simply determined by the carrier frequency while signal power spectral density at RF is determined
Fig. 2.1: FCC spectrum mask for UWB indoor and outdoor communications [2].

by the baseband pulse shapes. Carrier-less UWB directly transmits the modulated narrow pulses without an RF carrier. Two example pulse-based UWB signals with different pulse width and their power spectrum density (PSD) are shown in Fig. 2.2 and Fig. 2.3. An extremely short pulse of few nanoseconds has its spectrum crossed over very wideband. The spectrum width could be controlled by transmitting pulses with different pulse durations. The signal could be modulated using several different ways including pulse-position modulation (PPM), pulse amplitude modulation (PAM), pulse width modulation (PWM), on-off keying (OOK), or binary phase-shift keying (BPSK).

A great variety of architectures have been proposed for pulse-based IR-UWB receivers. Four types of them, which have the top potential to achieve the power constraints of cortical interfaces while delivering data rate of 100 Mbps or higher, are listed in Fig. 2.4: (a) Coherent pulse-template correlation receiver [49–52], (b) Non-coherent self-correlation receiver [53–56], (c) Super-regenerative receiver [57–59] and, (d) Injection-locked receiver [60].

It should be noted that the IR-UWB architectures are not only limited to these four types. Multi-band orthogonal frequency-division multiplexing (MB-OFDM) UWB chan-
nelizes the signal into 3 or more sub-bands [47, 48], has been primarily used for streaming video or wireless USB, in which the high-performance electronics are required. Another type is direct oversampling analog-to-digital converter (ADC) that requires very fast ADCs at Nyquist rate up to multi-gigahertz, leading to an extremely high power consuming [46]. Thus, these systems are not amenable to the energy constrained cortical interface application and will not be discussed in this dissertation.

### 2.1.1 Coherent Pulse-Template Correlation Receiver

Both coherent and non-coherent receivers correlate the received pulse first such that
the center frequency is down-converted to baseband. The difference is that, in a coherent receiver, the received pulse correlates with a local template pulse; in a non-coherent receiver, the received pulse correlates with itself [56].

Fig. 2.4a presents the main blocks of pulse-template correlation receiver. The received UWB pulses are correlated by the local template pulses which are generated according to the information acquired by the synchronization algorithms and/or channel estimation. Then it is integrated and further amplified by the variable gain amplifier (VGA) to a constant level for analog-digital conversion [49–52]. Since the receiver has no information about the arrival of transmitted signals, the phase synchronization with the incoming Tx pulses is a critical issue and difficult because accurate alignment between TX impulses and RX templates must be achieved [75]. Furthermore, the transmitted impulse from the channel and the antennas may be significantly distorted, increasing the difficulty in generating an accurate pulse template. Two modes of operation are taken in the synchronization: data acquisition and data reception. A known header is first sent during the acquisition mode. The receiver scans all the possible window positions for the header and measures the energy in each window. Once the header window is found, the receiver is locked to the transmitter and is then switched to the reception mode [75]. The synchronization algorithms are usually
Fig. 2.4: IR-UWB receiver architecture overview.

(a) Coherent pulse-template correlation receiver.

(b) Non-coherent self-correlation receiver.

(c) Super-regenerative receiver.

(d) Injection-locked receiver.
run in the digital backend, which controls the analog front-end.

The analog front-end of coherent receiver usually consumes more than 100 mW because of power-eating blocks like phase-locked loop (PLL) and multipliers which are required in this architecture. Zheng etc. [49] described a carrier-less receiver (200 Mbps/81 mW) analog front-end for wireless personal area networking, in which a PLL with a ring oscillator was used for clock generation while two cascaded delay-locked loops (DLL) were used for synchronization. However, the power of the separated ADC (4-bit/81 mW) and the baseband circuitry was not considered. Zhang etc. [51] reported a receiver front-end, which used two channels to receive signals in a broadband 3.1-9.5 GHz, consumed 156 mW while offering 100 Mbps. In this receiver front-end, 54% of the power was spent on the multipliers. In the receiver reported by [52], signals were operating in the sub-1 GHz which could relax the design burden of radio radio frequency (RF) blocks. The entire receiver cost 6.2 mW in the acquisition mode and 4.2 mW in the reception mode, while the maximum speed was 39 Mbps.

### 2.1.2 Non-Coherent Self-Correlation Receiver

Fig. 2.4b presents the typical architecture of non-coherent self-correlation receiver. The advantage of a non-coherent receiver is that it avoids the generation of a local pulse, the received signal is firstly mixed with its self at radio-frequency, and a power-eating delay-locked loop (DLL)/PLL is usually introduced to perform synchronization between the received pulse and local pulse, needing precision on the order of several tens of picoseconds [56]. Denis et al. proposed a scheme that, after the received signal is self-correlated, a windowed integrator and ADC at baseband generate a digital signal representing the total energy received in a given time window [55]. The ADC values are passed to a digital backend, which performs packet detection, synchronization and decoding. This receiver was reported to be 11 mW at 16 Mbps. Zhuo and David et al. proposed another synchronization scheme in which, a packet composed by preambles and payloads are transmitted with sufficient amplitudes covering the worst case channel at a low pulse repetition frequency for receiver acquisition [53]. At the receiver side, parameters such as the SNR and the channel delay
spread are estimated to determine the operation mode and baseband behavior for synchronization and data reception. The performance was demonstrated to be 16.3 mW at 33 Mbps in [53].

Xia et al. [56] proposed a symbol-level synchronization which further simplified the timing scheme. In her method, the received signal is firstly squared and integrated by the correlator, then a comparator compares it with a reference voltage and performs digital quantization. Only a sliding correlator is employed instead of DLL/PLL for the data synchronization. This design gives an admirable specification of 13.2 mW at 100 Mbps. However, it was particularly emphasized by the authors that, the inevitable large frequency offset between the baseband clock of the transmitter and receiver should be compensated by an additional digital baseband circuitry. Even though the same clock source was used in the measurement for the design, the receiver’s sensitivity was $-50 \text{ dBm}$ at the BER of $10^{-3}$, much lower than $-66 \text{ dBm}$ to $-99 \text{ dBm}$ compared to other designs in their comparison table. The large power-saving was mostly taken by the integrator circuit which was realized by only two capacitors.

2.1.3 Super-regenerative Receiver

Another attractive super-regenerative receiver architecture was proposed in [57–59] and its main blocks is shown in Fig. 2.4c. This receiver is built around an LC-VCO (inductor-capacitor voltage-controlled oscillator) which is driven between oscillating or non-oscillating states by an external command quench signal. An RF input pulse will drive the LC-VCO to oscillate quickly, meanwhile, the trapezoidal quench signal is triggered synchronously with the input. When the quench signal is switched off, the oscillations will be damped until the next RF input sample phase. By detecting the envelope of the LC-VCO output and comparing with an appropriate threshold, the bit sequence can be detected. Because no multiplier is required and the LC-VCO is working at unstable state sometimes, this architecture can save more power consumption. In [57] and [58], the analog front-end of the receiver could offer 10 Mbps while cost only 11.2 mW and 10.8 mW, respectively. Furthermore, Pelissier et al. [59] presented a UWB receiver front-end for Radio-frequency
Identification (RFID) application which cost only 5.4 mW with the data rate of 112 Mbps excluding the low-noise amplifier (LNA) and synchronization circuitry.

The super-regenerative receiver suffers from low sensitivity in comparison to the coherent and non-coherent receivers which are based on energy detection, because the quench signal slope duration, which is relatively close to the core oscillator time period, will build up the oscillation in the absence of a pulse, thus degrading the SNR [58]. Furthermore, the receiver requires synchronization for both Tx/Rx and quench signal/RF input.

### 2.1.4 Injection-locked Receiver

Injection locking is a frequency effect that can occur when a harmonic oscillator is disturbed by a second oscillator operating at a nearby frequency. When the coupling is strong enough and the frequencies near enough, the second oscillator can capture the first oscillator, causing it to have the essentially identical frequency as the second [76, 77]. The injection-locked receiver, which eliminates the clock data recovery circuitry, multiplexer and any synchronization on the digital backend, is another promising architecture proposed by Hu et al. [60]. As shown in Fig. 2.4d, the local oscillator is injection-locked to the incoming OOK-modulated pulses and hence is automatically phase-aligned with the transmitted clock. The ADC sampling requirements can be severely relaxed and can run at the actual data rate since the receiver clock is now injection-locked and synchronized with the transmitter. This local oscillator is called injection-locked VCO (IL-VCO). It was reported that such a receiver could deliver 500 Mbps while consuming only 45 mW [60].

### 2.1.5 Comparison of the Four Architectures

Fig. 2.5 compares the data rate and power consumption for the recently reported IR-UWB receivers. The solid symbols represent full receivers which include synchronization processes.

The energy/bit, which is shown in Fig. 2.6, is another popular figure-of-merit for transceivers because it relates directly to the energy required to transmit or receive one
bit of data. Compared with other types of architectures, the IL-VCO-based receiver offers the best energy efficiency among all the full receivers.

To sum all, the IL-VCO-based architecture has the best potential to offer very high data speed while respecting the rigorous power limits. The transceiver proposed in this dissertation is based on this type of architecture.

### 2.2 Proposed System Architecture and Link Analysis

#### 2.2.1 Proposed Architecture

However, there are three further concerns on the injection-locked receiver:

1. The data is OOK-modulated, a long string of empty data transitions would result in loss of phase synchronization.

2. The receiver’s SNR would be significantly deteriorated if the power is constraint within 10 mW, leading to a possible unacceptable BER performance (it was originally 45 mW in [60]).
3. As mentioned earlier, the amplitude of the received signal would be reduced significantly because of the coils’ low quality factor, further decreasing the receiver’s SNR.

To address these three concerns, a low-complexity error correction decoder is added to compensate the loss of SNR in our proposed architecture of the UWB system. As depicted in Fig. 2.7 [1,3], the data is first encoded by the error correction encoder and then modulated using OOK in the transmitter. The lumped model of the low-Q coils $L_1$ and $L_2$, by which the UWB pulses are inductively coupled, is shown in the center of the figure. The clock signal $ck$ generated by the IL-VCO is synchronized with the actual data rate, and is used as the sample clock by the 2-bit ADC. The data is recovered by an low-complexity error correction decoder, which is used to compensate the loss of SNR.

We use two pairs of coils to transmit power and data separately. The power and data coils can be made approximately independent of each other (carrier frequency for power coils $< 10$ MHz [32]), so that the data channel is isolated from the power channel. In practice, there is likely to be a small mutual inductance between all coils, leading to potential interference [38]. Even though the coils with GHz-order SRF perform a band-
pass filter behavior, considering the amplitude of power signals could be up to 400 V [38], a passive high-pass filter ($R_f, C_f$ shown in Fig. 2.7) is inserted between the LNA and the ADC in order to cancel the potential interference from power coils.

### 2.2.2 Link Transfer Function

The data coils are designed following the methods described in [36–39, 78]. The $LRC$ values and the estimated coupling coefficient $k$ are shown in Table 2.1, in which $k = M/\sqrt{L_1L_2}$, where $M$ is the mutual inductance. $R_2$ is designed to be a large value in order to decrease the quality factor. The coils’ self-resonance frequency is chosen to be 4 GHz at this theoretical analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coils in this work</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_1$ ($\mu H$)</td>
<td>0.16</td>
</tr>
<tr>
<td>$C$ ($f F$)</td>
<td>10.12</td>
</tr>
<tr>
<td>$R$ ($\Omega$)</td>
<td>1.83</td>
</tr>
<tr>
<td>$R_s$ ($\Omega$)</td>
<td>50</td>
</tr>
<tr>
<td>$k$</td>
<td>0.0841</td>
</tr>
</tbody>
</table>

Since $L_1$ is loosely coupled (small $k$) with $L_2$ and the current in tank $L_2C_2$ is very small, we can neglect the effect of $L_2C_2$ loading on the tank $L_1C_1$ to simplify the equation [24]. The transfer function in the S-domain is deduced as shown in (2.2) [3]:

$$T(j\omega) = \frac{V_{out}}{V_{in}} = \frac{1}{1 + j\omega L_1C_1}$$
\[ H_{12}(s) = \frac{V_{od}(s)}{V_{id}(s)} = \frac{sM_{12}}{R_{s1}L_{1}C_{1} s^2 + (R_{s1}R_{1}C_{1} + L_{1}) s + (R_{1} + R_{s1})} \]
\[ \times \frac{1}{L_{2}C_{2} s^2 + R_{2}C_{2} s + 1} \]  

\[ (2.2) \]

2.2.3 Link Impulse Response

The impulse response \( h(t) \) can be deduced from (2.2) (see the Appendix A). Suppose the input signal is \( x(t) \), the output signal \( y(t) \) is the convolution of \( x(t) \) and \( h(t) \):

\[ y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(t - \tau) \cdot h(\tau) d\tau \]  

\[ (2.3) \]

A carrier-base (carrier frequency \( f_c = 4 \) GHz) UWB pulse with 2 ns width is transmitted from coil \( L_1 \) to \( L_2 \) under high-\( Q \) and low-\( Q \) conditions as shown in Fig. 2.8. Another example is a carrier-less Scholtzs monocycle UWB pulse [79] with 2 ns width as the input signal as presented in Fig. 2.9. The bode plot of \( H_{12}(s) \) is shown in Fig. 2.10. The plots indicate that the low-\( Q \) coils have a flat frequency response which is good to minimize pulse distortion and ISI, but also introduce larger attenuation estimated to be about 20 dB.

In the next section, a pair of low-\( Q \) coils are designed, fabricated, and test in bio-tissue to observe the real attenuation and the implementation possibility.
Fig. 2.8: Time responses for carrier-base UWB pulse (input pulse width = 2 ns).
Fig. 2.9: Time responses for carrier-less UWB pulse (input pulse width = 2 ns).
2.3 Coil Fabrication and Measurement

2.3.1 Coil Fabrication

Wideband data transmission requires high-SRF coils thus demands very low parasitic capacitor and inductance since \( SRF = \frac{1}{2\pi \sqrt{LC}} \). The bandwidth of the UWB signals is in the order of GHz. Standard FR4 material, of which the substrate dielectric constant \( \varepsilon_r = 4.4 \) and the dielectric loss tangent \( \delta = 0.02 \), can hardly satisfy the requirement according to our experience. We therefore choose a high frequency circuit material RO4003C (1.5mm thick, Rogers Corporation, Rogers, CT) for fabricating our Tx and Rx coils. The corresponding parameters of this material is \( (\varepsilon_r, \delta) = (3.55, 0.0029) \). Both of the coils are squared shape with only one loop, very small width in order to lower the parasitic capacitor and inductance.

We used a commercial field solver, HFSS (Ansys, Canonsburg, PA), to design and simulate all the model parameters necessary for the coil design described in [39, 80]. The photo and the dimension of the Tx and Rx coils is shown in Fig. 2.12. \( L_t = 19.2 \) mm and \( L_r = 8.0 \) mm are the length of Tx and Rx coil, respectively. \( W_t = 2.4 \) mm and \( W_r = 1.0 \) mm are the width of Tx and Rx coil, respectively. A 2 kΩ resistor is added to the Rx coil to lower the
A great variety of techniques may be used to generate UWB waveforms. In order to test the specification of the fabricated coils before designing the chip, we adopt a low-cost architecture based on the transient response of passive filters excited by step signals [81]. The step signal with sufficiently short rise/fall times is directly applied to the 3rd-order band-pass filter, as shown in Fig. 2.11, then the UWB pulses can be generated. We realized the circuit, as exhibited in Fig. 2.12, by using standard component values ($L = 0.5$ nH, $C = 1.0$ pF).

