LINEARITY AND INTERFERENCE ROBUSTNESS IMPROVEMENT METHODS FOR ULTRA-WIDEBAND CMOS RF FRONT-END CIRCUITS

DISSEPTION

Presented in Partial Fulfillment of the Requirement for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Long Bu, M.S.

* * * * *

The Ohio State University

2008

Dissertation Committee:

Professor Joanne E. DeGroat, Advisor

Professor Steven B. Biblyk

Professor Patrick Roblin

Approved by

Adviser

Graduate Program in Electrical & Computer Engineering
ABSTRACT

As an emerging technology that finds a variety of applications from asset tracking, medical system to high speed data transfer, UWB technology draws tremendous amount of attention from academic, industry and government research groups. Differing substantially from conventional narrowband wireless communication systems, the UWB transceiver needs to process signal at very low power level over an extremely wide frequency range (multi-GHz) in which many other wireless transmissions with much higher power reside. The UWB system are supposed to have minimal interference on other communication while be able to coexists with whatever interferers in the air. This unique aspect of UWB communication makes the RF front-end susceptible to strong interferers which can potentially distort or block the desired signal. Thus in addition to low noise figure which is the major figure of merit of LNA for conventional transceivers, high linearity becomes a highly desirable feature for UWB front-end circuits.

Several LNA linearization theories and techniques for narrowband application have been reported and proved effective. However, no wideband LNA linearization techniques have been proposed to date. In this work, previous works on LNA linearity improvement are
reviewed. Theory and techniques are developed to improve the linearity of ultra-wideband LNA in parallel with the optimization of input matching, noise figure and gain. A novel wideband linearity improvement technique, Extended Effective Range Derivative Superposition Method (EERDS Method), has been developed in this work. A 3-10 GHz high linearity UWB LNA with active balun is designed and fabricated in IBM 0.13 μm CMOS process using the proposed EERDS method. This method can also be applied to other ultra-wideband RF front-end circuit such as mixer, RF VGA and PA.

Other than improving the linearity of the LNA, programmable band-select RF front-end circuit is another approach to enhance the robustness of the UWB transceiver to strong narrowband interferers that is proposed in this work. The idea is to grant the LNA the ability to switch pass band, such that the transceiver can detect and avoid the strong narrowband interferers, which greatly improve the ability of UWB RF front-end to survive strong narrowband interferers. Two 3-10 GHz band-select UWB LNAs with active balun, one configurable between narrowband (3-5 GHz) and wideband (3-10 GHz) mode and the other configurable between low band (3-5 GHz) and high band (6-10 GHz) mode, are designed and fabricated in IBM 0.13 μm CMOS process to demonstrate the proposed concept.
To my wife Yan
ACKNOWLEDGMENTS

First of all, I would like to express my greatest appreciation to my academic advisor Professor Joanne DeGroat. She has trusted and supported me throughout my endeavor in my PhD study and research work. Her invaluable mentorship has benefited me not only in my academic efforts, but also in work ethic and many other important things in life. I would not have been able to finish this work without her unreserved support.

I am deeply grateful to Professor Steve Bibyk for his help throughout the project. Since we collaborated with his group on the UWB Receiver project, I have had many opportunities to ask him questions and discuss with him on certain topics. His helpful advices and insightful comments have broadened my view and benefited me a lot in my research.

I am also sincerely thankful to Professor Patrick Roblin for being the committee member of my qualifier, candidacy exams and my final PhD defense. In addition, I am grateful for all the excellent courses and labs from him in RF circuits and systems. I have learned from him knowledge and techniques that have benefited me in my research work and will benefit me in my future career.
I am very thankful to Infoscitex Corp. (Systran Corp. back then) and the SBIR program of the U.S. Small Business Administration (SBA) Office for their support in this work. Particularly, I would like to thank Mr. Kien of Infoscitex Corp. for his consistent support on my work. We had many technical discussions on different topics especially on the package issues and board level design.

I am very thankful to Professor Valco for his help when I worked for him on my first Teaching Assistantship. He spent a lot of time helping me getting familiar with different equipment and procedures in the Microelectronics Cleanroom Laboratory. I am deeply grateful to him for all his help and the invaluable working ethic I learned from him.

I am very grateful to Bo Liang of Alereon Inc. for his help and many enlightening discussions. I also want to thank Brain Dupaix for his help in setting up and maintaining the IBM design kit. I would also like to express my appreciation here to Aaron Aufderheide and Edwin Lim for their IT support. Especially during the time that I conducted my research work remotely from Austin, TX, they helped me a lot in setting up the remote connection and maintaining the CAD tools.

Finally, I am deeply grateful to my beloved wife Yan for her encourage, patience and continuous support throughout this work. I wish her the best in the final stage of her PhD research work.
VITA

May 10, 1976.................................................. Born – Shenyang, China

1994-1998 ................................................. B.S. Materials Science
Shanghai Jiao tong University
Shanghai, China

2002-2004 ............................................... M.S. Electrical Engineering
University of Connecticut
Storrs, Connecticut

2004-2007 ............................................... Graduate Student,
Electrical Engineering
The Ohio State University
Columbus, Ohio

2007-present ......................................... RFIC Designer
Aleron Inc.
Austin, Texas

PUBLICATIONS


FIELD OF STUDY

Major Field: Electrical Engineering

Study in:

RF integrated circuits, Wireless Application, CMOS Front-end Circuits

viii
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT ........................................................................................................................ ii</td>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS ................................................................. v ..........................</td>
<td></td>
</tr>
<tr>
<td>VITA ............................................................................................................................ vii</td>
<td></td>
</tr>
<tr>
<td>TABLE OF CONTENTS ........................................................... ix ..........................</td>
<td></td>
</tr>
<tr>
<td>LIST OF TABLES ....................................................................................... xiv ........</td>
<td></td>
</tr>
<tr>
<td>LIST OF FIGURES .................................................................................. iii ..........</td>
<td></td>
</tr>
<tr>
<td>CHAPTERS .................................................................................................................</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION ......................................................................................... 1</td>
<td></td>
</tr>
<tr>
<td>1.1 Purpose of This Research ................................................................. 1</td>
<td></td>
</tr>
<tr>
<td>1.2 Dissertation Organization ............................................................... 3</td>
<td></td>
</tr>
<tr>
<td>2. BACKGROUND ......................................................................................... 6</td>
<td></td>
</tr>
<tr>
<td>2.1 A Brief History of UWB ................................................................. 6</td>
<td></td>
</tr>
<tr>
<td>2.2 The Standards War ............................................................................. 8</td>
<td></td>
</tr>
<tr>
<td>2.3 UWB, an Emerging Wireless Technology ....................................... 10</td>
<td></td>
</tr>
<tr>
<td>3. ULTRA-WIDEBAND FRONT-END CIRCUIT DESIGN TECHNIQUES AND CONSIDERATIONS ...................................................... 15</td>
<td></td>
</tr>
</tbody>
</table>
6.1 Introduction ..................................................................................................... 102
6.2 Circuit Design and Implementation ................................................................. 103
  6.2.1 Constant g_m Bias Circuit ......................................................................... 103
  6.2.2 High Linearity 3-10 GHz CMOS UWB LNA Core Circuit ....................... 104
  6.2.3 Active Balun Circuit ............................................................................... 105
  6.2.4 Output Buffer Circuit .............................................................................. 107
6.3 Results and Discussion ................................................................................... 109
  6.3.1 Impedance Matching ............................................................................... 110
  6.3.2 Gain Response ........................................................................................ 111
  6.3.3 Noise Figure ............................................................................................ 114
  6.3.4 Linearity .................................................................................................. 115
  6.3.5 Reverse Isolation and Stability ............................................................... 123
  6.3.6 Conclusions ............................................................................................. 125