![Fig. 2.11: The UWB pulse generator schematic.](image)

![Fig. 2.12: Photo of (a) UWB pulse generator; (b) Tx coil; (c) Rx coil.](image)
2.3.3 Coil Measurement

It should be noted that the coils presented here are non-optimized. The scattering parameters (S-parameters) of the fabricated coil measured in air and beef environment are shown in Fig. 2.13.

Fig. 2.14 shows the measurement setup in order to test the time response of the coil. We have used a digital oscilloscope (Tektronix TDS7704B) to measure the transient signal waveforms from both Tx and Rx coils, while they are mounted vertically at different coupling distances using the support frame made of a mini T-Slot style building system MicroRax (Twintec, Auburn, WA). Two plastic Ziploc bags (~ 45 µm thick), hanging from a horizontal clamp, were filled with the combination of saline solution (0.9% NaCl) and thin beef slice.

![Graphs showing measured S-parameters](image)

Fig. 2.13: Measured S-parameters of the fabricated coil.
(ranging from 5 to 15 mm), both of which were in room temperature. The Rx coil was sandwiched between two bags of beef tissue while the Tx coil was aligned with it, touching the outer surface of one bag.

A train of squared pulses, with amplitude of 3.3 V and rise/fall time of 130 ps, were obtained by a pulse generator (HP8133A 3 GHz) and applied to the UWB transmitter. The transmitted UWB pulses with peak-to-peak voltage $V_{tx} = 1.16$ V, the corresponding received pulses in air environment ($V_{rxa} = 46.3$ mV), received pulses in beef tissue environment ($V_{rxb} = 69.2$ mV), are shown in Fig. 2.15. The distance between the Tx and Rx coils is 10 mm and the maximum pulse rate is 166 Mbps. It should be noted that the value of $V_{tx}$ could be adjusted by modifying the amplitude of the squared pulses (by the pulse generator), thus the $V_{rxb}$ in tissue environment, for example, could be adjusted, not limited to 69.2 mV.

In order to observe and evaluate the system performance in the next section, a rational function-based model which can characterize the coils’ port behavior as a function of frequency is generated by using the Matlab RF Toolbox (Mathworks, Natick, MA), and the simulation results are also presented in Fig. 2.15. It is shown that the simulated output waveform matches well with the physical measurement. For further observation, 300 randomly selected input UWB signals are simultaneously fed into both the physical coils and the generated computational model, obtaining the output signals denoted by $Y_c$ and $Y_s$, respectively. The correlation coefficients between $Y_c$ and $Y_s$ in air environment ($r_a$) and in beef tissue environment ($r_b$), representing their similarity, are exhibited in Fig. 2.16. It can be proved that the model is in good agreement with the physical coils.

There is a drawback of this type of transmitter that the adjacent UWB pulses are 180° phase-shifted because of the opposite rise and fall steps, but it is sufficient for this measurement. In practice, the transmitter is generally fabricated in digital circuit with tunable power and spectrum [56], which is outside the discussion of this section. The transmitter design on our chip is discussed in Chapter. 4.

Fig. 2.17 shows the measured power spectral density for both the Tx and Rx pulses.
Fig. 2.14: Experimental setup for measuring time response of the inductive coil link between the beef tissue.

It indicates that the high frequency components of Tx pulses (> 1.8 GHz) are highly absorbable in tissue environment and are hard to pass through tissue conductor.

2.4 Receiver Performance Evaluation

The realistic physical channel has been put into the proposed UWB receiver model shown in Fig. 2.7. The sampling clock jitter is modeled as a Gaussian-distributed random error in the received pulse sampling time. The noise figure (NF) of the nonideal LNA ranged from 4-6 dB is also accounted. OOK modulation, together with a low-complexity convolutional Viterbi decoder, with the generator polynomial $G(x) = \left[1 + x^2, 1 + x + x^2\right]$ and rate $R=1/2$, are adopted for the performance evaluation. The BER is evaluated as the primary measure of the system reliability.

Fig. 2.18a shows the system performance when the root mean square (RMS) of the jitter is 35 ps, 40 ps and 45 ps, while the noise figure of LNA is constant at 5 dB. The results illustrate that the system cannot function reliably without the use of Error Correction Code (ECC) technique. The sampling jitter creates an obvious error floor that limits the system’s performance, independent of the SNR of received signals. With the use of error correction code, the error floor is effectively eliminated.

Fig. 2.18b shows the system performance when the LNA noise figure is 4 dB, 5 dB and
Fig. 2.15: Time response measurement of the coils.
Fig. 2.16: Correlation coefficients of 300 measured and simulated samples.

Fig. 2.17: Measured power spectral density for both pulses from Tx and Rx.
6 dB, while the jitter-RMS remains at 40 ps. It reveals that (1) the level of the error floor is determined by the sampling jitter, and (2) improving the LNA’s noise figure will not contribute to improve the system performance without the use of ECC.

Performance with two exaggerated jitter-RMS, 60 ps and 80 ps, are also evaluated in Fig. 2.18c, while the LNA NF = 5 dB. The result indicates that, as a matter of fact, the error floor still exits if the jitter-RMS is extremely large, but it could be expected that the performance would be improved by the ECC technique for more than four orders. We introduce a factor $\eta$ which could quantify the BER comparison (only hard decoding is considered), defined as:

$$\eta = \frac{BER_{Uncoded}}{BER_{Hard \ decoded}}$$

(2.4)

Fig. 2.18d shows the quantities of performance improved by the ECC when the received pulse amplitude is $V_{pprxb} = 7.6 \text{ mV}, 10.4 \text{ mV}$ and $24.7 \text{ mV}$. It demonstrates that the reliability will be improved exponentially with the jitter-RMS decreases.

It is a useful reference showing that ECC can offer large compensation to the UWB system. The burden of designing power-eating cells, like the LNA, VCO, and ADC, is greatly released thus more power would be saved.
Fig. 2.18: (a) System performance when the jitter RMS is 35 ps, 40 ps, 45 ps; NF of LNA is constant at 5 dB; (b) System performance when the LNA NF = 4 dB, 5 dB, 6 dB; jitter RMS is constant at 40 ps; (c) System performance when jitter RMS is 60 ps, 80 ps; LNA NF=5 dB; (d) Showing the quantities of performance improved by the ECC. The BER with ECC is unobservable when \( V_{rxb} = 24.7 \) mV and clock jitter < 55 ps.
CHAPTER 3
DESIGN METHODOLOGY FOR NANOMETER CMOS PROCESS

3.1 Introduction

In Chapter 2, the system-level architecture and performance are discussed and evaluated, however, the translation of these characteristics to a final design, especially in nanometer CMOS process, is a big challenge. Many design parameters trade with each other, such as gain, linearity, power consumption, noise, speed, input-output impedance, voltage swing, etc., making the design a multi-dimensional optimization problem. Power consumption constrains the design of the transceiver and makes the design for each circuit block of the chain more challenging. It is especially noticeable to the trade-off between power and the inherent noise of RF blocks. To exemplify, a very low-noise application should accept high power consumption, whereas a very low-power design would need to manage higher noise values.

The trade-off between power consumption and noise is strongly determined by the active element: Metal-Oxide-Semiconductor Field-Effect-Transistor, or MOSFET. For deep submicron or nanometer CMOS process, however, the short and narrow channel effects on the device characteristics bring not only larger noise and mismatch, smaller transconductance, but also limitations of conventional square-law based analog design flow [82]. The voltage-current (V-I) square-law relation of the transistor is given by [83]:

\[
I_D = \begin{cases} 
\mu_n C_{ox} \frac{W}{L} \left( (V_{GS} - V_{TH}) V_{DS} - \frac{V_{DS}^2}{2} \right), & V_{DS} \leq V_{GS} - V_{TH}, V_{GS} \geq V_{TH} \\
\frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 (1 - \lambda V_{DS}), & V_{DS} > V_{GS} - V_{TH}, V_{GS} \geq V_{TH}
\end{cases}
\] (3.1)

where \(I_D\) is the drain-source current, \(\mu_n\) is the mobility of charge carries, \(C_{ox}\) is the gate oxide capacitance per unit area, \((W/L)\) is the width-to-length ratio of the transistor.
$V_{GS}$, $V_{DS}$, $V_{TH}$ are the gate-source voltage, drain-source voltage and threshold voltage, respectively. $\lambda$ is the channel-length modulation coefficient which is reciprocal to the channel length $L$, i.e., $\lambda \propto (1/L)$ [83].

This square-law equation matches well with the transistor behavior when $V_{DS}$ is small and/or $L$ is large because of the linear relation between carrier velocity and horizontal electric field. However, the behavior of short channel devices deviates considerably from this model due to the velocity saturation effect. At high field, the carrier velocity approaches the thermal velocity, and when the electric field reaches a critical value $E_c$, the carrier velocity tends to saturate due to scattering effects [82]. In this case, the $V$-$I$ relation of the transistor is modified as:

\[
I_D = \begin{cases} 
\kappa(V_{DS})\mu_nC_{ox}\frac{W}{L} \left[ (V_{GS} - V_{TH})V_{DS} - \frac{V_{DS}^2}{2} \right], & V_{DS} \leq V_{GS} - V_{TH}, V_{GS} \geq V_{TH} \\
\kappa(V_{DS})\frac{1}{2}\mu_nC_{ox}\frac{W}{L} (V_{GS} - V_{TH})^2 \left( 1 - \lambda V_{DS} \right), & V_{DS} > V_{GS} - V_{TH}, V_{GS} \geq V_{TH} 
\end{cases} 
\]  

(3.2)

where

\[
\kappa(V_{DS}) = \frac{1}{1 + \left( \frac{V_{DS}}{E_cL} \right)} 
\]  

(3.3)

For large values of $L$ or small values of $V_{DS}$, $\kappa$ approaches 1 and (3.2) reduces to (3.1). For short channel devices $\kappa < 1$, the current is smaller than what would be expected, as shown in Fig. 3.1.

Equations (3.1) and (3.2) are considered to be valid for $V_{GS} \geq V_{TH}$. In reality, for $V_{GS} \approx V_{TH}$, a “weak” inversion layer still exist, and some current flows from drain to source. Even for $V_{GS} < V_{TH}$, $I_D$ is finite [83]. As seen in the plot of Fig. 3.2, where the $I_D$ versus $V_{GS}$ is plotted in a logarithmic scale. The logarithmic plot shows that below the threshold voltage $V_{TH}$ the current is not zero and has an exponential relation with the gate-source voltage. This current is often referred to as sub-threshold current.

Circuit features change as a function of the inversion region in which the MOSFET is biased. The MOSFET used in analog and radio frequency has been traditionally biased
in strong inversion (SI) region. This region is characterized by high power consumption as well as high MOSFET transition frequency \( f_T \) due to the small sizing of the MOSFET. In this zone, the \( V_{GS} \) is well above the threshold voltage \( V_{TH} \). The other two inversion regions can be distinguished: the weak inversion (WI) region, or sub-threshold region, is the zone where \( V_{GS} \) is well below \( V_{TH} \); and the moderate inversion (MI) region which is in the midst of weak and strong inversion, approximately “around” threshold. In weak and moderate regions, the values of \( V_{GS} \) and the overdrive voltage \( V_{OD} = V_{GS} - V_{TH} \) are very low, which make these zones very adequate for low supply voltage operation, thus consume much less power but at the cost of lower transition frequency and higher noise due to the increment in MOSFET sizes.

Therefore, in low power analog circuits, it is crucial to the use of the MOSFETs in moderate and/or weak inversion regions, taking advantage of the nanometer technologies proliferating nowadays. However, the difficulty of optimizing the MOSFETs in moderate or weak inversion regions and achieving the “best” specifications for a circuit block is significantly increased for short channel nanometer process because of (not limited to) the
considerable deviated $V$-$I$ square-law relation. In fact, because of the complicated Modern MOSFETs, any square-law driven design optimization will be far off from Spice results.

In 1996, Silveira et al. [84] proposed a powerful transconductance-to-drain current ($g_m/I_D$) technique to optimize an operational transconductance amplifier (OTA) that has since become the basis of many later developments in analog circuit design. The $g_m/I_D$ methodology was originally developed to help designers to size up transistors quickly with good accuracy and to calculate parameters such as small signal gain and bandwidth. Recently, this approach has been found in applications such as phase noise optimization of an LC-VCO [85], MOSFET nonlinearity characterization [86], and low noise amplifier design [87]. A book dedicated to $g_m/I_D$ methodology for the analog amplifier design has also been published [88]. Some top-rated universities (Stanford [89] and Berkley [90]) teach courses specifically focused on this design methodology. Utah State University has offered similar courses on this methodology for many years as well [91]. Typically, analog circuit designers reach their goal by taking advantage of their own experience while performing lots of simulations and optimizations until the circuit “somehow” meets the specifications. With $g_m/I_D$ methodology, designers can quickly iterate a systematic design through tens of different designs at one time (Fig. 3.3).

The basic idea of $g_m/I_D$ methodology is that each $g_m/I_D$ value is one to one related with the normalized current $i = I_D/(W/L)$. It can be observed from the fact that the $g_m/I_D$ ratio is equal to the derivative of the logarithmic of $I_D$ with respect to $V_{GS}$ as shown

![Diagram of $I_D$ versus $V_{GS}$ in logarithmic scale for an NMOS transistor with $W/L = 10\mu m/70\,\text{nm}$]
The design is essentially right on target! – Typical discrepancies are no more than 10-20%, due to $V_{DS}$ dependencies, finite output resistance, etc. We accomplished this by using pre-computed spice data in the design process. Even if discrepancies are more significant, there’s always the possibility to track down the root causes.

- Hand calculations are based on parameters that also exist in Spice, e.g. $g_m/I_D$, $f_T$, etc.
- Different from square law calculations using $\mu C_{ox}$, $V_{OV}$, etc.
- Based on artificial parameters that do not exist or have no significance in the spice model.

![Comparison of square-law based method and $g_m/I_D$-based method](image)

Fig. 3.3: Comparison of square-law based method and $g_m/I_D$-based method [4].

\[
\frac{g_m}{I_D} = \frac{1}{I_D} \frac{\partial I_D}{\partial V_{GS}} = \frac{\partial \ln I_D}{\partial V_{GS}} = \frac{\partial \left\{ \ln \left( \frac{I_D}{(W/L)} \right) \right\}}{\partial V_{GS}} = \frac{\partial \ln (i)}{\partial V_{GS}} \tag{3.4}
\]

Ideally, the normalized current $i$ is independent of transistor size according to (3.2), the $g_m/I_D$ ratio is also size independent (in reality, $i$ has a slight dependence with $W$ and $L$ which will be shown later). Therefore, the relationship between $g_m/I_D$ and normalized current $i$ is a unique characteristic for all transistors of the same type (NMOS or PMOS) [84].

With this approach, as illustrated in Fig. 3.4, we consider a range of $g_m/I_D$ between $4 \text{V}^{-1}$ (strong inversion) and $26 \text{V}^{-1}$ (weak inversion), and the drain current $I_D$ varying from $I_{D,\text{min}}$ to $I_{D,\text{max}}$. Then for each pair ($g_m/I_D$, $I_D$), the normalized current $i$, the transconductance $g_m$ and the transistor size $(W/L)$ are deduced. The transistor length $L$ is assumed to be a constant value, for example the technology minimum to reach the highest transition frequency $f_T$, then the width $W$ is solved.