7. WIDEBAND / NARROWBAND DUEL MODE CMOS UWB LNA WITH ACTIVE
   BALUN ........................................................................................................................... 127
  7.1 Introduction ..................................................................................................... 127
  7.2 Circuit Design and Implementation ................................................................. 128
    7.2.1 Wideband / Narrowband Duel Mode CMOS UWB LNA Core Circuit . 128
    7.2.2 Active Balun Circuit ............................................................................... 129
    7.2.3 Output Buffer Circuit .............................................................................. 130
  7.3 Results and Discussion ................................................................................... 131
    7.3.1 Impedance Matching ............................................................................... 132
LIST OF TABLES

Table                                                                 Page

Table 2.1: WiMedia Alliance OFDM physical band allocation .................. 12

Table 9.1: Summary of band-select LNA with active balun performance .......... 166

Table 9.2: Summary of LNA performance and comparison with other works ........ 166
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1: The first UWB receiver devised by G.F. Ross</td>
<td>7</td>
</tr>
<tr>
<td>Figure 2.2: Comparison of range and data rate of different wireless technologies</td>
<td>11</td>
</tr>
<tr>
<td>Figure 3.1: Noise model of two port network</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3.2: Typical amplifier topology</td>
<td>23</td>
</tr>
<tr>
<td>Figure 3.3: Common source with inductive source degeneration</td>
<td>26</td>
</tr>
<tr>
<td>Figure 3.4: LNA with 3rd order Chebychev band pass filter</td>
<td>27</td>
</tr>
<tr>
<td>Figure 3.5: Schematic of common source differential balun</td>
<td>32</td>
</tr>
<tr>
<td>Figure 3.6: Schematic of common source differential balun with RLC feedback</td>
<td>34</td>
</tr>
<tr>
<td>Figure 3.7: Common source CMOS balun with current reuse</td>
<td>35</td>
</tr>
<tr>
<td>Figure 4.1: Typical direct conversion receiver architecture</td>
<td>37</td>
</tr>
<tr>
<td>Figure 4.2: UWB and potential interferers transmit power</td>
<td>38</td>
</tr>
<tr>
<td>Figure 4.3: UWB receiver with ADC in frequency domain</td>
<td>39</td>
</tr>
<tr>
<td>Figure 4.4: Fundamentals and the 3rd order IM products</td>
<td>42</td>
</tr>
<tr>
<td>Figure 4.5: Effect of 3rd IM on UWB signal</td>
<td>43</td>
</tr>
<tr>
<td>Figure 5.1 1st, 2nd and 3rd order power series coefficients vs. Vgs</td>
<td>47</td>
</tr>
<tr>
<td>Figure 5.2 DC optimum biasing point</td>
<td>47</td>
</tr>
<tr>
<td>Figure 5.3 Composite CMOS for 3rd order distortion cancellation</td>
<td>48</td>
</tr>
</tbody>
</table>
Figure 5.4 3rd order coefficients of individual MOSFET and composition MOSFET ..... 48
Figure 5.5 IIP3 of LNA with different frequency and source degeneration inductance ... 50
Figure 5.6 Modified DS method circuit ................................................................. 50
Figure 5.7 Feed forward linearization topology ...................................................... 51
Figure 5.8 APD high linearity LNA ................................................................. 53
Figure 5.9 MOSFET DC test circuit for linearity analysis ........................................ 64
Figure 5.10: Normalized drain current of RF NFET vs. gate bias voltage (W=2.5 µm L=120nm) ................................................................. 65
Figure 5.11: Normalized 1st order power series coefficient of RF NFET drain current (W=2.5 µm L=120nm) ................................................................. 66
Figure 5.12: Normalized 2nd order power series coefficient of RF NFET drain current (W=2.5 µm L=120nm) ................................................................. 67
Figure 5.13: Normalized 3rd order power series coefficient of RF NFET drain current (W=2.5 µm L=120nm) ................................................................. 68
Figure 5.14: Low frequency A_{IIP3} derived from device DC characteristics (W=2.5 µm L=120nm) ................................................................. 69
Figure 5.15: Process variation of g3 and derived A_{IIP3} (W=2.5 µm L=120nm) .......... 71
Figure 5.16: Histogram of optimum V_{gs} over process variation (W=2.5 µm L=120nm) 71
Figure 5.17: Temperature variation of g3 and derived A_{IIP3} (W=2.5 µm L=120nm) ...... 72
Figure 5.18: Concept schematic of derivative superposition ........................................ 73
Figure 5.19: Superposition of g3 curves (W=2.5 µm L=120nm) .................................... 74
Figure 5.20: Derivative superposition using multiple auxiliary MOSFETs .................. 76
Figure 5.21: Normalized g3 curves of MOSFETs with different V_{gs} offset (W=2.5 µm) 77
Figure 5.22: The effect of channel length on threshold voltage ..................................... 79
Figure 5.23: Normalized g3 curves of MOSFETs with different length (W=2.5 µm) ....... 80
Figure 5.24: Proposed Extended Effective Range Derivative Superposition method ..... 81
Figure 5.25: Normalized $g_3$ curves of MOSFETs with EERDS Method (W=2.5 µm) ... 82
Figure 5.26: Calculated $A_{IP3}$ curves of MOSFETs with EERDS Method (W=2.5 µm). 82
Figure 5.27: Small signal model of nonlinear amplifier with source inductor .............. 84
Figure 5.28: 3rd order harmonic current cancellation test (no source inductor) .......... 88
Figure 5.29: 3rd order intermodulation current cancellation using EERDS method - magnitude (no source inductor) ......................................................................................................................... 90
Figure 5.30: 3rd order intermodulation current cancellation using EERDS method - phase (no source inductor) ................................................................................................................................. 91
Figure 5.31: Simulated IP3 at high frequency with EERDS method (no source inductor) ........................................................................................................................................... 92
Figure 5.32: 3rd order harmonic current cancellation test (with source inductor) ....... 94
Figure 5.33: 3rd order intermodulation current cancellation using EERDS method – magnitude (with 3nH source inductor) ............................................................................................................. 95
Figure 5.34: 3rd order intermodulation current cancellation using EERDS method - phase (with 3nH source inductor) .................................................................................................................. 96
Figure 5.35: Simulated IP3 at high frequency with EERDS method (with 3nH source inductor) ............................................................................................................................................... 96
Figure 5.36: 3rd order harmonic current cancellation using modified EERDS method (with source inductor) ........................................................................................................................................... 98
Figure 5.37: 3rd order intermodulation current cancellation using modified EERDS method –magnitude (with 3nH source inductor) ................................................................................................................................. 99
Figure 5.38: 3rd order intermodulation current cancellation using modified EERDS method -phase (with 3nH source inductor) .................................................................................................................................. 100
Figure 5.39: Simulated IP3 at high frequency with modified EERDS method (with 3nH source inductor) ........................................................................................................................................... 100
Figure 6.1: Constant $g_m$ bias circuit ......................................................................... 103
Figure 6.2: High linearity 3-10 GHz CMOS UWB LNA Core Circuit ......................... 104
Figure 6.3: High linearity 3-10 GHz CMOS active balun circuit .................................. 106
Figure 6.4: High linearity 3-10 GHz CMOS output buffer circuit ............................... 107
Figure 6.5: Layout of High linearity 3-10 GHz CMOS LNA with active balun

Figure 6.6: High linearity 3-10 GHz CMOS LNA input matching

Figure 6.7: High linearity 3-10 GHz CMOS LNA output matching

Figure 6.8: High linearity 3-10 GHz CMOS LNA gain

Figure 6.9: High linearity 3-10 GHz CMOS LNA gain imbalance

Figure 6.10: High linearity 3-10 GHz CMOS LNA phase imbalance

Figure 6.11: High linearity 3-10 GHz CMOS LNA noise figure

Figure 6.12: IIP3 of high linearity 3-10 GHz CMOS LNA

Figure 6.13: OIP3 of high linearity 3-10 GHz CMOS LNA

Figure 6.14: High linearity 3-10 GHz CMOS LNA IP3 temperature variation

Figure 6.15: High linearity 3-10 GHz CMOS LNA IP3 supply voltage variation

Figure 6.16: High linearity 3-10 GHz CMOS LNA IP3 process variation

Figure 6.17: The effect of tone spacing of two tone test on linearization

Figure 6.18: P1dB of high linearity 3-10 GHz CMOS LNA with EERDS method

Figure 6.19: P1dB of high linearity 3-10 GHz CMOS LNA without EERDS method

Figure 6.20: S12 of high linearity 3-10 GHz CMOS LNA

Figure 6.21: Simulated Rollet’s stability factor of high linearity 3-10 GHz CMOS LNA

Figure 7.1: 3-10 GHz CMOS Band-select (wideband/narrowband) UWB LNA circuit

Figure 7.2: 3-10 GHz CMOS active balun circuit

Figure 7.3: 3-10 GHz CMOS output buffer circuit

Figure 7.4: Layout of 3-10 GHz band-select (narrowband / wideband) CMOS LNA with active balun

Figure 7.5: Input impedance matching of 3-10 GHz band-select (narrowband / wideband) CMOS LNA
Figure 7.6: Output impedance matching of 3-10 GHz band-select (narrowband / wideband) CMOS LNA ................................................................. 134

Figure 7.7: Gain response of 3-10 GHz band-select (narrowband / wideband) CMOS LNA ......................................................................................... 135

Figure 7.8: 3-10 GHz band-select (narrowband / wideband) CMOS LNA gain imbalance .......................................................................................... 136

Figure 7.9: 3-10 GHz band-select (narrowband / wideband) CMOS LNA phase imbalance ...................................................................................... 137

Figure 7.10: 3-10 GHz band-select (narrowband / wideband) CMOS LNA noise figure .......................................................................................... 138

Figure 7.11: IIP3 of 3-10 GHz band-select (narrowband / wideband) CMOS LNA ..... 139

Figure 7.12: OIP3 of 3-10 GHz Band-select (narrowband / wideband) CMOS LNA ... 139

Figure 7.13: P1dB of 3-10 GHz Band-select (Narrowband mode) CMOS LNA .......... 140

Figure 7.14: P1dB of 3-10 GHz Band-select (Wideband mode) CMOS LNA............. 141

Figure 7.15: S12 of 3-10 GHz Band-select (narrowband / wideband) CMOS LNA...... 142

Figure 7.16: Rollet’s stability factor kf of 3-10 GHz Band-select (narrowband / wideband) CMOS LNA\ ................................................................. 142

Figure 8.1: 3-10 GHz CMOS Band-select (high band/low band) UWB LNA circuit... 146

Figure 8.2: Variable gain 3-10 GHz CMOS active balun circuit ............................... 147

Figure 8.3: Layout of 3-10 GHz band-select (low band / high band) CMOS LNA with active balun................................................................. 149

Figure 8.4: Input impedance matching of 3-10 GHz Band-select (low band / high band) CMOS LNA................................................................. 150

Figure 8.5: Output impedance matching of 3-10 GHz Band-select (low band / high band) CMOS LNA................................................................. 151

Figure 8.6: Gain response of 3-10 GHz Band-select (low band / high band) CMOS LNA ......................................................................................... 152

Figure 8.7: Gain response of 3-10 GHz Band-select (low band mode) CMOS LNA at different gain settings .............................................................. 153
Figure 8.8: Gain response of 3-10 GHz Band-select (high band mode) CMOS LNA at
different gain settings...................................................................................................... 153

Figure 8.9: 3-10 GHz Band-select (low band / high band) CMOS LNA gain imbalance
......................................................................................................................................... 154

Figure 8.10: 3-10 GHz Band-select (low band / high band) CMOS LNA phase imbalance
......................................................................................................................................... 155

Figure 8.11: 3-10 GHz Band-select (low band / high band) CMOS LNA noise figure 156

Figure 8.12: IIP3 of 3-10 GHz Band-select (low band / high band) CMOS LNA........ 157

Figure 8.13: OIP3 of 3-10 GHz Band-select (low band / high band) CMOS LNA....... 157

Figure 8.14: P1dB of 3-10 GHz Band-select (low band / high band) CMOS LNA ....... 158

Figure 8.15: Effect of output buffer on P1dB (high band mode)................................. 159

Figure 8.16: S12 of 3-10 GHz Band-select (low band / high band) CMOS LNA........ 160

Figure 8.17: Rollet’s stability factor kf of 3-10 GHz Band-select (low band / high band)
CMOS LNA.................................................................................................................... 160
CHAPTER 1

INTRODUCTION

1.1 Purpose of This Research

The unconventional way of transmitting and receiving RF signals in UWB system presents many design challenges to the front-end RF circuit design, such as wide bandwidth, low noise figure and high linearity. In this work, these design challenges are evaluated and discussed from a system perspective and their significances and tradeoffs are discussed. The theories and techniques are investigated in an effort to improve the overall performance of the UWB RF front-end circuits.

Among these tasks, the main objective of this research is to develop theory and techniques of Ultra-wideband linearization. Several LNA linearization methods have been proved to be effective in some works reported recently [1-3]. However, these linearization techniques typically have two major limitations: small effective $V_{gs}$ range
and narrow bandwidth. Small effective $V_{gs}$ range not only limits the large signal linearity of the amplifier, but also make the design sensitive to process and temperature variations, which limits the effectiveness in practical applications. The narrowband nature, on the other hand, prevents these linearization techniques from being directly applied in UWB applications. In this work, research efforts are focused on the improvement on these two aspects. Starting from the device model of the typical short channel MOSFET, an analytical model is developed to describe the linearity of the MOSFET in both the weak inversion and the strong inversion region. Based on this model, an Extended Effective Range Derivative Superposition (EERDS) method is developed and its effectiveness at both DC and high frequency is examined. At high frequency, the distortion behavior and the linearity degradation mechanism are studied using 3rd order intermodulation current cancellation technique. With necessary modification for RF application, the EERDS method is proved to be effective over a wide frequency range.

As an integrate part of this work, a 3-10 GHz high linearity UWB LNA with active balun was designed and fabricated in IBM 0.13 μm CMOS process using the proposed EERDS method. The simulation data indicate that the EERDS method effectively improves the linearity of the LNA over 7 GHz bandwidth.