Fiorelli et al. [85] well explored the $g_m/I_D$ design methodology for the LC-VCO optimization. This chapter details some of their design flows and creates the necessary database utilized throughout the methodology. The database includes:

1. $g_m/I_D$ versus normalized current $i = I_D/(W/L)$;
2. $f_T$ versus $g_{m}/I_D$;

3. $g_{ds}/I_D$ versus $g_{m}/I_D$;

4. $g_{mb}/g_{m}$ versus $g_{m}/I_D$;

5. Normalized capacitance $C_{ij}$ versus $g_{m}/I_D$, $ij = \{gs, gd, gb, bd, bs\}$;

6. On-chip inductors.

3.2 MOSFET $g_{m}/I_D$ Model

Before explaining the database, a “lumped” MOSFET model needs to be stated in advance. Fig. 3.5 shows the NMOS and PMOS transistor symbols with the gate (G), drain (D) and source (S). Since in most circuits the bulk terminals of NMOS and PMOS transistors are tied to ground and $V_{DD}$, respectively, the bulks (B) are omitted in the symbols.

The capacitance exits between every two of the four terminals of a MOSFET and is also shown in Fig. 3.5. $C_{gs}, C_{gd}, C_{gb}, C_{bd}, C_{bs}$ are the gate-source, gate-drain, gate-bulk, bulk-drain and bulk-source capacitances of the MOSFET, respectively. The capacitance between S and D is negligible [83].

3.2.1 $g_{m}/I_D$ Versus Normalized Current $i = I_D/(W/L)$

The $g_{m}/I_D$ versus normalized current $i = I_D/(W/L)$ is extracted for the transistor length $L = 70$ nm which is widely used in the following circuits design in order for the max-
Fig. 3.5: (a) NMOS transistor; (b) PMOS transistor; (c) MOSFET capacitances.

imum transition frequency $f_T$. Physically for a MOSFET layout, as presented in Fig. 3.6, if the transistor finger width is $w_n$ and the number of fingers is $n_f$, the total transistor width is determined by $W = w_n \times n_f$. For the data acquisition, a set of $W = \{1, 2, ..., 160\}$ $\mu$m is simulated, which is the pairwise product of $w_n = \{1, 2, 3, 4, 5\}$ $\mu$m and $n_f = \{1, 2, 4, 8, 16, 32\}$. This range is enough for covering well the variations of $g_m/I_D$ versus $i$ in the whole transistor region. The data is presented in Fig. 3.7, indicating that the $g_m/I_D$ versus $i$ plot is width independent. Even though the curves have a slight variation especially in the strong inversion, they are accurate enough for quickly sizing up transistors at the beginning of the design by deploying only one curve of them.

Fig. 3.8 shows the $g_m/I_D$ versus $i$ plots when transistor width is constant at $W = 10$ $\mu$m and the length $L = \{70, 80, 90, ..., 300\}$ nm. In this case, larger variations exist especially in the weak inversion region.

In order for a full observation, the $g_m/I_D$ versus $i$ plot variations with the $V_{ds}$ voltage are presented in Fig. 3.9. The transistor length is constant at $L = 70$ nm. $V_{ds}$, $w_n$ and $n_f$ are swept in the range $\{0.3, 0.6, 0.9, 1.2\}$ V, $\{1, 2, 3, 4, 5\}$ $\mu$m and $\{1, 2, 4, 8, 16, 32\}$, respectively. Note that the drain-source voltage for PMOS is $V_{sd}$. It can be observed that the variations of $g_m/I_D$ versus $i$ plot with transistor size and $V_{ds}$ do not modify considerably. Therefore,
Fig. 3.6: Layout of an NMOS transistor with $W = 6 \, \mu\text{m}$ ($n_f = 4, W_n = 1.5 \, \mu\text{m}$), $L = 70 \, \text{nm}$.

- $L = 70 \, \text{nm}$
- $V_{ds} = 0.6 \, \text{V}$
- $n_f = \{1, 2, 4, 8, 16, 32\}$
- $W = \{1, 2, \ldots, 160\} \, \mu\text{m}$

Fig. 3.7: $g_m/I_D$ versus $i$ for different transistor width $W$. 

- $L = 70 \, \text{nm}$
- $V_{sd} = 0.6 \, \text{V}$
- $n_f = \{1, 2, 4, 8, 16, 32\}$
- $W = \{1, 2, \ldots, 160\} \, \mu\text{m}$
Fig. 1.8: \( \frac{g_m}{I_D} \) versus \( i \) for different transistor length \( L \).

Fig. 3.8: \( \frac{g_m}{I_D} \) versus \( i \) for different transistor length \( L \).
the “$g_m/I_D$ versus $i$” plot can be used as the transistors intrinsic property throughout the circuit design process.

In reality, only one curve, for example the curve for $L = 70 \text{ nm}, W = 10 \mu\text{m}$, is enough with a good accuracy throughout the design process. It is not complicated, however, to use the curve for a specific $L$ for a better accuracy in the Matlab scripts generated for the circuit parameter calculation, which will be discussed in Chapter 4.

### 3.2.2 $f_T$ Versus $g_m/I_D$

For an MOS transistor, the transition frequency $f_T$ is defined as the frequency where the magnitude of the short-circuit, common-source current gain falls to unity [82]:

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd} + C_{gb})}$$  \hspace{1cm} (3.5)

This parameter predicts, not accurately, the frequency limit of one single transistor, and can be considered as a figure of merit for a transistor. Inaccuracy comes from the inaccurate “lumped” transistor model and many high order effects at higher frequencies, as well as the layout design and device connections. In terms of experience, the “lumped” transistor model is accurate enough up to about $f_T/5$ [4].

The $f_T$ versus $g_m/I_D$ plot shown in Fig. 3.10 indicates that, for an NMOS transistor biased in strong inversion, $f_T$ can surpass 100 GHz and PMOS transistor reaches 80 GHz. In the weak inversion, the $f_T$ drop down to levels below the gigahertz. This check simply gives rough limitations of the technology in terms of frequency in each inversion.

An interesting observation is the design trade-off between $g_m/I_D$ and $f_T$. Fig. 3.11 shows the $g_m/I_D$ and $f_T$ versus overdrive voltage $V_{ov}$, which is defined as the $V_{GS}$ in excess of the threshold voltage $V_{TH}$, i.e., $V_{ov} = V_{GS} - V_{TH}$. In weak inversion, $g_m/I_D$ is large but $f_T$ is small; in strong inversion, $g_m/I_D$ is small but $f_T$ is large. The product of $g_m/I_D$ and $f_T$, which is shown in Fig. 3.2.2, peaks in moderate inversion. Operating the transistor in moderate inversion is optimal when we value speed and power efficiency equally.
moderate inversion is optimal when we value speed and power efficiency equally.

in the weak inversion, the transistor model is accurate enough up to about $f_T/5$ [1].

In the strong inversion, the magnitude of the short-circuit, common-source current gain falls to unity [2]:

$$H = \frac{I_{DS}}{I_{DS}}$$

This parameter predicts, not accurately, the frequency limit of one single transistor,

and can be considered as a figure of merit for a transistor. Inaccuracy comes from the

circuit parameter calculation, which will be discussed in Chapter 1.

use the curve for a specific

of the threshold voltage $V_T$.

$V_T$ is small; in strong inversion,

$V_T$ drop down to levels below the gigahertz. This check simply

can surpass 100 GHz and PMOS transistor reaches 80 GHz.

The $f_T$ versus overdrive voltage

for NMOS and PMOS transistors.

The $f_T$ versus $g_m/I_D$ plot shown in Fig. 1.10 indicates that, for an NMOS transistor

biased in strong inversion, $f_T$ ... inversion. Operating the transistor in

is optimal when we value speed and power efficiency equally.
Fig. 3.11: $g_m/I_D$ and $f_T$ versus overdrive voltage $V_{ov}$.

Fig. 3.12: The product of $g_m/I_D$ and $f_T$ peaks in moderate inversion [4].
3.2.3 \( g_{ds}/I_D \) Versus \( g_m/I_D \)

The drain current \( I_D \) also varies with the drain-source voltage \( V_{ds} \) due to the channel-length modulation. A very important MOSFET parameter needed for analog circuit design is the drain-to-source conductance \( g_{ds} \), defined by [83]:

\[
g_{ds} = \frac{\partial I_D}{\partial V_{ds}} = \frac{1}{r_o} = \lambda I_D
\]  

(3.6)

where \( r_o \) is the output resistance of the transistor, \( \lambda = g_{ds}/I_D \) is the channel-length modulation coefficient as mentioned in (3.1).

The \( g_{ds}/I_D \) versus \( g_m/I_D \) plot for four drain-source voltages is presented in Fig. 3.13. The transistor length \( L = 70 \) nm, while the transistor width is swept in the range of \( \{1, 2, 3, 4, 6, ..., 160\} \) \( \mu \)m. The variations are clearly appreciable. The \( g_{ds}/I_D \), i.e. the \( \lambda \), remains low when \( g_m/I_D \) is small in which case the transistor is in a strong inversion region, leading to a higher output resistance \( (r_o = 1/(\lambda I_D)) \) at a constant \( I_D \). Attention could be paid that, the transistor intrinsic gain \( A_i \), as defined in (3.7), is reciprocal to the curve slope in Fig 3.13.

\[
A_i = g_m r_o = \frac{g_m}{g_{ds}} = \frac{g_m/I_D}{g_{ds}/I_D}
\]  

(3.7)

It is worth observing the \( g_{ds}/I_D \) versus \( g_m/I_D \) plot for different transistor length \( L \). As shown in Fig. 3.14, the \( g_{ds}/I_D \) decreases when \( L \) increases. It is true because \( g_{ds}/I_D \), or say \( \lambda \), is reciprocal to the channel length \( L (\lambda \propto (1/L)) \) in the first order model approximation [83].

The \( g_{ds} \) parameter has a considerable spread with variations in the drain-source voltage and different transistor length. These facts make us expect some differences between computational data and simulations if the correct drain voltage is not chosen.

3.2.4 \( g_{mb}/g_m \) Versus \( g_m/I_D \)

The bulk potential influences the threshold voltage and hence the gate-source overdrive. This is called bulk effect which mainly factored in as an effective increase in the threshold
voltage $V_{TH}$. The threshold voltage is given by [83]:

$$V_T = V_{TH0} + \gamma \left( \sqrt{2\phi_F + V_{SB}} - \sqrt{2\phi_F} \right)$$  \hspace{1cm} (3.8)

where $V_{TH0}$, $\gamma$ and $\phi_F$ are technology related parameters. This equation states that the threshold voltage is a function of the technology and the applied source-bulk voltage $V_{SB}$. Define the transconductance $g_{mb} = \partial I_D / \partial V_{BS}$ and its relation with $g_m$ is [83]:

$$g_{mb} / g_m = \eta = \frac{\gamma}{2\sqrt{2\phi_F + V_{SB}}}$$  \hspace{1cm} (3.9)

This equation suggests that incremental bulk effect becomes less pronounced as $V_{SB}$ increases. $g_{mb}$ is much smaller than $g_m$, as it is shown in Fig 3.15. However, it should be extracted in the database especially when the bulk effect is important for the circuit behavior.
Fig. 3.14: $g_{ds}/I_D$ versus $g_m/I_D$ for different transistor length $L$. 

PMOS $g_{ds}/I_D$

NMOS $g_{ds}/I_D$

$V_{ds} = 0.6\, \text{V}$

$W = 10\, \mu\text{m}$
3.2.5 Normalized Capacitance $\overline{C}_{ij}$ Versus $g_m/I_D$

At high frequencies, the intrinsic capacitances of the MOSFET, which is mainly composed of $\{C_{gs}, C_{gd}, C_{gb}, C_{bd}, C_{bs}\}$, must be taken into account since they substantially influence the circuit behavior. These capacitances change with the inversion level and the transistor size [83] and are difficult to model or predict before simulation. Fiorelli et al. [85] proposed a semi-empirical method to model the MOSFET intrinsic capacitance used for $g_m/I_D$ methodology. In this method, the intrinsic capacitances are considered to be proportional to the gate area ($WL$). This can be done because these capacitances are proportional to the oxide capacitance $C_{ox}$ which is itself proportional to ($WL$) [82,83].

The normalized capacitance is defined by:

$$\overline{C}_{ij} = \frac{C_{ij}}{WL}, ij = \{gs, gd, gb, bd, bs\} \tag{3.10}$$

Fig 3.16 and 3.17 shows the normalized capacitance $\overline{C}_{ij}$ versus $g_m/I_D$ of for NMOS and
PMOS transistors with a set of width and $V_{ds}$, $C_{ij}$ are used to estimate the capacitances of the transistors in the considered width range, only multiplying by the respective $WL$. A rough estimation can be obtained by extracting a middle-sized MOSFET with a middle bias of $V_{ds}$.

A summary of their basic properties is:

- $C_{gs}$: (1) varies between 4-10 fF/µm² through all regions; (2) the largest capacitance that should be mainly taken into account in the circuit design.

- $C_{gd}$: (1) mostly remains at around 4 fF/µm² though all regions; (2) can be as large as $C_{gs}$ when $V_{ds}$ is low in the strong inversion.

- $C_{gb}$: (1) varies between 1-3 fF/µm² through all regions; (2) values are close to $C_{gd}$.

Therefore, $C_{gd}$ and $C_{gb}$ are the second largest capacitances that sometimes need to be considered.

- $C_{bd}$: (1) no more than 0.05 fF/µm² through all regions; (2) can be ignored in most cases.

- $C_{bs}$: (1) no more than 1 fF/µm² through all regions; (2) can be ignored in most cases.

### 3.2.6 Data Acquisition Scheme

In order to acquire these data, the test circuits of Fig. 3.18 is utilized. The MOSFET gate, drain and source nodes are connected to a DC voltage source. The bulk nodes are connected either to the ground (NMOS) or the supply voltage $V_{DD}$ (PMOS). $V_{GS}$ is swept from 0 to $V_{DD}$. $V_D$ and $V_S$ are set around their expected DC value. The power supply is 1.2 V in the process.

### 3.3 On-Chip Inductor $g_m/I_D$ Model

In radio frequency design, on-chip passive inductors need to be correctly characterized since they strongly determine the performance of the circuit. For example, in a VCO design,
NMOS, \( L = 70 \text{ nm}, W = \{1, 2, ..., 160\} \mu \text{m} \)

Fig. 3.16: Normalized capacitance \( C_{ij} \) versus \( g_m / I_D \) for NMOS transistor, \( ij = gs, gd, gb, bd, bs \).
Fig. 3.17: Normalized capacitance $C_{ij}$ versus $g_m/I_D$ for PMOS transistor, $ij = gs, gd, gb, bd, bs$. 

PMOS, $L = 70 \text{nm}$, $W = \{1, 2, ..., 160\} \ \mu\text{m}$
Fig. 3.18: Test circuits to acquire $g_m/I_D$, $g_{ds}/I_D$, $g_{mb}/g_m$ and $C_{ij}$.

if the tank inductor models a parasitic parallel resistor much lower than the real one, the VCO will fail to oscillate.

Inductor models can be extracted by using the 3-D electromagnetic simulators by solving Maxwell’s equations numerically, such as HFSS (Ansys, Canonsburg, PA), ADS Momentum (Keysight, Santa Rosa, CA) or ASITIC [92]. Although accurate, they have three main drawbacks: (1) they need to have the technological data provided by the foundry to obtain accurate descriptions; (2) the simulations are computationally intensive, both in memory and time, and (3) the use of these tools complicates the interface between the inductor model and the circuit simulator (such as SPICE).

In order to speed-up the design, the inductor models in this dissertation are extracted by utilizing the parameterized library provided by the foundry. The inductors in the library are simulated using AC analysis at the working frequency to obtain their equivalent complex impedance.

The schematic in Fig. 3.19 shows the AC analysis for single-ended and differential inductor. For a single-ended inductor, one of its ports is connected to the source, and the other is grounded. For a differential inductor, its middle port is AC grounded, and the other two are connected to identical sources with a phase difference of 180° in order to consider the differential voltages between these ports.
Fig. 3.19: AC analysis for (a) single-ended inductor and (b) differential inductor.

A extracted inductor can be modeled as an equivalent ideal inductor with a parasitic resistor, either in series or parallel, at the working frequency $f_0$, as presented in Fig. 3.20.

![Fig. 3.20: (a) Serial inductor modeling and (b) parallel inductor modeling.](image)

The impedance of an inductor is defined by:

$$Z_{ind} = R_s + j\omega L_s = R_p / j\omega L_p$$

(3.11)

The quality factor, $Q$, is:

$$Q = \frac{\omega L_s}{R_s} = \frac{R_p}{\omega L_p}$$

(3.12)
At the working frequency, the serial combination of Fig. 3.20a can be converted to its equivalent parallel configuration of Fig. 3.20b, by the relationship:

\[
\begin{align*}
L_p &= L_s \left(1 + \frac{1}{Q^2}\right) \\
R_p &= R_s \left(1 + Q^2\right)
\end{align*}
\]

Fig. 3.21: Layout of a differential inductor with its parameters: coil width \(w\), number of turns, and diameter \(d\).
Fig. 3.22: Series resistance, parallel resistance and quality factor ($Q$) versus $L$ at $f_0 = 1.6$ GHz.
CHAPTER 4
UWB TRANSCEIVER CIRCUITS IMPLEMENTATION

This chapter describes the design and implementation of each circuit block, focusing primarily on ultra-low power target and the $g_m/I_D$ methodology for circuit optimization.

The proposed UWB transceiver block diagram is shown in Fig 4.1, consisting of a transmitter with error correction encoder, a receiver with a common-gate low noise amplifier (CG-LNA), an injection-locked VCO (IL-VCO), a phase shifter and a divide-by-16 frequency divider (/16). The phase shifter is used to align the phase between the Clock generated from the divider and the amplified data pulses from the LNA.

The FPGA (field-programmable gate array) is used for test purpose, including generating PRBS (pseudorandom binary sequence) data to the TX and collecting decoded (recovered) data from the RX for the BER calculation. Considering the properties of the fabricated coils described in Chapter 2, 1.6 GHz was selected as the carrier frequency, which implies that the target sample clock generated for the 2-bit ADC is 100 MHz thus the target data rate is 100 Mbps. Due to the fabrication deadlines, however, the 2-bit ADC was not implemented on-chip.

4.1 Low Noise Amplifier

4.1.1 Brief Introduction

There are two candidate topologies for the LNA: common-source (CS) and common-gate (CG). The basic CS-LNA and CG-LNA circuits are depicted in Fig. 4.2.

The input impedance of the CG-LNA stage $Z_{in,CG}$ is approximate $1/g_m$, ease of design to match to 50 Ω, while that of CS-LNA $Z_{in,CS}$ is:

$$Z_{in,CS} = s(L_g + L_s) + \frac{1}{sC_{gs}} + \left(\frac{g_m}{C_{gs}}\right)L_s$$

(4.1)
Fig. 4.1: Block diagram of the UWB transceiver prototype.
As can be seen, $Z_{m,CS}$ is sensitive to the input parasitic parameters, i.e. the process, voltage and temperature (PVT) variations.

The effective transconductance of the CS-LNA stage, with the input matched to $R_s$, is:

$$G_{m,CS} = \frac{1}{2R_s} \left( \frac{\omega_T}{\omega_0} \right)$$  \hspace{1cm} (4.2)

where $\omega_T$ is the transition frequency, $\omega_0$ is the operating frequency.

The effective input transconductance of the CG-LNA with same perfect matching conditions is:

$$G_{m,CG} = \frac{1}{2R_s}$$  \hspace{1cm} (4.3)

Typically, the $\omega_T/\omega_0$ lies in the range of 5-10 depending on the operating frequency and the process details. Therefore, the CS-LNA has higher gain than the CG-LNA [93].
The theoretical minimum noise figure of CS-LNA, with the optimum $Q$ of the input resonant circuit, is \[93\]:

\[
F_{\text{min,CS}} = 1 + \frac{\gamma}{\alpha} \left( \frac{\omega_0}{\omega_T} \right) \frac{2\delta\alpha^2}{\gamma^2} Q_{\text{opt}}
\]

\[Q_{\text{opt}} = \sqrt{1 + 2|c| \sqrt{\frac{5\gamma}{\delta\alpha^2}} + \frac{5\gamma}{\delta\alpha^2}}\] \hspace{1cm} (4.4)

$\alpha$, $\gamma$ and $\delta$ are bias-dependent parameters, $c$ is the correlation coefficient between the gate noise and drain noise of the MOSFET \[94\].

While for the CG-LNA, the noise figure approximately equals to:

\[
F_{\text{CG}} = 1 + \frac{\gamma}{\alpha}
\]

\hspace{1cm} (4.5)

The noise figure of CG-LNA is constant with respect to $\omega_T/\omega_0$, while that of CS-LNA is linear with $\omega_T/\omega_0$.

The CS-LNA also exhibits inferior reverse isolation and stability due to the Miller effect because the $C_{gd}$ provides a feed-forward path between input and output.

Table 4.1 represents the comparison of the two basic LNA topologies. In Hu’s design \[60\], two CS-LNAs were deployed in order to achieve a broadband frequency response from 3 to 5 GHz. The first LNA was centered at $f_1 = 3.5$ GHz and second one was centered at $f_1 = 4.5$ GHz. This solution did have a good gain and noise performance, however, at the cost of large power consumption (45 mW for RX in \[60\]).

For the design in this dissertation, a single stage CG-LNA is adopted to meet the

<table>
<thead>
<tr>
<th>Specification</th>
<th>CS-LNA</th>
<th>CG-LNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Matching</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>Gain</td>
<td>$\checkmark$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Noise Figure</td>
<td>$\checkmark$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>Power</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>Reverse Isolation</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>Parasitic</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
</tbody>
</table>
critical power constraint. Furthermore, the capacitor cross-coupled gain-boosting technique is used to improve the noise figure and gain performance. Gain-boosting LNA designs have been described in many publications [93,95,96]. Since the difficulty to accurately model the parasitic parameters in radio frequency, most of the designs were based on simplified models and equations. This section is focused on the full modeling construction and optimization by using $g_m/I_D$ methodology in all transistor regions.

4.1.2 Gain-Boosted CG-LNA

The basic idea of gain-boosting is introducing an inverting amplification, $A$, between the source and gate terminals, resulting in the effective transconductance $G_{m,\text{eff}}$ increased to $(1 + A)g_m$, as depicted in Fig. 4.3.

In a differential topology, the inverting gain is naturally available. Fig. 4.4 shows the
complete gain-boosted CG-LNA circuit with capacitive cross-coupling ($C_0$ and $C_1$). In a conventional CG-LNA, the gates are AC shorted to ground. In the gain-boosted CG-LNA, however, the gate of $M_0$, for example, is connected to the source of $M_1$ through $C1$.

As presented in the signal model of $M_0$ in Fig. 4.5, the gate $G_0$ is not AC grounded, thus the gate-source voltage $V_{gs}$ is some voltage ratio derived from $V_{s1} - V_{s2} = -2V_s$, which can be proved to be:

$$\frac{V_{gs}}{-2V_s} = \frac{C_c}{C_c + C_{gs}}$$

Hence the drain current of $M_0$ is now:

$$I_{d0} = -g_m \left[ \frac{2C_c}{C_c + C_{gs}} \right] V_s$$

Thus the effective transconductance is:

$$G_{m,eff} = \left[ \frac{2C_c}{C_c + C_{gs}} \right] g_m = (1 + A_b) g_m$$

where $A_b = (C_c - C_{gs}) / (C_c + C_{gs})$. In this case, $C_c \gg C_{gs}$ resulting $A_b \approx 1$ and $G_{m,eff} \approx 2g_m$.

### 4.1.3 Signal Modeling

The equivalent half circuit model of the differential CG-LNA is presented in Fig. 4.6. It consists of a common gate source-degenerated transistor $M$ with its effective transconductance $G_m = (1 + A_b)(g_m + g_{mb})$ (bulk effect included). The input signal is fed to the source terminal via the source impedance $R_s$ (mostly 50 Ω), the off-chip capacitor $C_{ext}$, the bond wire and the pad. $C_{ext}$ is included to facilitate the input matching condition with feasible low-value source inductor $L_s$. The output stage consists an inductor $L_d$ connected to the power supply and a varactor $C_v$ for inter-stage matching. The load $Z_L$ is the total input impedance of the following stages (IL-VCO and buffer).
Fig. 4.4: Complete gain-boosted CG-LNA.

Fig. 4.5: Effective transconductance calculation of capacitive cross-coupled CG-LNA.
In fact, the parasitic parameters of the pad, bond wire, the package leads and even the PCB (Printed Circuit Board) traces should be considered and modeled to make sure the total input impedance of the LNA, $Z_{in,CG}$, is resistively equal to $R_s$. This can be done by simulating and extracting those parasitics from the Process Design Kit (PDK) or information offered by the foundry. However in this dissertation, only the circuit design considerations are discussed, all other parasitics like bond wire, pad, etc., are out of the scope of the discussion (but they were considered in the actual chip design). Therefore, in this case, $C_{ext}$ is moved forward and included in the input impedance $Z_{in}$ as shown in Fig. 4.6.

The small signal model for half of the differential CG-LNA is depicted in Fig. 4.7. $R_{ps}$ and $R_{pd}$ are the parallel parasitic resistances of $L_s$ and $L_d$, respectively. Note that serial parasitic resistance of $L_s$ should be small in order to limit the power consumed (wasted) by $L_s$. Thus, the $R_{ps}$ is much larger than $R_s$. $Z_{in}$ and $Z_{out}$ are the input and output impedance
of the CG-LNA, respectively. $Z_{d,mos}$ is the output impedance seen into the transistor from the drain, and $Z_d$ is the input impedance of the output stage.

**Input Impedance Match**

The input matching condition is $Z_{in} = R_s + j0$, hence computing $Z_{in}$ we have:

$$Z_{in}^{-1} = j\omega_0 (C_{ext} + C_{gs} + C_{bs}) + \frac{1}{j\omega_0 L_s} + \left[ \frac{1}{R_{ps}} + \frac{1 + G_m r_o}{r_o + Z_d} \right] = \frac{1}{R_s} + j0 \quad (4.9)$$

Solving for $Z_d$ and $C_{ext}$ gives:

$$Z_d = Z_{d,match} = R_s \left( 1 + \frac{G_m}{g_{ds}} \right) \left( 1 - \frac{R_s}{R_{ps}} - \frac{1}{g_{ds}} \right) \quad (4.10)$$

$$C_{ext} = \frac{1}{\omega_0^2 L_s} - C_{gs} - C_{bs} \quad (4.11)$$

Here, $r_o = 1/g_{ds}$, $\omega_0$ is the center frequency of the input signal. $Z_d$ must be fixed to $Z_{d,match}$ in (4.10) to achieve the input matching.

**Output Impedance Match**

Short $V_{in}$ to AC ground, and notice that $(C_{ext} + C_{gs} + C_{bs})$ and $L_s$ have been canceled out at the center frequency $\omega_0$, thus the output impedance is:

$$Z_{out}^{-1} = j\omega_0 (C_v + C_{bd} + C_{gd}) + \frac{1}{j\omega_0 L_d} + \left( \frac{1}{R_{pd}} + Z_{d,mos}^{-1} \right) \quad (4.12)$$

where $Z_{d,mos}$ can be obtained after some calculations:

$$Z_{d,mos} = \frac{1}{g_{ds}} + \left( 1 + \frac{G_m}{g_{ds}} \right) \left( R_s / R_{ps} \right) \quad (4.13)$$

The output impedance matching condition is that $Z_{out}$ is conjugate-matched to the load $Z_L$, i.e. $Z_L = Z_{out}^*$. Therefore, $Z_d$ can be expressed as:
\[
Z_d^{-1} = \frac{1}{j \omega_0 L_d} + j \omega_0 (C_v + C_{bd} + C_{gd}) + \frac{1}{R_{pd}} + (Z_{out}^*)^{-1} 
\] (4.14)

Substituting (4.12) into (4.14) gives:

\[
Z_d^{-1} = \frac{2}{R_{pd}} + Z_{d,mos}^{-1} 
\] (4.15)

Note that \(Z_d\) must be set as \(Z_d,match\) in (4.10) to facilitate the input matching condition, hence we have:

\[
\frac{1}{R_{pd}} = \frac{1}{2} \left( Z_{d,match}^{-1} - Z_{d,mos}^{-1} \right) 
\] (4.16)

**Power Gain**

For convenience, the signal model for the LNA is sketched as shown in Fig. 4.8. Under perfect matching (at \(\omega_0\)), the input capacitance and inductance are canceled out. The input impedance of the input stage \(Z_{in}\) thus equals to \(R_s\), gives \(V_s = (1/2)V_{in}\). Part of the input
current $I_{in}$ is shunted into $R_{ps}$ which is denoted as $I_{ps}$, while the other part is transferred to the output stage ($I_{out}$).

According to the Kirchhoff’s current law (KCL), the $I_{out}$ can be deduced as:

$$I_{out} = \left( 1 - \frac{R_s}{R_{ps}} \right) I_{in} \quad (4.17)$$

The input power is:

$$P_{in} = I_{in}^2 R_s \quad (4.18)$$

At the output stage, $I_{out}$ is separated into $Z_{out}$ and $Z_L$ where $Z_L$ is conjugate-matched to $Z_{out}$ because of output matching. If $Z_{out} = a + jb$, then $Z_L = a - jb$. Only the real part consumes power thus the load power is:

$$P_{out} = a \cdot I_{out}^2 / 4 \quad (4.19)$$

The power gain of the CG-LNA can be written as:

$$G = 10 \log \left( \frac{P_{out}}{P_{in}} \right) = 10 \log \left[ \frac{a}{4R_s} \left( 1 - \frac{R_s}{R_{ps}} \right)^2 \right] \quad (4.20)$$

The real part can be obtained from (4.12) and (4.16):

$$a^{-1} = \frac{1}{R_{pd}} + Z_{d,mos}^{-1} = \frac{1}{2} \left( Z_{d,match}^{-1} + Z_{d,mos}^{-1} \right) \quad (4.21)$$

At last, the noise factor is given by [93]:

$$F = 1 + \frac{\gamma}{\alpha (1 + A_b) G_m R_s} \quad (4.22)$$

4.1.4 $g_m/I_D$ Design Verification

With the obtained input/output impedance matching, power gain and noise factor as functions of the MOSFET model, this section details the $g_m/I_D$ design flow to optimize the
power gain, noise figure within a power constraint.