In addition to linearization, another path taken in this research to improve the robustness of the front-end circuit is to add flexibility in frequency domain. All the UWB LNAs reported to date have fixed bandwidth. Thus if a strong narrowband interferer falls in
band and severely distort the front-end circuits, which is very likely to happen due to the wide bandwidth of UWB receiver, the receiver can suffer degraded performance or even get blocked. However, if the bandwidth of the front-end circuit can be adjusted, the receiver can switch to other band if a strong narrowband interferer is detected. A detect and avoid scheme like this makes the receiver to be smarter and more adaptable to the interferers. Further, since different regions allocate different frequency band for UWB operation, it would be desirable to have the capability to disable the frequency bands that are not being used in order to avoid unnecessary interference. Towards this end, two 3-10 GHz band-select UWB LNAs with active balun, one configurable between narrowband (3-5 GHz) and wideband (3-10 GHz) mode and the other configurable between low band (3-5 GHz) and high band (6-10 GHz) mode, are designed and fabricated in IBM 0.13 μm CMOS process to demonstrate this concept.

1.2 Dissertation Organization

CHAPTER 2 gives a brief history of the UWB and an introduction of the current status of the UWB technology.

CHAPTER 3 discusses the major design considerations and tradeoffs for UWB LNA. Several active balun topologies and their application in UWB receiver are also discussed.
In CHAPTER 4, the effect and significance of the linearity of the front-end circuit of the UWB receiver is investigated from a system perspective.

In CHAPTER 5, LNA linearization theory and techniques are review and their merits and limitations are discussed. An analytical model is developed to describe the linearity of the MOSFET in both the weak inversion and the strong inversion region. Based on this model, an Extended Effective Range Derivative Superposition (EERDS) method is developed to improve the effective gate bias range and the bandwidth of the linearization. The effectiveness of the EERDS method is examined at both DC and high frequency.

In, CHAPTER 6, a 3-10 GHz high linearity UWB LNA with active balun designed using the proposed EERDS method is presented. The simulation data indicates the EERDS method effectively improves the linearity of the LNA over 7 GHz frequency range.

In CHAPTER 7, a different way of improving the robustness of the front-end circuit is discussed. The band-select capability of the front-end circuit allows the receiver to detect and avoid strong narrowband interferers. A 3-10 GHz band-select UWB LNAs with active balun configurable between narrowband (3-5 GHz) and wideband (3-10 GHz) mode is presented and the corresponding simulation results are discussed.

For better interferer suppression, a band-select LNA with non-overlapping passband is highly desirable. In CHAPTER 8, 3-10 GHz band-select UWB LNAs with active balun
configurable between low band (3-5 GHz) and high band (6-10 GHz) mode is presented and the corresponding simulations result are discussed.

CHAPTER 9 summarizes this research work and draws conclusions.
CHAPTER 2

BACKGROUND

2.1 A Brief History of UWB

Ultra-wideband (UWB) technology, which appears to be a recent technology breakthrough to many people, actually has a history as long as the wireless communication. In 1887, Hertz conducted his famous experiment which proved the existence of the electromagnetic wave. Using a couple of spark gaps, coils and a primitive dipole antenna, he built a transmitter which can radiate electromagnetic wave and a receiver which can sense the wave and acknowledge the reception with a visible spark. Though without any intention of pursuing this end, the signal generated by Hertz’s instrument has a very wide bandwidth and is legitimately a UWB signal.

The spark gap transmitter caught the interests of many people and continued to be studied and developed over the following decades. On December 12, 1901, Marconi
accomplished his historic cross-Atlantic transmission using spark gap technique, the signal of which has, inevitably, very rich spectrum. With the rapid development of wireless communication in that era, people started to realize that such unregulated wideband signal can cause severe interference. Thus the carrier based wireless communication became the mainstream in 1920s and has held the position till today. UWB, on the other hand, remains active in many areas.

The concept of UWB in modern sense was developed in early 1960s by Bennett and Ross in their research work on time domain electromagnetics [4], the study of electromagnetic-wave propagation from a time domain perspective. In 1973, Ross filed the first fundamental patent on UWB communications systems. Figure 2.1 shows the circuit diagram of the UWB receiver in this patent. In 1978, Ross demonstrated the first UWB communication system that worked in free space.

Figure 2.1: The first UWB receiver devised by G.F. Ross [5]
Interestingly, when Ross invented the first UWB communication system, the term “baseband” is used instead of “UWB”. The term “UWB” made its debut in the radar research conducted under the Defense Advanced Research Projects Agency (DARPA) in 1990 [6, 7]. The initial intended usage of this term is to differentiate radar applications utilizing short-pulse waveforms having a large fractional bandwidth (>25%) from their traditional counterparts.

2.2 The Standards War

By late 90s, numerous patents on UWB communication had been filed. However, no legislation had been made by FCC on this topic, which presents a huge uncertainly that cast a cloud on the research and commercialization of the UWB. In 1998, FCC released the Notice of Inquiry [8] as its first public effort to investigate the possibility of permitting the operation of ultra-wideband (UWB) radio systems on an unlicensed basis under Part 15 of its rules. Four years later, a February 14, 2002 Report and Order by the FCC [9] authorizes the unlicensed use of UWB in 3.1–10.6 GHz. The FCC power spectral density emission limit for UWB emitters operating in the UWB band is -41.3 dBm/MHz, which is the same limit that applies to unintentional emitters in the UWB band under the Part 15 rules. In March 2007, FCC released an important waiver on the measurement of the emission power of the UWB transmitter.
The fast development of UWB technology calls for an industrial standard. IEEE 802.15.3a attempted to provide a higher speed UWB PHY enhancement amendment to IEEE 802.15.3 for applications which involve imaging and multimedia. This effort, however, ignited a standard war between two large industrial alliances with fundamentally different ways in implementation of UWB communication. The WiMedia Alliance, which is led by Intel proposed the use of Multiband OFDM (orthogonal frequency division multiplexing) for communication devices. On the other side, the UWB Forum led by Freescale advocates DS (direct sequence) architecture.

Comparing with DS-UWB architecture, the Multiband OFDM architecture has many inherent advantages. It has better range and causes less interference. It also allows the integration of multiple radio front-ends and can support multi-protocol using the same transceiver hardware. Most importantly, the ability to freely choose which bands to be used allows this technology to comply with different regulations all over the world. This ability also enable the “detect and avoid” (a mandatory feature in Japan) of other radio signals in presence in order to minimize potential interference. The penalty that comes with these merits of Multiband OFDM architecture is higher level of complexity, slightly higher power consumption and longer time to market.

After an extended effort, IEEE 802.15.3a task group (TG3a) failed to work out a joint proposal between the two alliances. On January 19, 2006, the task group dismissed itself and decided to leave this choice to the market. In this competition, the USB
Implementors' Group, the European standards maker ECMA and the Bluetooth SIG elected to favor WiMedia over UWB Forum. In 2006, Freescale, the founder and the leader of UWB Forum, decided to leave the group. At that point, the wide acceptance of WiMedia appeared to be inevitable.

2.3 UWB, an Emerging Wireless Technology

Despite of its rather long history, UWB technology is, in many senses, an emerging wireless technology that draws tremendous interests from academic, industry and government research groups. Thanks to the breakthroughs in wireless communication, semiconductor processing and integrated circuit design, UWB technology has gone through fast development and finds applications in many different areas, including short range high speed data transfer, smart home, security system, through wall imaging, medical system, surveillance radar and WPAN (wireless personal network). Among these applications, wireless USB as a short range high speed UWB application is considered the most promising commercial application and draws the most attention. According to the latest WiMedia standard, the wireless USB is supposed to provide very high speed (480 Mbit/s within 3 meter and 110 Mbit/s within 10 meter) wireless link among different devices used in an office or a digital home. Figure 2.2 shows a comparison of different wireless technologies in terms of transmission distance and data rate, which indicates a substantial speed advantage in short range. In addition, UWB is supposed to be more cost
and power efficient per Megabyte of data transfer compared with other wireless technologies. These advantages make UWB the best technology for next generation wireless personal networks (WPAN).