Before that, minimum transistor length $L_{\text{min}}$ is chosen to be 70 nm in order to maximize the $f_T$. Find a set of $L_s$ with large parallel parasitic resistance in the inductor database. In this section, 10 pairs of $(L_s, P_{ps})$ are selected. The working frequency $f_0$ is 1.6 GHz. $C_v$ can be temporarily out of consideration; it has to be included in the system simulation when the LNA is connected with the following well-designed circuit blocks. The $g_m/I_D$ design algorithm is organized in Algorithm. (1) as below.
Algorithm 1 $g_m/I_D$ Methodology For Gain-boosted Common-gate LNA Design

**Input:** $(L_s, P_{ps})$: A pair from inductor database; an acceptable Power.

**Output:** Power Gain (G).

1: **for** each Power **do**
2:   Compute total current $2I_D$;
3:   **for** $g_m/I_D = 3 : 1 : 24$ **do**
4:     Find the normalized current $i$, compute transistor width $W$;
5:     Compute $C_{ij} = WLC_{ij}$;
6:     Find the output transconductance $g_{ds}$;
7:     Compute $Z_{d,match}$ from (4.10), $C_{ext}$ from (4.11), $Z_{d,mos}$ from (4.13);
8:     Compute $R_{pd}$ from (4.16), $a$ from (4.21);
9:     Find a $L_d$ with the parasitic parallel resistance equal or close to $R_{pd}$, if it does not exit, return an *error* flag;
10:   Compute power gain $G$ from (4.20);
11: **end for**
12: **end for**

Table 4.2 shows three different designs with $g_m/I_D=6, 12, \text{and} 18$, all targeting the power gain of 10 dB. Similar results are appreciated according to the comparison of the Matlab calculations and the SpectreRF simulations. The power gains have less than 20 % difference. Considering the noise figure, Table 4.2 also reveals a better noise performance when moving towards the strong inversion.

<table>
<thead>
<tr>
<th>$gm/Id$ (V$^{-1}$)</th>
<th>$Id$ (mA)</th>
<th>$W$ (µm)</th>
<th>$Ld$ (nH)</th>
<th>G (dB)</th>
<th>NF (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>6.37</td>
<td>2.50</td>
<td>2.49</td>
<td>24.24</td>
<td>24.24</td>
</tr>
<tr>
<td>12</td>
<td>12.43</td>
<td>1.60</td>
<td>1.61</td>
<td>72.56</td>
<td>72.6</td>
</tr>
<tr>
<td>18</td>
<td>18.54</td>
<td>1.05</td>
<td>1.03</td>
<td>180.8</td>
<td>182</td>
</tr>
</tbody>
</table>
SpectreRF simulation results of $S_{11}$, $S_{22}$, power gain, and noise figure (NF) for the three CG-LNA designs are shown in Fig. 4.9.

The variable capacitance $C_v$ can be adjusted off-chip by tuning a control signal $V_{tune}$ from 0-1.2 V to compensate the process variation. The voltage gains with $V_{tune,LNA}=0$ V and 1.2 V are plot in Fig. 4.10. The average power for the LNA is 3.37 mW.

4.2 Injection-Locked VCO

4.2.1 Brief Introduction

Oscillator injection locking is widely used in communication systems such as frequency division [6, 97–99], phase-locked loop [100], quadrature generation [101, 102]. When an external signal is applied to an oscillator, the latter stops to be an autonomous circuit and synchronizes to the external signal. This phenomenon has been investigated by Adler in 1946 [76], and since then by many other authors, more recently in [77, 103–105].

Generally, a free-running oscillator contains a gain block $H(j\omega)$ and a filter $F(j\omega)$ as presented in Fig. 4.11a. According to the Barkhausen criteria [83], the oscillation can be sustained at a given frequency $\omega_0$ if the negative feedback has a loop gain and phase that satisfy:

$$|H(j\omega)F(j\omega_0)| \geq 1$$

$$\angle H(j\omega)F(j\omega_0) = \pi \quad (4.23)$$

Suppose that the gain block is capable of satisfying the gain criteria in 4.23 at any frequency, then the oscillation occurs at $\omega_0$ where the filter satisfies the phase criteria in 4.23. $\omega_0$ is the free-running frequency of the oscillator.

Consider the case in Fig. 4.11b, an external signal $V_{inj}$ at frequency $\omega_{inj}$ is applied to the oscillator. The phase shift across the loop at $\omega_0$ is no longer $\pi$, and the oscillation cannot be at $\omega_0$. However, under some conditions, the phase response of the filter will compensate an extra phase shift, and the Barkhausen’s phase criteria is satisfied at $\omega_{inj}$, instead of $\omega_0$ [77]. We say that the oscillator is injection-locked by $V_{inj}$ and its oscillation
Fig. 4.9: SpectreRF simulation results of $S_{11}$, $S_{22}$, power gain, and noise figure (NF) for the three CG-LNA designs.
frequency moves from the natural frequency $\omega_0$ to the injection frequency $\omega_{inj}$. The range of frequencies $\omega_0 \pm \omega_L$ where synchronization occurs defines the locking range of the oscillator. This is called first-harmonic injection locking [103].

In fact, if the injection signal contains harmonic frequency component that is close to $\omega_0$, the oscillator will oscillate at a fractional or multiple of the input frequency $\omega_{inj}$. To be more clear, a sketch map of the injection-locking phenomenon is shown in Fig. 4.12. $\omega_{out}$ is the output frequency of the oscillator being injection-locked. Note that $n$ is an integer ($n=\{1, 2, 3, \ldots\}$).

- If $\omega_{inj}$ is close enough to $\omega_0$ (in the locking range of the oscillator), the oscillation will be sustained at $\omega_{inj}$ instead of $\omega_0$. This is called first-harmonic injection locking.

- If $\omega_{inj}$ is close enough to $n\omega_0$, or say, $(1/n)\omega_{inj}$ is close enough to $\omega_0$, the oscillation will be sustained at $(1/n)\omega_{inj}$. This is called super-harmonic injection locking and is widely used for frequency dividers [6, 97–99, 103, 106].

- If $\omega_{inj}$ is close enough to $(1/n)\omega_0$, or say, $n\omega_{inj}$ is close enough to $\omega_0$, the oscillation will be sustained at $n\omega_{inj}$. This is called sub-harmonic injection locking which is popular for frequency multipliers [100].
This section presents the analysis and modeling of a first-harmonic injection-locked VCO, and deduce necessary equations for $g_m/I_D$ methodology.

### 4.2.2 Signal Modeling

The detailed view of the IL-VCO is exhibited in Fig. 4.13 [60]. It shows the complementary VCO composed by NMOS $M_5$ and $M_6$, PMOS $M_7$ and $M_8$, with its LC tank which consists of $L$ and a digital controlled capacitor bank $C_{bank}$. In order to differentiate it from the whole IL-VCO for convenience, we give this part an alias—VCO Core, or Core.

$M_1$ and $M_2$ are the injection differential pair with the isolation transistors $M_3$ and $M_4$. The VCO Core is biased at $I_{osc}$ while the injection pair is biased at $I_{inj}$. The bias current $I_{bias}$ is adjusted by an off-chip variable resistor $R_v$. $M_{10}$ and $M_{13}$ exhibit small resistance since their gates are at $V_{DD}$ voltage level. $M_{11}$ and $M_{14}$ act as a MOS capacitance. Together, they form a low-pass filter to suppress noise from the reference current $I_{bias}$.

To simplify the analysis, we assume that the four transistors in the VCO Core are sized so that their transconductances are the same, i.e. $g_{m5,6} = g_{m7,8} = g_m$.

#### Cross-Coupled Structure

The cross-coupled transistors $M_5/M_6$ and $M_7/M_8$ need to be carefully analyzed before
Fig. 4.12: Injection-locking phenomenon: (a) first-harmonic injection locking; (b) super-harmonic injection locking; (c) sub-harmonic injection locking.
Fig. 4.13: Injection-locked VCO.
the complete equations being constructed. It only needs to check the NMOS pair \( M_5 \) and \( M_6 \) since the case is similar to the PMOS complementary \( M_7 \) and \( M_8 \).

Fig. 4.14 shows the NMOS cross-coupled circuit and the modeling simplification steps from the complete small signal model to the reduced model. Since \( M_5 \) and \( M_6 \) are matched, it holds that \( C_{ij5} = C_{ij6} \) in which \( ij = \{gd, db, ds, gb, gs\} \), as well as \( g_{ds5} = g_{ds6} \). Thus, as shown in Fig. 4.14, we have:

\[
C_{t5} = C_{t6} = C_{db5,6} + C_{ds5,6} + C_{gb5,6} + C_{gs5,6} \quad (4.24)
\]

\[
C_{nmos} = 2C_{gd5,6} + \frac{1}{2}(C_{db5,6} + C_{ds5,6} + C_{gb5,6} + C_{gs5,6}) \quad (4.25)
\]

where the \( C_{nmos} \) is the equivalent parasitic capacitances for the NMOS cross-coupled circuit. Similarly, for PMOS cross-coupled circuit:

\[
C_{pmos} = 2C_{gd7,8} + \frac{1}{2}(C_{db7,8} + C_{ds7,8} + C_{gb7,8} + C_{gs7,8}) \quad (4.26)
\]

Note that the final reduced model in Fig. 4.14 contains a negative resistance \(-2/g_m\), thus this cross-coupled circuit is also called “negative-\( G_m \) oscillator” [83]. The analysis based on a negative feedback perspective could also be found in [83].

**Complete IL-VCO Model**

Now the complete model for the \( VCO \) Core can be constructed as shown in Fig. 4.15. The inductor \( L \) is modeled as a ideal inductor \( L_{ind} \) paralleled with its parasitic resistance \( R_{ind} \) (the transconductance form \( g_{ind} = 1/R_{ind} \)). As can be seen, the \( VCO \) Core can finally be modeled as a RLC network: a positive resistance \( 1/g_{core} \), a negative resistance \(-1/g_m\), an ideal inductor \( L_{ind} \) and the parasitic capacitance \( C_{core} \), where:

\[
C_{core} = C_{nmos} + C_{pmos} + C_{bank} \quad (4.27)
\]
Fig. 4.14: Cross-coupled NMOS small signal analysis.
\[ g_{\text{core}} = \frac{1}{2} g_{ds5,6} + \frac{1}{2} g_{ds7,8} + g_{\text{ind}} \]  

(4.28)

Now consider the injection differential pair which consists of \( M_1, M_2, M_3 \) and \( M_4 \). The cascode structure is used for reverse isolation purpose. Intuitively, without \( M_3 \) and \( M_4 \), the output resistances and capacitances seen from the drain of \( M_1 \) and \( M_2 \) will be directly applied to the \( \text{Core} \), and affects the \( \text{Core} \)'s behavior. Stacking the transistors \( M_3 \) and \( M_4 \) on \( M_1 \) and \( M_2 \), respectively, will not only increase the applied resistances, but also decrease the applied capacitances.

To see this, and also to deduce the equations for the \( g_m/I_D \) design flow, we first observe the small signal model of the cascode circuit, which is illustrated in Fig. 4.16. \( C_1 \) includes \( C_{gd1}, C_{bd1}, C_{bs1} \) and \( C_{gs3}. \) \( C_3 \) includes \( C_{bd3} \) and \( C_{gd3}. \) Bulk effect of \( M_3 \) is considered in this case.

In order to calculate the partial impedance \( Z'_{inj} \) shown in the last step in Fig. 4.16, apply a virtual output voltage \( v_o \) and output current \( i_o \) at node \( D_3 \), thus the voltage at node \( D_1 \) is:

\[
V_{D1} = [i_o + (g_{m3} + g_{mb3})V_{D1}] \left( \frac{1}{sC_1}/r_{o1}/ \frac{1}{g_{m3} + g_{mb3}} \right)
\]
\[
= [i_o + (g_{m3} + g_{mb3})V_{D1}] Z_{D1}
\]

(4.29)

Dividing both sides by \( i_o \), and notice that \( Z_{D1}^{-1} = sC_1 + 1/r_{o1} + (g_{m3} + g_{mb3}) \), we can obtain:

\[
\frac{V_{D1}}{i_o} = \frac{1}{Z_{D1}^{-1} - (g_{m3} + g_{mb3})} = \frac{r_{o1}}{1 + sr_{o1}C_1} = r_{o1}/ \frac{1}{sC_1}
\]

(4.30)

Refer again to Fig. 4.16 and take a look at the voltage and current applied to \( r_{o3} \), it holds that:

\[
[i_o + (g_{m3} + g_{mb3})V_{D1}] r_{o3} = v_o - V_{D1}
\]

(4.31)
Fig. 4.15: Complete VCO Core small signal model.

Fig. 4.16: Small signal model for cascode structure.
Dividing both sides by $i_o$ and making some rearrangement, $Z'_{inj} = v_o/i_o$ can be deduced as:

$$\frac{v_o}{i_o} = r_{o3} + \left[1 + (g_{m3} + g_{mb3})r_{o3}\right] \left[\frac{1}{sC_1}\right]$$

$$= r_{o3} + A_c \left[\frac{1}{sC_1}\right]$$

(4.32)

Here, $A_c = 1 + (g_{m3} + g_{mb3})r_{o3}$ is the gain introduced by the cascode structure. $Z'_{inj}$ can be considered as a resistor $r_{o3}$ in series with an parallel network, which consists a resistor $A_c r_{o1}$ and a capacitor $C_1/A_c$. As can be seen, the applied resistances to the Core are increased by a factor of $A_c$, while the capacitances are reduced by a factor of $A_c$, minimizing the impacts from the injection differential part to the VCO Core, and also prevents the large oscillation output signal of the Core feeding back through. Because $M_3$ and $M_4$ are used for isolation purpose, their length could be selected larger in order to increase their output resistance $r_{o3,4}$. Finally, the total output impedance of the cascode circuit $Z_{inj}$ equals to:

$$Z_{inj} = Z'_{inj}/\left(\frac{1}{sC_3}\right)$$

(4.33)

To sum up, the complete small signal model of the whole IL-VCO is illustrated in Fig. 4.17. Since $M_3$ and $M_4$ are biased in the triode region, $r_{o3}$ and $r_{o4}$ are small thus can be ignored for simplicity. $C_T$ and $g_T$ are the total parasitic capacitance and total positive transconductance of the IL-VCO which are given by:

$$\begin{align*}
C_T &= C_{core} + 2C_{3,4} + \frac{2}{A_c}C_{1,2} \\
g_T &= g_{core} + \frac{2}{A_c}g_{ds1,2}
\end{align*}$$

(4.34)

where $C_{1,2} = C_{gd1,2} + C_{bd1,2} + C_{bs1,2} + C_{gs3,4}$, $C_{3,4} = C_{bd3,4} + C_{gd3,4}$. 
Fig. 4.17: Complete small signal model of IL-VCO.

**Oscillation Start-up Criteria**

For oscillation built-up, the absolute value of the positive resistance should be larger than the negative one [83], i.e.:

\[ g_m \geq g_T \]  \hspace{1cm} (4.35)

In order to ensure oscillation in the presence of temperature and process variations, typically the \( g_m \) is selected larger than the theoretical criteria shown in (4.35) [83]. A safety margin could be set by defining a coefficient \( k_{osc} \) to transform the inequality [107]:

\[ g_m = k_{osc} g_T \]  \hspace{1cm} (4.36)

where \( k_{osc} \) is generally in the range of 1.5-3.

The MOSFET intrinsic gain \( A_i \), as defined in (3.7) in Chapter 3, is rewritten as below:

\[ A_i = g_m r_o = \frac{g_m}{g_{ds}} = \frac{g_m/|I_D|}{g_{ds}/|I_D|} \]  \hspace{1cm} (4.37)

Suppose \( g_{m3,4} = \eta g_{m3,4} \), the \( g_T \) can thus be transformed into:
\[ g_T = g_{core} + \frac{2}{A_c} g_{ds1,2} \]

\[ = \frac{1}{2} g_{ds5,6} + \frac{1}{2} g_{ds7,8} + g_{ind} + \frac{2}{1 + (g_{m3,4} + g_{mb3,4}) r_{o3,4}} g_{ds1,2} \]

\[ = \frac{1}{2} g_{ds5,6} + \frac{1}{2} g_{ds7,8} + g_{ind} + \frac{2}{1 + (1 + \eta) g_{m3,4} / g_{ds3,4}} g_{ds1,2} \]  
(4.38)

\[ = g_{ind} + \frac{g_m}{2} \left( \frac{1}{A_{i5,6}} + \frac{1}{A_{i7,8}} \right) + \frac{2g_{m1,2}}{A_{i1,2} [1 + (1 + \eta) A_{i3,4}]} \]  
(4.39)

Substituting (4.36) into (4.39), we can get:

\[ g_{ind} = \beta_1 g_m + \beta_2 g_{m1,2} \]  
(4.40)

where

\[ \beta_1 = \frac{1}{k_{osc}} - \frac{1}{2A_{i5,6}} - \frac{1}{2A_{i7,8}} \]  
(4.41)

\[ \beta_2 = \frac{-2}{A_{i1,2} [1 + (1 + \eta) A_{i3,4}]} \]  
(4.42)

Hence we have:

\[ g_m = \frac{1}{\beta_1} (g_{ind} - \beta_2 g_{m1,2}) \]  
(4.43)

The free-running oscillation frequency of the IL-VCO is:

\[ f_{vco} = \frac{1}{2\pi \sqrt{L_{ind} C_T}} \]  
(4.44)

The locking range of a sine wave IL-VCO is described in [76, 77]:

\[ \omega_L = \frac{\omega_{out}}{2Q} \frac{I_{inj}}{I_{osc}} \frac{1}{\sqrt{1 - \frac{I_{inj}^2}{I_{osc}^2}}} \]  
(4.45)
where $Q$ represents the quality factor of the tank. Suppose the injection occurs once every $N$ cycles, the locking range needs to be modified as [103]:

$$\omega_L = \frac{\omega_{out}}{2Q} \cdot \frac{I_{inj}}{I_{osc}} \cdot \frac{1}{N} \cdot \sqrt{1 - \frac{I_{inj}^2}{N^2 I_{osc}^2}} \simeq \frac{\omega_{out}}{2Q} \cdot \frac{I_{inj}}{I_{osc}} \cdot \frac{1}{N}$$

(4.46)

where the total quality factor $Q$ of the tank is given by:

$$Q = \frac{1}{2\pi f_0 \cdot g_T \cdot L_{ind}}$$

(4.47)

In our case, $f_0 = 1.6$ GHz and the target data rate is $DR = 100$ Mbps. Therefore, $N$ can be calculated by:

$$N = \frac{1}{DR} \cdot \frac{1}{f_0} = \frac{f_0}{2 \cdot DR} = 8$$

(4.48)

A short discussion is worth taken to bring the results of the above analysis back into perspective. The VCO Core takes the responsibility to oscillate in a quiescent state. The Core is an RLC lossy tank, and the cross-coupled NMOS/PMOS pairs supply the energy to start and maintain oscillation by a form of, intuitively, a negative resistance $-1/g_m$. In the above analysis, all the capacitances and inductance are modeled for the oscillation frequency estimation, and the transconductances, or resistances, are modeled for the lossy parts which are the actual energy consumers. $g_m$ value should be “strong” enough to cover the loss introduced by the resistances. The injection part should be simultaneously considered when designing the VCO Core because of the unignorable parasitic capacitances of $M_3$ and $M_4$. However, the impacts from $M_1$ and $M_2$ are well suppressed and could be neglected because of the $A_c$ introduced by the cascode structure.

### 4.2.3 $g_m/I_D$ Design Verification

With the analytical expressions deduced above, the $g_m/I_D$ design flow can be constructed. Firstly, set the minimum transistor length $L_{min} = 70$ nm for $M_1, M_2, M_5, M_6,$
$M_7$, and $M_8$, set the transistor length $L_{\text{min}} = 250 \text{ nm}$ for $M_3$ and $M_4$, and define the working frequency $f_0$ is 1.6 GHz. Secondly, pick up a high-Q inductor $L_{\text{ind}}$ in the database, and find its parallel equivalent resistance $R_{\text{ind}}$ hence get its $g_{\text{ind}} = 1/R_{\text{ind}}$. Finally, suppose $I_{\text{osc}} = 2I_D$, and $I_{\text{inj}} = \epsilon I_{\text{osc}}$ where $\epsilon > 1$. The design algorithm is organized in Algorithm 2 as below.

**Algorithm 2** $g_m/I_D$ Methodology For Injection-locked VCO Design

**Input:** Transistor length; high-Q $L_{\text{ind}}$; $k_{\text{osc}} = 2$

**Output:** Tunable oscillation centered at $f_0 = 1.6 \text{ GHz}$; Power consumption;

1: for each $I_D = (1 : 1 : 10) \times 50 \mu\text{A}$ do;
2: for each $(g_m/I_D) = 3 : 1 : 24$ do
3: Find normalized current of $i_{5,6}$ and $i_{7,8}$;
4: Find $(g_{ds}/I_D)_{5,6}$ and $(g_{ds}/I_D)_{7,8}$;
5: Compute intrinsic gain $A_{i5,6}$ and $A_{i7,8}$ from (4.37);
6: Find normalized capacitance $C_{ij5,6}$ and $C_{ij7,8}$;
7: for each $\epsilon = 1.5 : 0.5 : 5$ do
8: for each $(g_m/I_D) = 3 : 1 : 24$ do
9: Find normalized current of $i_{1,2}$ and $i_{3,4}$;
10: Compute $g_{m1,2}$ and $g_{m3,4}$;
11: Find $g_{m3,4}$ and compute $\eta = g_{m3,4}/g_{m3,4}$;
12: Find $(g_{ds}/I_D)_{3,4}$;
13: Compute $A_{3,4}$ from (4.37);
14: Find normalized capacitance $C_{ij1,2}$ and $C_{ij3,4}$;
15: Compute $\beta_2$ from (4.42);
16: Compute $g_m$ from (4.43);
17: Compute transistor width $W_{5,6}$ from $i_{5,6}$ and $I_D$;
18: Compute transistor width $W_{7,8}$ from $i_{7,8}$ and $I_D$;
19: Compute transistor width $W_{1,2}$ from $i_{1,2}$ and $\epsilon I_D$;
20: Compute transistor width $W_{3,4}$ from $i_{3,4}$ and $\epsilon I_D$;
21: Compute $C_{ij1,2}$, $C_{ij3,4}$, $C_{ij5,6}$ and $C_{ij7,8}$;
22: Compute $A_c$ from (4.32);
23: Compute $C_T$ from 4.44, $C_{\text{core}}$ from (4.34);
24: Compute $C_{nmos}$ from (4.25), $C_{pmos}$ from (4.26);
25: Compute $C_{\text{bank}}$ from (4.27);
26: Compute Power=$V_{DD} \times (I_{\text{bias}} + 2I_D + 2\epsilon I_D)$;
27: end for
28: end for
29: end for
30: end for
Comparing the Matlab calculations and the SpectreRF simulations, similar design parameters are obtained with small differences, as shown in Table 4.3. The simulated average power is 1.21 mW.

Table 4.3: MATLAB CALCULATIONS AND SPECTRERF SIMULATIONS COMPARISON FOR THE IL-VCO.

<table>
<thead>
<tr>
<th>Transistors</th>
<th>$g_m/I_D$ (V$^{-1}$)</th>
<th>Width (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calc.</td>
<td>Sim.</td>
</tr>
<tr>
<td>$M_1, M_2$</td>
<td>15</td>
<td>16.4</td>
</tr>
<tr>
<td>$M_3, M_4$</td>
<td>10</td>
<td>9.1</td>
</tr>
<tr>
<td>$M_5, M_6$</td>
<td>20</td>
<td>21.8</td>
</tr>
<tr>
<td>$M_7, M_8$</td>
<td>20</td>
<td>18.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$I_{osc}$ (µA)</th>
<th>$I_{inj}$ (µA)</th>
<th>Average Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>92</td>
<td>1000</td>
</tr>
</tbody>
</table>

Without an injected signal, the IL-VCO runs at its free-running frequencies which can be coarsely tuned by the 4-bit capacitor bank (shown in the next section). When the controlling bits $b_3 b_2 b_1 b_0$ change from 0000 to 1111, the IL-VCO runs freely at totally 16 sub-bands, as shown in Fig. 4.18.

When a sinusoid signal is injected, the IL-VCO will automatically be locked at the frequency of the signal. If the injected signal is a carrier-based pulse chain, the IL-VCO transits between locked and unlocked status, as shown in Fig. 4.19, leading to a higher phase noise, i.e., a higher clock jitter. Fig. 4.20 presents a locking range of 1.52-1.66 GHz of the IL-VCO. The average power of the IL-VCO is 1.61 mW.

4.2.4 Digital Controlled Capacitor Bank

When designing a VCO with wide tuning range, it is better to divide the full tuning range into multiple overlapping bands. In this case, binary control signals are used to select the band of interest (coarse tuning), and analog signal fine-tunes the free-running oscillation frequency in the selected band.
Fig. 4.18: IL-VCO free-running status: transient signal and coarse tunable frequencies from $b_3b_2b_1b_0 = 0000$ to $b_3b_2b_1b_0 = 1111$. 
Fig. 4.19: IL-VCO running status with injected signal.

Fig. 4.20: IL-VCO locking range 1.52-1.66 GHz.
To implement this scheme, a bank of 4-bit binary-weighted switched capacitors ($C_d$) and a small differential accumulation mode varactor ($C_v$) are employed to realize the $C_{\text{bank}}$ computed previously, as shown in Fig. 4.21. The capacitors $nC_d$ ($n = 1, 2, 4, 8$) are connected to each output node and are switched to ground by switches $M_{pd}$. Each switch contributes additional loss to the tank due to its finite resistance. Thus, minimum-length NMOS devices are utilized and made as wide as can be tolerated with regards to the resulting parasitic drain-to-bulk capacitance, which ultimately limits the achievable tuning range [108].

### 4.3 Polyphase Filter

In order to control the clock phase, quadrature signals are required. One way to generate in-phase and quadrature signals is using a VCO operating at twice the frequency of interest and divide-by-2 circuit, performed either in the digital or in the analog domain [109]. Fig. 4.22 represents a basic implementation with two master-slave flip-flops. The outputs of a set of divide-by-2 flip-flops are triggered by opposite phases of a 50% duty cycle clock, the outputs are in quadrature but at half the clock frequency. Important drawbacks of this scheme are the significant increase in power consumption because of the high-frequency-running flip-flops, and the need for accurate 50% duty cycle in the oscillator output.

Another way to obtain quadrature signals through the use of a VCO design capable of directly delivering such signals. A ring oscillator fulfills the requirement [110]. However, not to say the notorious high phase noise of the ring oscillator, an additional oscillator is obviously not acceptable because of the constraint power limit. A third way for quadrature generation is coupling two symmetric LC-VCOs to each other, as exemplified in Fig. 4.23, the combination of a direct connection and a cross (inverting) connection forces the two VCOs to oscillate in quadrature [111]. As mentioned, any additional oscillator will significantly give more burden on the power budget.

An attractive method to generate quadrature signals is to use $RC - CR$ network, or polyphase filter, as presented in Fig. 4.24 [112–114]. The balanced quadrature signals are directly generated from the differential signal ($R_i = R$, $C_i = C$, $i = 1, 2, 3, 4$).
Fig. 4.21: Capacitor bank.
Fig. 4.22: Quadrature generation by divide-by-2 flip-flops.

Fig. 4.23: Basic digram for quadrature VCO.

One stage polyphase filter exhibits quadrature phase shift only around the pole frequency, \( f_p = 1/2\pi RC \). To achieve a broadband response, several stages must be cascaded. The expense of the multi-stage polyphase filter is its lossy characteristic. When one stage loads the output without buffering, the resulting voltage division by 2 lowers the 3 dB of gain, i.e. 3 dB loss for one stage [113]. Therefore, amplifier(s) is required for multistage polyphase filter [113].

Phase inaccuracy results in larger phase sweeping steps for the phase shifter in a specific phase range, thus leading to worse phase align between the clock and data, degrading the BER performance. However, in our case, very accurate quadrature phases are not necessary. As long as the phase shifter can tolerate the 3 dB loss of I/Q signals, the polyphase filter can be directly connected to the phase shifter (AC coupled) without the additional amplifier(s).

In one stage, the phase error between the I/Q signals is given by [112]:

\[
\Delta \varphi = -2 \arctan (\omega_0 RC)
\] (4.49)
For the frequency variation of $1.6 \text{GHz} \pm 150 \text{MHz}$, the phase error is $5^\circ$ which is acceptable.

A big challenge is that, the polyphase filter is sensitive to the $RC$ mismatch. Assuming a relative resistor mismatch of $\alpha$ and capacitor mismatch of $\beta$, the phase shift $\Delta \varphi$ can be expressed as:

$$
\Delta \varphi = \frac{\pi}{2} - \left[ \arctan \left( \omega_0 RC \left( 1 + \alpha \right) \left( 1 + \beta \right) \right) - \arctan \left( \omega_0 RC \right) \right] \\
\simeq \frac{\pi}{2} - \frac{\alpha + \beta}{2}
$$

(4.50)

Therefore, for the resistor and capacitor matching of 1%, the phase error would be $0.6^\circ$. If the mismatch is 10%, the phase error would be $5^\circ$.

Careful layout is very important to the polyphase filter. Although common-centroid structures can alleviate mismatch to a certain extent, special attention is still necessary for the process-sensitive block. Large component results in better matching for both the
resistors and capacitors. Resistors achieve better matching, and the dominant source of performance degradation is the capacitor matching [115]. Also, large resistor values will result in larger input impedance, hence result in less loading of the IL-VCO resonant tank. Therefore, the resistor values are set at the largest possible values, and the capacitors are designed to match the frequency requirement.

A layout sketch for the polyphase filter is shown in Fig. 4.25. Each resistor \( R_r \) \((r = 1, 2, 3, 4)\) are separated into four identical ones: \( R_{ia}, R_{ib}, R_{ic} \) and \( R_{id} \). Those that are without any notations are the dummy resistors.