![Figure 2.2: Comparison of range and data rate of different wireless technologies](image)

In WiMedia standard, the 7.5 GHz bandwidth from 3.1 GHz to 10.6 GHz that is allocated for UWB operation is divided in to 5 band groups, with each band group is further divided into 3 bands with 528 MHz bandwidth (except for band group 5, which has 2 bands only) as shown in Table 2.1. The UWB device is operating in one band group at any given time slot and constantly hopping between 3 bands in order to further spread the
power in frequency domain and minimize the interference to other wireless communication.

<table>
<thead>
<tr>
<th>Band Group</th>
<th>Band</th>
<th>Lower Frequency</th>
<th>Center Frequency</th>
<th>Upper Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3168 MHz</td>
<td>3432 MHz</td>
<td>3696 MHz</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3696 MHz</td>
<td>3960 MHz</td>
<td>4224 MHz</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4224 MHz</td>
<td>4488 MHz</td>
<td>4752 MHz</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4752 MHz</td>
<td>5016 MHz</td>
<td>5280 MHz</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5280 MHz</td>
<td>5544 MHz</td>
<td>5808 MHz</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5808 MHz</td>
<td>6072 MHz</td>
<td>6336 MHz</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>6336 MHz</td>
<td>6600 MHz</td>
<td>6864 MHz</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>6864 MHz</td>
<td>7128 MHz</td>
<td>7392 MHz</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>7392 MHz</td>
<td>7656 MHz</td>
<td>7920 MHz</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>7920 MHz</td>
<td>8184 MHz</td>
<td>8448 MHz</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>8448 MHz</td>
<td>8712 MHz</td>
<td>8976 MHz</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>8976 MHz</td>
<td>9240 MHz</td>
<td>9504 MHz</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>9504 MHz</td>
<td>9768 MHz</td>
<td>10032 MHz</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>10032 MHz</td>
<td>10296 MHz</td>
<td>10560 MHz</td>
</tr>
</tbody>
</table>

Table 2.1: WiMedia Alliance OFDM physical band allocation

UWB frequency band allocation is different in different regions. United States have the most relaxed regulation on UWB. All 14 bands are available for UWB transceiver operation. In Europe, band group 3, 4 can be used for UWB with the exception of band 12. Band group 1 can be used for UWB on the condition that DAA (Detect and Avoid) is implemented. In Japan and Korea, band 9-13 can be used for UWB communication. Same as European regulation, band group 1 can be used for UWB only if DAA (Detect and Avoid) is implemented.
The very high data rate achieved by the UWB technology comes from its ultra-wide bandwidth. In traditional narrowband communication system, the effort to improve the data rate has been focused on improving the SNR (signal to noise ratio) of the receiver. According to Shannon’s channel capacity theorem, the maximum achievable bandwidth efficiency in a channel contaminated with additive white Gaussian noise can be written as [10]:

\[ C = B \cdot \log_2 \left( 1 + \frac{S}{N} \right) \]  

(2.1)

Where \( C \) is the channel capacity in b/s and \( B \) is the transmission bandwidth in Hz. Thus the channel capacity increases logarithmically with SNR and saturates when SNR reaches a certain level, beyond which further increase of SNR negligibly improves the channel capacity.

The UWB technology, on the other hand, takes a different route in the effort of improving the data rate. Equation (2.1) also indicates that the channel capacity increases linearly with bandwidth. Thus even with very lower transmission power, which means lower SNR, very high data rate can be achieved by aggressively increasing the bandwidth.

In addition to high throughput and low power, another important merit of Wireless USB that is less often cited is the high device capacity. The WiMedia standard allows 127 devices to connect to a host simultaneously. In addition, because of the short range nature of UWB, the cluster of devices in the next room can reuse the spectrum without causing
any interference. Thus Wireless USB is inherently suitable for environment with a high density of devices. 802.11n, which is considered a promising next generation WLAN technology, would face many challenges when the number of devices within the transmission range increases.
3.1 Input Matching

In a receiver, the antenna is connected to the input port of LNA directly or through a RF filter. In either case, the interface typically has an impedance of 50 $\Omega$. In order to minimize the power loss due to the reflection at this interface, the input matching of the LNA should be optimized.

The $S_{11}$ represents the input reflection coefficients of a two port network. It is defined by the ratio of the reflected wave vector to the input wave vector when the output is matched.

$$S_{11} = \frac{b_1}{a_1} \bigg|_{a_2=0}$$  \hspace{1cm} (3.1)
$a_i = \frac{1}{2\sqrt{Z_0}}(V_1 + Z_0 I_1) \quad b_i = \frac{1}{2\sqrt{Z_0}}(V_1 - Z_0 I_1)$

$S_{11}$ is a vector which can be expressed in complex form or magnitude and angle. In most cases, it is referred as its magnitude in decibel, which can be expressed as:

$$S_{11B} = 20 \log(|S_{11}|)$$

At multi-GHz frequency range, the signal power is scarce and maximum power transfer is highly desirable. For LNA design, in order to maximize the power transfer from the antenna to the LNA, the reflection power at the input of the LNA should be minimized. Thus, a good input matching (small value of $S_{11}$) is highly desirable.

Due to the inductive and capacitive elements in the input matching network of the LNA, the $S_{11}$ is usually a strong function of frequency. An important task in LNA design is to guarantee the $S_{11}$ to be sufficiently small in the designed frequency band. In most cases, the $S_{11}$ value of around -10 dB is sufficient. $S_{11}$ can be expressed as:

$$S_{11} = \frac{Z_{in} - Z_s^*}{Z_{in} + Z_s^*}$$

(3.2)

Where $Z_{in}$ is the input impedance of the LNA and $Z_s$ is the source impedance, which typical has the value of 50 $\Omega$. Thus in order to achieve a small $S_{11}$, $Z_{in}$ has to be conjugate matched to $Z_s$. 

16
3.2 Noise Figure

For given RF receiver, the required noise figure is dictated by equation (3.3):

\[ S_{\text{min}} = -174 + 10 \cdot \log_{10}(B) + NF + SNR_{\text{min}} \]  

(3.3)

Where \( S_{\text{min}} \) is the minimum detectable signal level and \( SNR_{\text{min}} \) is the minimum required SNR to retrieve the data from the received signal. An important task of receiver design is to minimize the sensitivity (\( S_{\text{min}} \)) which leads to higher dynamic range. Towards this end, a very low noise figure is highly desirable.

For a well designed receiver, the noise figure of the LNA typically constitutes the majority part of, and thus determines, that of the receiver. As a result, the optimization of the noise figure is typically the most important goal in LNA design. For UWB LNA, due to the lower achievable gain and ultra wide bandwidth required, the noise optimization is much more challenging than that of conventional narrowband LNA.

3.2.1 Classic 2 Port Noise Model

Classic Two Port Noise Model has long been used to analyze and optimize the noise of the RF amplifier. As described in [11], the amplifier is modeled as noiseless two port network with an input referred voltage noise source in series and an input referred current noise source in shunt. As shown in Figure 3.1.
The noise figure can be simply expressed as the ratio of the overall noise at the input to the noise due to the source.

\[ F = \frac{\bar{i}_{c}^2 + |i_{n} + Y_{s}e_{n}|^2}{i_{s}^2} \]  

(3.4)

Here the noise source \( i_{n} \) and \( e_{n} \) can be correlated. In order to take this effect into account, \( i_{n} \) can be expressed as:

\[ i_{n} = i_{c} + i_{u} \]

Where \( i_{c} \) is completely correlated with \( e_{n} \) while \( i_{u} \) is uncorrelated to \( e_{n} \). The noise parameters are defined as:

\[ R_{n} = \frac{\bar{e}_{n}^2}{4KT\Delta f} \quad G_{u} = \frac{\bar{i}_{u}^2}{4KT\Delta f} \quad G_{s} = \frac{\bar{i}_{s}^2}{4KT\Delta f} \quad Y_{c} = \frac{i_{c}}{e_{n}} \]

It can be proved that

\[ F' = 1 + \frac{G_{u} + Y_{c} + Y_{s}^2 R_{n}}{G_{s}} \]  

(3.5)

Solving for the condition for minimum noise figure, it arrives
\[ \text{Re}\{Y_{\text{opt}}\} = G_{\text{opt}} = \frac{G_u + G_c^2}{R_n} \] (3.6)

\[ \text{Im}\{Y_{\text{opt}}\} = B_{\text{opt}} = -B_c \] (3.7)

When this condition is satisfied, the noise figure reaches its minimum value of:

\[ F_{\text{min}} = 1 + 2R_n \left[ \frac{G_u + G_c^2}{R_n} \right] \] (3.8)

The noise figure of arbitrary input admittance can be expressed as:

\[ F = F_{\text{min}} + \frac{R_n}{G_s} |Y_s - Y_{\text{opt}}|^2 \] (3.9)

Here, noise parameter \( R_n \) indicates the sensitivity of noise figure to the deviation of \( Y_s \) to its optimum value. A large \( R_n \) implies that a small deviation of \( Y_s \) from \( Y_{\text{opt}} \) results in a large noise figure increase, and vice versa.

### 3.2.2 MOSFET Noise Model

There are several different mechanisms that generate noise in a MOSFET [6]. First, the thermal noise in the channel contributes to the drain current noise, which can be expressed as:

\[ \overline{i_{\text{nd}}}^2 = 4KT\gamma g_{\text{d0}}\Delta f \] (3.10)

Where \( \gamma \) is a process dependent constant and \( g_{\text{d0}} \) is the transconductance at zero \( V_{\text{ds}} \).
Second, the drain current also exhibit flicker noise due to the trap and release of the carriers at the interface defects. This noise can be expressed as:

\[
\bar{i}_{\text{nd,flicker}}^2 = \frac{K}{f} g_m^2 \frac{W}{L} C_{\text{ox}}^{-2} \Delta f
\]  

(3.11)

It can be seen that the flicker noise is inversely proportional to the frequency. For high frequency application such as LNA that operates at multi-GHz frequency range, the effect of flicker noise can be neglected.

Third, the noise current in the channel causes potential variation, which in turn causes gate current through the capacitive coupling of the gate capacitor. This gate noise current can be expressed as:

\[
\bar{i}_{\text{ng}} = 4K T \hat{\delta} g_v \Delta f \quad \text{where} \quad g_v = \frac{\omega^2 C_{gs}^2}{5g_{d0}}
\]

(3.12)

It can be seen from the equation that the gate noise is proportional to the square of the frequency. Thus the gate noise is a significant contributor at high frequency.

Fourth, the finite resistance of the substrate also contributes thermal noise. The variation of the substrate potential modulates the drain current due to the body effect and manifests itself at the output. The noise drain current due to substrate resistance can be expressed as:

\[
\bar{i}_{\text{ng,sub}}^2 = \frac{4K T R_{\text{sub}} g_m^2}{1 + (\omega R_{\text{sub}} C_{cb})^2} \Delta f
\]

(3.13)
The substrate noise is attenuated by the low pass filter consists of $R_{\text{sub}}$ and $C_{\text{cb}}$ at high frequency. The corresponding cutoff frequency in typical process is about 1 GHz. Thus for LNA operates at GHz range, the effect of substrate noise is not significant.

The noise in MOSFET can also be described using two port noise model. It can be shown that the corresponding noise parameters are [6]:

$$ R_n = \frac{\gamma}{\alpha} g_m $$
$$ G_u = \frac{\delta \omega^2 C^2_{gs}}{5g_{d0}} (1 - |\delta|^2) $$

$$ G_c \approx 0 $$
$$ B_c = \omega C_{gs} (1 - \alpha |\delta| \sqrt{\frac{\delta}{5\gamma}}) $$

The optimum source conductance and susceptance are:

$$ G_{\text{opt}} = \alpha \omega C_{gs} \sqrt{\frac{\delta (1 - |\delta|^2)}{5\gamma}} $$
$$ B_{\text{opt}} = -\omega C_{gs} (1 - \alpha |\delta| \sqrt{\frac{\delta}{5\gamma}}) \quad (3.14) $$

With optimum source impedance, the minimum noise figure that can be achieved for a given device and dc operating point is:

$$ F_{\text{min}} = 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_r} \sqrt{\gamma \delta (1 - |\delta|^2)} $$

(3.15)
3.3 UWB LNA Design Techniques and Tradeoffs

3.3.1 Typical LNA Topologies

First consider the common source stage which is a commonly used amplifier topology at low frequency, as shown in Figure 3.2-a. One problem of this topology is that the impedance seen at the input is $C_{gs}$ and $C_{gd}$ modified by Miller effect in parallel only. Thus it is difficult for such a topology to provide a real impedance of 50 $\Omega$. Transmission line and lump element can be used to match the input to 50 $\Omega$. However, such matching typically has very narrow bandwidth.

A direct method to provide a 50 $\Omega$ input termination for common source LNA is to use a 50 $\Omega$ shunt resistor at the input, as shown in Figure 3.2-b. This method provides reliable wideband input matching. However, there are many drawbacks associated with this topology. First, the shunt resistor attenuates the signal by 2 before the transistor, which reduces the gain of the LNA. Second, the shunt resistor adds significant thermal noise to the input of the transistor. Third, this topology can not provide an input impedance looking into the gate by the transistor that equals $Z_{opt}$, which is necessary for noise optimization. The combination of these effects results in rather high noise figure. Ignoring the gate current noise, the lower bound of the noise figure of this topology can be shown to be [6]:

Ignoring the gate current noise, the lower bound of the noise figure of this topology can be shown to be [6]:

\[ F_N = \frac{4kT}{50} \]

Where $k$ is the Boltzmann constant and $T$ is the absolute temperature.
\[ F \geq 2 + \frac{4\gamma}{\alpha} \cdot \frac{1}{g_m R} \quad (3.16) \]

For a practical design with typical short channel devices, this leads to a noise figure greater than 9 dB.

---

In order to alleviate the severe noise problem in the common source LNA with shunt resistor, resistive feedback amplifier topology is sometimes used to provide broadband
input matching, as shown in Figure 3.2-c. In this topology, the resistance seen at the input can be expressed as:

\[ R_{\text{in}} = \frac{R_F}{1 - A_y} \]  

(3.17)

Where \( A_y \) is the gain of the amplifier without feedback. Thus a much higher resistance value can be used for \( R_F \) to provide a 50 \( \Omega \) input termination. As a result, the thermal noise contribution of the feedback resistor is reduced to some extent, but still constitutes a non-negligible portion of the total noise. In addition, as in the case in common source stage with shunt resistor, this topology can not provide an impedance looking into the source from the transistor that equals \( Z_{\text{opt}} \). Thus, the noise optimization is still limited in this topology. However, the broadband matching capability makes it a practical choice for wideband application.