### 4.4 Phase Shifter

The phase shifter uses a current-steering digital-to-analog converter (DAC) that supplies tail current to the two differential pairs while sharing the same resistive loading [116]. A simplified schematic of the phase shifter is depicted in Fig. 4.26. \( I_1 \) and \( I_2 \) are complementary current supply, i.e. the sum of them is a constant: \( I_1 + I_2 = I \). All the four transistors are in the same size and in the saturation region. Suppose \( I_1 = \eta I \), in which \( \eta \) is some ratio between 0 and 1, thus we have \( I_2 = (1 - \eta)I \). Further suppose the signal amplitude is \( A \), signal frequency is \( \omega \) and the phase of \( I_+ \) is \( \alpha \), we can have:

\[
\begin{align*}
I_+ &= A \sin(\omega t + \alpha + 0) \\
I_- &= A \sin(\omega t + \alpha + \pi) \\
Q_+ &= A \sin(\omega t + \alpha + \frac{\pi}{2}) \\
Q_- &= A \sin(\omega t + \alpha + \frac{3\pi}{2})
\end{align*}
\]

According to the first-order squared-law, the transconductance of the four transistors are:

\[
\begin{align*}
g_{m1,2} &= \sqrt{2\mu_n C_{ox}} \frac{W}{L} \left( \frac{1}{2} \eta I \right) = K\sqrt{\eta I} \\
g_{m3,4} &= \sqrt{2\mu_n C_{ox}} \frac{W}{L} \left( \frac{1}{2} (1 - \eta) I \right) = K\sqrt{(1 - \eta) I}
\end{align*}
\]

The output voltage of \( V_{o1} \) can be obtained as:
Fig. 4.25: Layout sketch for the polyphase filter.
Fig. 4.26: Phase shifter working principle.

\[
V_{o1} = R \left[ g_{m1,2} \cdot (I_+) + g_{m3,4} \cdot (Q_+) \right]
\]
\[
= R \left[ K \sqrt{\eta} \cdot A \sin (\omega t + \alpha) + K \sqrt{(1 - \eta)} \cdot A \sin \left( \omega t + \alpha + \frac{\pi}{2} \right) \right]
\]
\[
= RAK \sqrt{I} \left[ \sqrt{\eta} \sin (\omega t + \alpha) + \sqrt{(1 - \eta)} \cos (\omega t + \alpha) \right]
\]
\[
= RAK \sqrt{I} \sin (\omega t + \alpha + \Delta \varphi)
\]

(4.53)

where

\[
\Delta \varphi = \arctan \sqrt{\frac{1 - \eta}{\eta}}
\]

(4.54)

Similarly, \( V_{o2} \) can also be expressed as:
\[ V_{o2} = R \left[ g_{m1,2} \cdot (I_-) + g_{m3,4} \cdot (Q_-) \right] \]
\[ = R \left[ K \sqrt{\eta I} \cdot A \sin (\omega t + \alpha + \pi) + K \sqrt{(1-\eta)I} \cdot A \sin (\omega t + \alpha + \frac{3\pi}{2}) \right] \]
\[ = RAK \sqrt{I} \left[ -\sqrt{\eta} \sin (\omega t + \alpha) - \sqrt{(1-\eta)} \cos (\omega t + \alpha) \right] \]
\[ = -RAK \sqrt{I} \sin (\omega t + \alpha + \Delta \phi) \quad (4.55) \]
\[ = -V_{o1} \quad (4.56) \]

As can be seen, \( \Delta \phi \in [0, \pi/2] \) if \( \eta \in [0, 1] \), i.e., choosing different current weight \( \eta \) results in a phase shift in between the \( I_+ \) and \( Q_+ \).

The complete phase shifter circuit is shown in Fig. 4.27. \( I \) and \( Q \) are the input I/Q signals. \( M_{I+}, M_{I-}, M_{Q+}, \) and \( M_{Q-} \) are switches controlled by \( I_{\text{sel}} \) and \( Q_{\text{sel}} \) which are digital signals in order to select the quadrant, thus the phase can be shifted from 0 to 360°. The two identical DACs are binary-weighted which are controlled by the 7-bit bus \( A[6 : 0] \). The phase shift in one quadrature \( (I_{\text{sel}}=0, Q_{\text{sel}}=0) \) is presented in Fig. 4.28 with an average power of 1.52 mW.

### 4.5 Frequency Divider

#### 4.5.1 Brief Introduction

Frequency dividers are widely used in high-speed communication circuitry and have been the subject of extensive study [76]. In the digital domain, a binary counter can be used for power-of-2 integer division, clocked by the input signal [117]. An arrangement of flip-flops is a classic method for integer-n division. Shifter register network that is clocked by the input signal is another popular structure. The last register’s complemented output is fed back to the first register’s input. The digital realization for the frequency divider is convenient to design, robust to process variation, but suffers from comparatively high power consumption and is hard to process RF signals.
Fig. 4.27: Complete phase shifter circuit.
Therefore, analog frequency divider is widely used in RF domain. A very conventional architecture is regenerative frequency divider, also known as Miller frequency divider [5, 118, 119]. As shown the basic structure in Fig. 4.29 [5], Miller frequency divider contains a mixer to produce new signals which contain the frequency components of \( f_{in} + f_{out} \) and \( f_{in} - f_{out} \). If \( f_{out} = 1/2f_{in} \), it can realize divide-by-2 frequency division after filtering the high frequency part by the low pass filter (LPF). However, mixer and LPF are power consuming. One such design for an OFDM UWB system in [5] cost 47 mW in 0.18-\( \mu \)m CMOS process.

![Mixer and LPF diagram](image)

**Fig. 4.29:** Regenerative (Miller) divider [5].
Another popular structure is based on divide-by-2 current-mode logic (CML) D flip-flops (DFFs), as presented in Fig. 4.30 \[106,120\]. For each CML, the \(Q\) terminals are connected back to the \(D\) terminals in reversed polarity, resulting in a divide-by-2 operation. Four stages are cascaded to implement the 16 division ratio. As can be seen, if the input frequency is 1.6 GHz which is in our case, the first CML toggles at 800 MHz. The second, third and fourth CML toggles at 400 MHz, 200 MHz and 100 MHz, respectively. Thus, the maximum input frequency of the entire frequency divider is limited by the maximum toggling frequency of the first CML in the chain. High-speed CML dividers are usually power hungry because the CML of the first stage has to handle very high frequencies \[106,120\].

As was discussed in Section 4.2 (IL-VCO), the super-harmonic injection locking phenomenon can be used as frequency dividers \[6,97–99,103,106\]. This is a popular structure since it cost very less power. The injection-locked frequency divider (ILFD) can be based on either LC-tank oscillator (\(LC\)-Type) \[6,97,103\] or ring oscillators (\(Ring\)-Type) \[98,99,106\]. However, \(LC\)-Type divider, as shown in Fig. 4.31 \[6\], suffers from large die area because of the passive inductor. The \(Ring\)-Type divider, which includes divide-by-2 circuits based on DFFs with negative feedback, is robust to process variation and has a wide frequency locking range, leading to its dominant position in industry designs \[120,121\].

A fully differential, divide-by-8 ILFD based on ring oscillator proposed by Cheng et al. \[106\] gives a tuned locking range of 4-18 GHz with a power of 3.6 mW in 0.18-\(\mu\)m CMOS.

Fig. 4.30: Conventional divide-by-16 CML frequency divider.
technology. This section presents a modified divide-by-16 ILFD based on Cheng’s design and solves correlated equations and steps for $g_m/I_D$ methodology to optimization the divider.

### 4.5.2 Signal Modeling

A modified divide-by-16 frequency divider is used in this dissertation, based on Cheng’s design, is shown in Fig. 4.32. It consists of an eight-stage ring of latches, each of which is a CML D-latch. The output of the last latch (CML8) is connected to the input of the first latch (CML1) with inverted polarity to achieve extra phase-shift of 180°. The clock terminals of the eight CMLs are tied together and used to inject the differential input signal which is previously buffered after the phase shifter. The output signal can be taken from any of the eight $Q_+/Q_-$ terminals.

The modified resettable CML is shown in Fig. 4.33. It is a CML latch with PMOS $M_9$ and $M_{10}$ operating in triode region. When $RST$ is low (no reset), the CML works as a DFF. That is, when ($CK_+ = 1, CK_- = 0$), $M_7$ is ON and $M_8$ is OFF; and hence
the CML behaves as a sensing differential amplifier with $M_1$, $M_2$, $M_9$ and $M_{10}$ devices. When $(CK_+ = 0, CK_- = 1)$, $M_7$ is OFF and $M_8$ is ON; and hence the CML behaves as a differential pair with positive feedback, $M_3$ and $M_4$. The positive feedback latches the output signal $(Q_+/Q_-)$ at their current level.

When reset signal $RST$ is high, $M_5$ pulls $Q_+$ to an inverse voltage potential to $CK_+$ ($Q_+ = 0$ if $CK_+ = 1$, $Q_+ = 1$ if $CK_+ = 0$). Therefore, the latch stores a static state according to the input clock signal. Only one of the CMLs needs reset input (CML1) which is controlled off-chip, other reset inputs are OFF for CML2 to CML8 ($RST=$Ground and $RST=VDD$, permanently).

The speed of this CML is determined by the bias current of $M_{11}$ together with the load resistance and capacitance seen at the output nodes, $Q_+$ and $Q_-$. A higher current value can charge the load quickly, hence it allows higher frequency operation.

Fig. 4.34 illustrates the timing diagram of the input $CK$ and output $Q_1$ to $Q_8$. The transition edges of $Q_1$ to $Q_8$ are delayed by $t_p$ from the rising edge of $CK$. $t_p$ is the propagation time for the CML to toggle state after when $CK$ is high.
Fig. 4.33: Modified resettable CML circuit.
Fig. 4.34: Timing diagram of the divide-by-16 frequency divider.

Suppose that the $CK$ keeps at high voltage level all the time, the divider works as a ring oscillator with the free-running frequency at:

$$f_0 = \frac{1}{2N t_p} \quad (4.57)$$

where $N$ is the number of stages. As shown in [122], the free-running frequency of a $N$-stage ring oscillator is given by:

$$f_0 \approx \frac{1}{2NR_L C_L \ln 2} \quad (4.58)$$

where $R_L$ and $C_L$ are respectively the equivalent resistance and capacitance at the output of each CML cell. Hence we have:

$$t_p = R_L C_L \ln 2 \quad (4.59)$$

Further suppose the input clock ($CK$) frequency is $f_{ck}$, hence the period is $T_{ck} = 1/f_{ck}$. In the first half period of $CK$, it must leave an enough time space (at least $t_p$) for $Q_1$ to
toggle its state, as shown in Fig. 4.35a. This gives:

\[
\frac{T_{ck}}{2} > t_p
\]  

(4.60)

Fig. 4.35: Zoomed-in timing diagrams: (a) minimum requirement for \( T_{ck} \); (b) maximum requirement for \( T_{ck} \).

At the other hand, \( T_{ck} \) cannot be too long, otherwise \( Q_2 \) would have time to toggle its state. In other words, \( Q_1 \) cannot propagate to the next stage in half of the clock period as presented in Fig. 4.35b. That is:

\[
\frac{T_{ck}}{2} < 2t_p
\]  

(4.61)

From (4.60) and (4.61), \( t_p \) must satisfy:

\[
\frac{1}{4f_{ck}} < t_p < \frac{1}{2f_{ck}}
\]  

(4.62)
Substituting (4.59) into (4.62) gives:

\[
\frac{1}{(4 \ln 2) f_{ck}} < R_L C_L < \frac{1}{(2 \ln 2) f_{ck}} \tag{4.63}
\]

Notice that \(M_3\) and \(M_4\) are cross-coupled NMOS pair whose model has been detailed in Fig. 4.14. Considering the \(M_1\) and \(M_9\) as shown in Fig. 4.36 and substituting (4.25), the \(R_L\) and \(C_L\) for the CML circuit can be written as:

\[
R_L^{-1} = R_{1,2} + g_{ds1,2} + g_{ds3,4} - g_{m3,4} \tag{4.64}
\]

\[
C_L = (C_{gd1,2} + C_{bd1,2}) + (4C_{gd3,4} + C_{ds3,4} + C_{gb3,4} + C_{gs3,4}) \tag{4.65}
\]

The buffers shown in Fig. 4.32 is a complementary self-biased differential amplifier (CSDA) [7] which is presented in Fig. 4.37.

In the CSDA, transistors \(M_5\) and \(M_6\) operate in the linear region. Thus the output swing of the amplifier can be very close to the difference between the two supply rails. This large output swing makes interfacing the CSDA to the CMOS logic gates straightforward since it provides a large margin for variation in the logic threshold of the gates [7]. Furthermore, the linear-region operation of \(M_5\) and \(M_6\) can provide output switching currents momentarily large while keeping its quiescent current very small. This feature makes CSDA can fast charge and discharge the output capacitive loads without consuming a large amount of power.

With the resulting set of equations derived, it is now to present the \(g_m/I_D\) design flow. The transistor length for \(M_i\) (i=1,2,...,8) is set as the minimum \(L_{min} = 70\, \text{nm}\). Suppose \(g_{m1,2,3,4} = g_m\) and \(I_{bias} = 2I_D\), the algorithm is presented in Algorithm. (3). The output clock generated from the frequency divider at different input frequencies are presented in Fig. 4.38 with the average power shown in Fig. 4.39.
Fig. 4.36: Small signal model to calculate $R_L$ and $C_L$ for the CML circuit.

4.6 Mixer

The mixer basically includes analog switches $SW_1$, $SW_2$, $SW_3$ and $SW_4$ which are composed of a NMOS and a PMOS transistor, as presented in Fig. 4.40. A IL-VCO in the TX gives a constant sinusoidal carrier wave at the frequency of 1.6 GHz. When $Data=1$, $SW_1/SW_2$ are ON while $SW_3/SW_4$ are OFF, the output signal $V_{op}$ and $V_{on}$ are differential sinusoidal waves with a width equals to the $Data$ width. When $Data=0$, $SW_1/SW_2$ are OFF while $SW_3/SW_4$ are ON, the output signal $V_{op}$ and $V_{on}$ are at a DC potential biased by the signal $Bias$. Capacitors $C_1$ and $C_2$ give a AC path to ground in order to suppress the fed through signal when $SW_1/SW_2$ are OFF.

4.7 Error Correction Encoder and Decoder

The design of low-power error correction decoder is crucial in this system. Convolutional Viterbi decoder (VD) is chosen since its high-speed and low-complexity features. Usually, there are two methods used to extract the decoded bits: the trace-back (TB) and
Fig. 4.37: Buffer: a complementary self-biased differential amplifier (CSDA) [7].
Fig. 4.38: Output clock from the frequency divider.
Fig. 4.39: Average power for the frequency divider.

Fig. 4.40: Complete mixer circuit.
Algorithm 3 \( g_m/I_D \) Methodology For Injection-locked Frequency Divider Design

**Input:** Current \( I_D \);
**Output:** All transistor dimensions; Power.

1: \textbf{for} each \( I_D = (1 : 1 : 15) \times 20 \mu A \) \textbf{do}
2: \hspace{1em} \textbf{for} each \( g_m/I_D = 3:1:24 \) \textbf{do}
3: \hspace{2em} Compute \( g_m3,4 \);
4: \hspace{2em} Find normalized current \( i_n = i_{1,2,3,4} \);
5: \hspace{2em} Find \( (g_{ds}/I_D)_{1,2,3,4} \) and compute \( (g_{ds})_{1,2,3,4} \);
6: \hspace{2em} Find normalized capacitance \( C_{ij1,2,3,4} \);
7: \hspace{2em} Compute transistor width \( W_n = W_{1,2,3,4} \) from \( i_n \) and \( I_D \);
8: \hspace{2em} Compute \( C_{ij1,2,3,4} \);
9: \hspace{2em} Compute \( C_L \) from (4.65);
10: \hspace{2em} Compute \( R_L \) from (4.64);
11: \hspace{2em} Compute \( R_L C_L \). If it does not satisfy (4.63), return \textit{error}.
12: \hspace{2em} Compute total power = \( V_{DD} \times (8 \times 2I_D) \).
13: \hspace{1em} \textbf{end for}
14: \textbf{end for}

the register-exchange (RE) \cite{123}. The TB method is acceptable for trellis with a large number of states, whereas the RE approach is more suitable for trellis with a small number of states. In this work, a low-complexity encoder with constraint length \( K = 3 \), code rate \( R = 1/2 \) and generator polynomial \( G(x) = [1 + x^2 \ 1 + x + x^2] \) is adopted, thus the RE-based Viterbi algorithm (VA) is chosen for decoding process. The encoder has two memory units and generates two encoded bits for each incoming data bit, i.e., only one parity-check bit is added for each data bit.