Common gate stage as shown in Figure 3.2-d is another frequently used topology that provides broadband input matching. The impedance looking into the source is approximately:

\[ R_{\text{in}} = \frac{1}{g_m + g_{mb}} \]  

(3.18)

Thus \( R_{\text{in}} \) can be set to 50 \( \Omega \) by careful selection of the device size and the bias current. Further, this impedance is not sensitive to the frequency, which makes it a favorable choice for wideband application. Since no explicit resistor is added, the noise performance of this topology is much better. It can be shown that the lower bound of the noise figure of common gate stage is:
\[
F \geq 1 + \frac{\gamma}{\alpha}
\]  
(3.19)

For a practical design with typical short channel devices, this leads to a noise figure of about 6 dB or higher.

### 3.3.2 Ultra-Wideband Input Matching

Common source LNA with inductive degeneration (as shown in Figure 3.3) is a widely used topology due to its capability to provide good input impedance match and optimize noise figure simultaneously. In this topology, the real part of the input impedance is obtained from phase lag between the drain current and the input voltage caused by the source inductor. The input impedance can be expressed as:

\[
Z_{in} = sL + \frac{1}{sC_{gs}} + \omega_r L \quad \text{where} \quad \omega_r = \frac{g_m}{C_{gs}}
\]  
(3.20)

In this equation, we can find that the real impedance of \( \omega_r L \) is generated without the thermal noise contribution of a real resistor. Further, by careful selection of gate width and source inductor, simultaneous input and noise match can be achieved. This greatly alleviates the difficult tradeoff between noise figure and input match.

In this topology, bandwidth of the input matching is determined by the quality factor of the matching network, which is defined as \( Q_s = \frac{1}{\omega C_{gs} R_s} \). Given the fixed real impedance of 50Ω and the minimum device size necessary for reasonable gain, the quality factor Q
can not be made arbitrarily small to meet the bandwidth requirement. With current state of art CMOS process, a bandwidth of up to 2 GHz can be achieved with this topology. For applications with wider bandwidth, extra input matching network becomes necessary.

In many recent works [12-14], wideband passive band pass filter is used together with the inductive degenerated common source amplifier in order to provide ultra-wideband simultaneous input and noise match. Figure 3.4 shows a LNA using 3rd order Chebychev band pass filter as matching network. In this topology, the gate capacitance and source inductance are part of the matching network. An extra capacitor $C_t$ is placed in parallel with the $C_{gs}$ in order to adjust the total capacitance to the value required for simultaneous input and noise match. Good input matching ($S_{11} < -10\text{dB}$) and low noise figure (NF < 5dB) can be achieved over ultra wide frequency range from 3 to 10 GHz.

Figure 3.3: Common source with inductive source degeneration
3.3.3 Ultra-Wideband Noise Optimization

For most CMOS process, the optimum source impedance looking from the transistor is inductive in nature. For simultaneous input and noise match, $G_{\text{opt}}$ has to be tuned to 1/50 S. In addition, the input impedance seen at the gate of the core transistor must exhibit a real part of 50 $\Omega$ and an inductive imaginary part that equals the conjugate of $B_{\text{opt}}$ in order to achieve simultaneous input and noise match.

In equation (3.14), $\alpha$ is a parameter that reflects the degree to which the operation of the device deviates from the long channel regime, which is defined as:
\[ \alpha = \frac{g_m}{g_{d0}} \]  

At high bias voltage which is required for linearity enhancement, the short channel effect becomes more magnificent and \( \alpha \) takes smaller value. In order for \( G_{\text{opt}} \) to remain at the level of 1/50 S, \( C_{gs} \) has to be increased by increasing the device width. On the other hand, high biasing voltage results in high current density which mandates that the device width to be reduced in order to keep reasonable power consumption. This problem can be solved by choosing smaller device width to reduce the drain current and adding a capacitance \( C_t \) across the gate to achieve simultaneous input and noise match.

Several considerations and tradeoffs are important in UWB LNA design. First, the noise performance of MOSFET degrades as channel length shrinks. Using larger channel length improves the noise figure. Longer channel length, however, degrades gain and linearity. Second, high Q inductor should be used for the gate inductor. The noise figure of the LNA is very sensitive to the thermal noise at the input. The parasitic resistance of the gate inductor can add noise to the LNA. By using high Q inductor with smallest possible inductance, this effect can be minimized. Third, when other conditions remain the same, higher transconductance of the \( g_m \) stage improves noise figure. This is because higher gain reduces the input referred noise of the drain current noise. Fourth, though its effect is less significant, the substrate resistance should be minimized in order to reduce its thermal noise contribution. Good layout practice is crucial to reduce the substrate noise.
3.3.4 Ultra-Wideband Gain Response

In RF design, the power gain is often of most concern instead of voltage gain because the signal to noise ratio is determined by the power of the signal to that of the noise. The gain of a component in RF system is often referred to the power gain. In S-parameter measurement, the gain is measured by $S_{21}$, which is expressed as:

$$S_{21} = \frac{b_2}{a_1} \bigg|_{a_2=0}$$

(3.22)

Where $a_1 = \frac{1}{2\sqrt{Z_0}}(V_1 + Z_0 I_1)$ \quad $b_2 = \frac{1}{2\sqrt{Z_0}}(V_2 - Z_0 I_2)$

The main purpose of the LNA stage is to provide sufficient gain to suppress the noise of the following noisy stages. The noise factor of the cascaded stages can be calculated as:

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + ......$$

(3.23)

Since the gain of the first stage (LNA) $G_1$ shows up in the denominator of all terms except the first, it is important to make $G_1$ high enough to minimize the effect of the noise in the following stages. The level of gain that is required also depends on the noise figure and gain of the following stages.

The gain of the LNA can be improved through several methods. First, the bias current of the LNA can be increased to improve $g_{m}$, which can be expressed as:

$$g_m = \sqrt{2\mu C_{ox} \frac{W}{L} I_D}$$

(3.24)
Assume device size keeps unchanged, the $g_m$ increases with the square root of the drain current. Thus a higher gain can be achieved by increasing the drain current. However, this improvement is at the expense of higher power consumption. In addition, for short channel device in strong inversion region, the $g_m$ increases even slower with increasing $I_D$ due to the short channel effect. Thus the improvement of gain by increasing bias current can be limited in many cases.

Second, the gain of the LNA can also be improved by using larger W/L ratio while keeping drain current constant. This relationship can also be seen from the expression of $g_m$. However, this method results in a lower gate source voltage $V_{gs}$, which degrades the linearity of the LNA. In the extreme case when the LNA enters weak inversion region, both linearity and noise figure degrade severely.

Third, the gain of the LNA can also be improved by increasing the quality factor of the LNA. Since the input signal is amplified by the quality factor $Q$ before it is seen by the transistor, the voltage gain of the LNA is proportional to $Q$. Thus increasing the quality factor $Q$ by a factor of two can potentially increase the gain by 6 dB. However, the bandwidth will be reduced by half accordingly. The choice of $Q$ involves the tradeoff between the gain and bandwidth.

Fourth, the choice of load impedance is another tradeoff between gain and bandwidth. Higher load impedance increases the gain of the LNA, but reduce the bandwidth at the
same time for a given load capacitance. Shunt peaking is a frequently used technique to improve the bandwidth by pushing the onset of the gain roll off to higher frequency. For a given band width, higher gain and flatter gain response can be achieved with the application of shunt peaking.