The trellis diagram of the encoder is depicted in Fig. 4.41. The state is indicated as a pair of bits, with the first bit representing the least significant bit (lsb) while the second representing the most significant bit (msb). The transitions between two states are represented by branches which are along the state diagram indicating input/output values. In VA, the branch is assigned a weight, referred to as branch metric (BM) that are accumulated along the path, forming a path metric (PM). When two branches enter the same state, the branch with the smaller PM survives while the other one is discarded. For simple hard decision, the weight is the Hamming distance (number of bits differ) between the encoded bits and the received bits.

The units for calculating the BMs and PMs are required for both TB and RE methods.
Fig. 4.41: Trellis diagram for the encoder with constraint length $K = 3$, generator polynomial \( G(x) = [1 + x^2, 1 + x + x^2] \).

The difference is the methodology of extracting the decoded output bits. In TB method, the decoded bit is extracted by tracking backward from the state with minimum PM at Stage \( t + \Gamma \) (\( \Gamma \) is the decoding depth), by following the survivor path, to the original state \( t \). It is necessary to trace backward through the trellis once for each output. In RE approach, a register is assigned to each state. The register contains the decoded output sequence along the path from initial state \( t \) to the final state \( t + \Gamma \). The decoded output sequence is selected in the register assigned to the state with the minimum PM (double buffering of the data may be necessary so that the contents in registers are not lost during the copying from one state to another [123]). Since the RE method does not need to be traced back, it is faster than the TB method.

The VD composed of four functional units is shown in Fig. 4.42. The BM unit (BMU) calculates the BMs. The add-compare-select unit (ACSU) adds the BMs to the PMs along the path, compares the PMs, and then stores the minimum PMs in the PM memory unit (PMU). Meanwhile, the ACSU appends the associated decoded bit in the survivor-path memory unit (SMU). All the units are driven by the CONTROL block. It should be noted that it is possible to make an incorrect decoding decision on a finite decoding depth, which is called a truncation error. The truncation error is typically small enough to ensure
negligible performance degradation if a decoding depth of about four to five constraint length is employed [124, 125]. The decoding depth $\Gamma$ is $12 (4 \times K)$ in this design with the tradeoff of power and performance.

The power estimation method, which has been measured and verified with the absolute modeling error of $5.2\%$ and standard deviation of $6.6\%$ [126], is adopted for estimating the power dissipation of the VD. The VD is designed by Verilog hardware description language (Verilog-HDL). Gate mapping is carried out by a synthesis tool (Cadence RTL Compiler). The delay information of the circuit is written in a standard delay format (.sdf) file. The HDL simulator, Cadence SimVision, is used to create the value change dump (.vcd) file. Here, the .vcd file and the .sdf file are used to carry out cycle-accurate power simulation. By post-processing the data, the energy distribution can be extracted [126]. The estimated power of the entire VD is $51.3\mu W$ at the speed of 100 Mbps in this 65-nm CMOS process.

The decoder creates a speed limit for the system. It should be noted here that this Viterbi decoder is not the only option for UWB systems used for cortical interfaces. We selected the Viterbi decoder as the option because it requires least area while still operating above 100 Mbps. In general, higher speed can be achieved through parallel processing, either by duplicating the Viterbi decoder or by using a more complex option like the Gallager decoder. Recently, error correction circuits with sub-threshold operation have been reported in [127]. With this technology, for example, a Gallager-A (GA) decoder which has a highly parallel architecture consumes only $0.66\, mW$ when operating at 200 Mbps in a 65-nm low power high threshold (LP-HVT) CMOS process [127]. Furthermore, analog

![Block diagram of the Viterbi decoder.](image-url)
decoding circuits [128] may be another alternative solution since the power-consuming ADC will be removed in this situation.
CHAPTER 5
CONCLUSIONS AND FUTURE WORK

This work proposed a novel high-speed ultra-wideband wireless communication system for the cortical interface application. Compared to the traditional narrowband techniques, the wideband solution can offer order(s) of transmission speed higher than its existing narrowband counterpart. An IR-UWB receiver system with pulse-injection-locking method is deployed. The injection-locking technique eliminates the clock data recovery circuitry, leading to significant power reduction.

However, the injection-locked IR-UWB system also introduces many difficulties to the communication between the devices in and out of human tissue. The first challenge is to design and fabricate a pair of low-Q inductive coils with high self-resonance frequency in the order of GHz. Secondly, the low-Q coils and the tissue absorption in high-frequency band result in a significant signal attenuation, i.e., a much lower SNR is introduced. Thirdly, the injection-locking technique brings a higher clock jitter to the system compared to other PLL-based designs. All these considerations lead to many uncertainties for the system performance. To address these difficulties, a pair of low-Q coils is fabricated by using a high-frequency material RO4003C, and the physical communication channel through both the air and biological tissue is measured. Moreover, a rational function-based model that can well characterize the coils’ port behavior is built in Matlab for the system reliability evaluation of which the BER is evaluated as the primary measure. Critical results show that:

1. The sampling clock jitter creates an error floor that is independent of the SNR of the received signals. With the use of ECC, the error floor is eliminated;

2. The level of the error floor is determined by the sampling jitter. An LNA with a better noise figure will not contribute to improving the system performance without
the use of ECC;

3. ECC can improve the system performance more than four orders.

With the ECC technique, the burden of designing power-consuming cells, like the LNA, VCO, and ADC, is greatly released thus more power can be saved.

Another contribution of this work is the exploration of $g_m/I_D$ design methodology for circuit optimization in all-inversion regions. The design methodology easily presents the devices’ sizing of the circuit. Therefore, the design period is significantly reduced since only a little adjustments are needed after the Matlab calculation. The works for this part includes:

1. A database is created including the MOSFET semi-empirical models based on the AC characteristics of the device: $g_m/I_D$ versus the normalized current $i$, the drain-to-source conductance over $I_D$, the bulk-to-source conductance over $I_D$, five capacitance of the quasistatic model. The simplified models for passive inductors are also obtained by using the AC simulation at the working frequency;

2. Specific design methodologies are derived for the CG-LNA, the IL-VCO, and the frequency divider circuit block. Its efficiency has been proved since the devices modeling data is collected once and can be reutilized for other circuits designs.

3. A summary of the average power for each circuit block of the receiver is presented in Table 5.1.

<table>
<thead>
<tr>
<th>Circuit Block</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNA</td>
<td>3.37</td>
</tr>
<tr>
<td>IL-VCO</td>
<td>1.61</td>
</tr>
<tr>
<td>Phase Shifter</td>
<td>1.52</td>
</tr>
<tr>
<td>Frequency Divider</td>
<td>0.67</td>
</tr>
<tr>
<td>ECC</td>
<td>0.05</td>
</tr>
<tr>
<td>Total</td>
<td>7.22</td>
</tr>
</tbody>
</table>

Future works could be applied in:

1. The carrier frequency is determined by the characteristics of the coils. Higher carrier frequency can not only increase the data throughput but also can decrease the power
of the LNA and the VCO. Therefore, further coil optimization is worth taken in the future;

2. For the ECC, higher speed can be achieved through parallel processing, either by duplicating the Viterbi decoder or by using a more complex option like the Gallager decoder;

3. Since the system can tolerate RF blocks with the higher noise level, the LC-tank VCO can be replaced with an injection-locked ring oscillator which has less power than its LC-based counterpart at the cost of a higher noise level. In this case, the expensive on-chip inductor can be saved for both the transmitter and the receiver.
Bibliography


APPENDIX
Impulse Response Calculation for the Coils

In order to obtain the impulse response equation, re-write and re-arrange the function (2.2) as:

\[ H_{12}(s) = \frac{sM_{12}}{s^2 + (R_{s1}R_{1}C_1 + L_1)s + (R_1 + R_{s1})} \times \frac{1}{L_2C_2s^2 + R_2C_2s + 1} \]

\[ = \frac{R_{s1}L_1C_1L_2C_2}{s^2 + \frac{R_{s1}R_1C_1 + L_1}{R_{s1}L_1C_1}s + \frac{R_1 + R_{s1}}{R_{s1}L_1C_1}}(s^2 + \frac{R_2}{L_2}s + \frac{1}{L_2C_2}) \]

\[ = \frac{s\Theta}{(s^2 + 2\zeta_1\omega_1s + \omega_1^2)(s^2 + 2\zeta_2\omega_2s + \omega_2^2)} \quad (A.1) \]

where

\[ \Theta = \frac{M_{12}}{R_{s1}L_1C_1L_2C_2} \quad (A.2) \]

\[ \omega_1^2 = \frac{R_1 + R_{s1}}{R_{s1}L_1C_1} \quad (A.3) \]

\[ \omega_2^2 = \frac{1}{L_2C_2} \quad (A.4) \]

\[ \zeta_1 = \frac{R_{s1}R_1C_1 + L_1}{2R_{s1}\sqrt{L_1C_1}} \sqrt{\frac{R_{s1}}{R_1 + R_{s1}}} \quad (A.5) \]

\[ \zeta_2 = \frac{R_2}{2} \sqrt{\frac{C_2}{L_2}} \quad (A.6) \]

Notice that the denominator of (A.1) is combined with two two-order systems. Assume the roots of them are \((r_{11}, r_{12})\) and \((r_{21}, r_{22})\), function (A.1) can be further written as:
\[ H_{12}(s) = \frac{s\Theta}{(s^2 + 2\zeta_1\omega_1s + \omega_1^2)(s^2 + 2\zeta_2\omega_2s + \omega_2^2)} = \frac{s\Theta}{(s - r_{11})(s - r_{12})(s - r_{21})(s - r_{22})} = \frac{A_1}{s - r_{11}} + \frac{A_2}{s - r_{12}} + \frac{B_1}{s - r_{21}} + \frac{B_2}{s - r_{22}} \quad (A.7) \]

where

\[ A_1 = (s - r_{11})H_{12}(s) \big|_{s=r_{11}} = \frac{\Theta r_{11}}{(r_{11} - r_{12})(r_{11} - r_{21})(r_{11} - r_{22})} \quad (A.8) \]

\[ A_2 = (s - r_{12})H_{12}(s) \big|_{s=r_{12}} = \frac{\Theta r_{12}}{(r_{12} - r_{11})(r_{12} - r_{21})(r_{12} - r_{22})} \quad (A.9) \]

\[ B_1 = (s - r_{21})H_{12}(s) \big|_{s=r_{21}} = \frac{\Theta r_{21}}{(r_{21} - r_{11})(r_{21} - r_{12})(r_{21} - r_{22})} \quad (A.10) \]

\[ B_2 = (s - r_{22})H_{12}(s) \big|_{s=r_{22}} = \frac{\Theta r_{22}}{(r_{22} - r_{11})(r_{22} - r_{12})(r_{22} - r_{21})} \quad (A.11) \]

\[ r_{11,12} = \begin{cases} 
-\zeta_1\omega_1 \pm \omega_1\sqrt{\zeta_1^2 - 1}, \zeta_1 \geq 1 \\
-\zeta_1\omega_1 \pm j\omega_1\sqrt{1 - \zeta_1^2}, \zeta_1 < 1 
\end{cases} \quad (A.12) \]

\[ r_{21,22} = \begin{cases} 
-\zeta_2\omega_2 \pm \omega_2\sqrt{\zeta_2^2 - 1}, \zeta_2 \geq 1 \\
-\zeta_2\omega_2 \pm j\omega_2\sqrt{1 - \zeta_2^2}, \zeta_2 < 1 
\end{cases} \quad (A.13) \]

I If \((\zeta_1 \geq 1, \zeta_2 \geq 1)\) In this case, all roots as well as \(A_1, A_2, B_1, B_2\) are real. The impulse response is the reverse Laplace Transform of \((A.7)\):

\[ h(t) = A_1e^{r_{11}t} + A_2e^{r_{12}t} + B_1e^{r_{21}t} + B_2e^{r_{22}t} \quad (A.14) \]
II If \( \zeta_1 < 1, \zeta_2 < 1 \) In this case, all the roots as well as \( A_1, A_2, B_1, B_2 \) are complex. Notice that \((A_1, A_2)\) and \((B_1, B_2)\) are conjugate:

\[ A_1 = A_2^* \quad (A.15) \]

\[ B_1 = B_2^* \quad (A.16) \]

Assume that: \( A_1 = a_1 + jb_1, A_2 = a_1 - jb_1, B_1 = a_2 + jb_2, B_2 = a_2 - jb_2 \). Re-write the complex roots as:

\[ r_{11,12} = -\zeta_1 \omega_1 \pm j\omega_1 \sqrt{1 - \zeta_1^2} = -\tau_1 \pm j\omega_d \quad (A.17) \]

\[ r_{21,22} = -\zeta_2 \omega_2 \pm j\omega_2 \sqrt{1 - \zeta_2^2} = -\tau_2 \pm j\omega_d \quad (A.18) \]

The reverse Laplace Transform of the first two-order system is:

\[
\mathcal{L}_1^{-1} = \mathcal{L}^{-1} \left[ \frac{A_1}{s - r_{11}} + \frac{A_2}{s - r_{12}} \right]
\]

\[
= \mathcal{L}^{-1} \left[ \frac{A_1}{s + \tau_1 - j\omega_d} + \frac{A_1^*}{s + \tau_1 + j\omega_d} \right]
\]

\[
= A_1 e^{(-\tau_1 + j\omega_d) t} + A_1^* e^{(-\tau_1 - j\omega_d) t}
\]

\[
= e^{-\tau_1 t} \left[ A_1 e^{j\omega_d t} + A_1^* e^{-j\omega_d t} \right] \quad (A.19)
\]

The expansion of \((A.19)\) is given by:

\[
\mathcal{L}_1^{-1} = e^{-\tau_1 t} \left\{ (a_1 + jb_1) [\cos(\omega_d t) + j\sin(\omega_d t)] + (a_1 - jb_1) [\cos(\omega_d t) - j\sin(\omega_d t)] \right\}
\]

\[
= 2e^{-\tau_1 t} [a_1\cos(\omega_d t) - b_1\sin(\omega_d t)] \quad (A.20)
\]
Similarly, the reverse Laplace Transform of the second two-order system is:

\[
\mathcal{L}_2^{-1} = 2e^{-\tau_2 t} [a_2 \cos(\omega d_2 t) - b_2 \sin(\omega d_2 t)]
\] (A.21)

So we can get the whole impulse response:

\[
h(t) = \mathcal{L}_1^{-1} + \mathcal{L}_2^{-1}
\]

\[
= 2e^{-\tau_1 t} [a_1 \cos(\omega d_1 t) - b_1 \sin(\omega d_1 t)] + 2e^{-\tau_2 t} [a_2 \cos(\omega d_2 t) - b_2 \sin(\omega d_2 t)]
\] (A.22)

III If \((\zeta_1 < 1, \zeta_2 > 1)\) In this case, \((r_{11}, r_{12}, A_1, A_2)\) are complex, \((r_{21}, r_{22}, B_1, B_2)\) are real. Similarly, the impulse response could be obtained by:

\[
h(t) = 2e^{-\tau_1 t} [a_1 \cos(\omega d_1 t) - b_1 \sin(\omega d_1 t)] + B_1 e^{r_{21} t} + B_2 e^{r_{22} t}
\] (A.23)

IV If \((\zeta_1 > 1, \zeta_2 < 1)\) In this case, \((r_{11}, r_{12}, A_1, A_2)\) are real, \((r_{21}, r_{22}, B_1, B_2)\) are complex. The impulse response is given by:

\[
h(t) = A_1 e^{r_{11} t} + A_2 e^{r_{12} t} + 2e^{-\tau_2 t} [a_2 \cos(\omega d_2 t) - b_2 \sin(\omega d_2 t)]
\] (A.24)
CURRICULUM VITAE

Yi Luo

Published Journal Articles


Published Conference Papers