The improvement methods and tradeoffs described above are all subject to the constraint of power, noise, stability, and linearity. Most importantly, the impedance seen by the transistor at the input should be matched to the $Z_{opt}$ which optimize the noise figure. These constraints prevent very high gain to be realized for LNA. In most cases, the gain of the UWB LNA falls in the range from 10 to 20 dB.

### 3.4 Single-ended to Differential Conversion

#### 3.4.1 The Necessity of On-chip CMOS Active Balun

In order to minimize the number of expensive off-chip components, off-chip balun is typically undesirable. In addition, a single-ended RF input port simplifies the board level design. Thus single-ended LNA is a much more desirable solution for RF front-end circuit. On the other hand, however, the following mixer stage needs balanced signal in order to minimize the direct signal feed through. As a result, an on-chip balun is necessary to convert the single-ended signal to a balanced differential signal. This balun
has to provide relatively flat gain response with minimal gain and phase imbalance over the desired bandwidth, which is a rather challenging task to achieve with on-chip components.

### 3.4.2 Common Source Differential Balun

The simplest type of active balun is shown in Figure 3.5. It is basically a common source differential amplifier with one differential input tied to RF input and the other tied to AC ground. Assuming a perfect current source with infinite output impedance, the AC currents in the two differential paths are forced to have same amplitude but different direction. Thus a differential signal is achieved at the output node.

![Schematic of common source differential balun](image)

Figure 3.5: Schematic of common source differential balun [15]

This topology, however, has many limitations. Among them the most severe problem is that it is impossible to have a current source with very high output impedance at multi-
GHz frequency. In reality, the parasitics can greatly reduce the impedance at node X at high frequency. As a result, a significant part of AC current leaks to the ground instead of injecting into the source of M2, which leads to a much larger gain of left branch than that of the right branch. The signal imbalance with this topology can be greater than 10 dB in some cases.

### 3.4.3 Common Source Differential Balun with Compensation Feedback

In order to solve the problem described above, an improved design was developed and demonstrated in some recent works [16]. The major difference here is the connection of the other differential input. Instead of being tied to the AC ground, an RLC feedback network is used to form a feedback path from the output of the left path to the differential input on the right, as shown in Figure 3.6. Since the output of the common source stage has a 180° phase difference from the input, the feedback signal can be used to generate the negative signal. The inductor here is used to resonant out the gate capacitance and compensate the phase delay. This network, however, is frequency dependent and is difficult to accommodate very wide bandwidth.
3.4.4 Common Source Balun with Current Reuse

In order to solve the bandwidth problem and reduce the power, a common source CMOS balun with current reuse is proposed in [17], as shown in Figure 3.7. In this circuit, a NMOS is used as common source amplifier and a PMOS transistor is used as a common gate amplifier. Both transistors share the same DC bias current thus reduces the total power consumption. The transistors are sized such that the two outputs have equal gain but opposite phase.
This circuit was designed in 0.18u CMOS process. According to the author, this balun exhibits less than 4 dB of gain imbalance and 8 degree of phase imbalance from DC up to 10 GHz, which makes it a promising topology for UWB application.
4.1 Significance of Front-end Linearity in UWB Receiver

As is commonly known, the noise figure is primarily determined by the earlier stages in an RF system, while the linearity is primarily determined by the later stages. This is why the noise figure has long been used as the major figure of merit for LNA. Then a natural question that follows is: why should we add the optimization of linearity to the already complicated task of UWB LNA design?

In addition to providing sufficient gain to suppress the noise in the following noisy stages (active balun and mixer) while adding minimal noise of itself, the LNA must also amplify the signal with minimal distortion. When only anticipated signal presents, this is typically not an issue since the received signal level in UWB communication is usually not high enough to cause severe distortion. However, in presence of strong interferers (e.g.
802.11a), the nonlinearity of the LNA will result in different adverse effects such as desensitization, cross modulation and intermodulation, which distort the signal and degrade the performance of the receiver.

![Typical direct conversion receiver architecture](image)

**Figure 4.1**: Typical direct conversion receiver architecture

In most conventional receiver architecture as shown in Figure 4.1, there is only one signal path from the front-end to the ADC. In addition, the out-of-band interferers are attenuated significantly at the input of LNA. In this case, the latter stages in the receiver path usually become saturated before the LNA does. The IIP3 of the overall receiver can be expressed as:

\[
\frac{1}{IIP_{3,6_{\text{out}}}} = \frac{1}{IIP_{3,1}} + \frac{G_1}{IIP_{3,2}} + \frac{G_1 G_2}{IIP_{3,3}} + \frac{G_1 G_2 G_3}{IIP_{3,4}} + \text{......} \quad (4.1)
\]

Thus the linearity of the overall receiver is mainly determined by the later stages but only slightly affected by the linearity of the LNA. When a strong narrowband interferer is present in band, this type of receiver can only reduce the gain of VGA to trade noise for IIP3, in which case the linearity can be limited by the mixer or the lowpass filter.
In narrowband system, the chance that a strong narrowband interferer falls in the signal band is small. RF filter can be used to efficiently protect the front-end circuit from the interference of other radio transmissions. However, in UWB communication system, the bandwidth of the signal can easily span from 500MHz to several GHz. Thus it is practically impossible to prevent strong interferers from falling into this ultra wide passband. Notch filters can be used to attenuate known narrowband filters. But unexpected interferers, which are present from time to time especially in environment with abundance of RF transmissions, can not be dealt with in this manner.

![Figure 4.2: UWB and potential interferers transmit power](image)

Further, since the FCC part 15 mandates that the transmit power of the UWB system to be lower than -41.3 dBm in 3.1-10.6 GHz frequency range, the signal level of UWB
system is much smaller than typical narrowband wireless communication standard. As shown in Figure 4.2, the power level of 802.11b is 42 dB higher than (or 15,850 times) that of the UWB system while the power level of 802.11a is 34 dB higher than (or 2,510 times) that of the UWB system. Thus the RF front-end circuit with poor linearity can easily be blocked with these potential interferers.

Many recent research works proposed receiver architectures that divide wideband signal into channels or subbands and process them separately [18-21]. For instance, an OFDM UWB receiver that employs the concept of ADC in frequency domain is proposed in [22]. The architecture of the receiver is shown in Figure 4.3.

In this type of receiver architecture, if the strong narrowband interferers are present, the base band circuit for one or several subband(s) may be blocked, while other baseband
circuit can still process the incoming signal using other subbands. Thus with well
designed communication protocol, the data stream can be kept uninterrupted. However, if
the interferer is strong enough to saturate the LNA which processes the signal in all
frequency bands, the whole receiver will be blocked. In these cases, the linearity of the
LNA determines the robustness of the receiver to strong interferers. Thus a UWB low
noise amplifier (LNA) with high linearity is very desirable for UWB receiver.

4.2 Effect of Non-linearity in UWB Front-end

Due to its ultra wide bandwidth, the UWB receiver differs in many ways from its
narrowband counterparts. Thus it is worthwhile to reevaluate the effect of non-linearity in
UWB LNA. In conventional narrowband receiver, the main purpose of linearization is to
minimize the IM products that can potentially fall in band (e.g. 3\textsuperscript{rd} order IM product).
However, the wide bandwidth of UWB makes the receiver front-end prone to IM
products of different orders as well as strong narrowband interferers. This problem of
narrowband nature can be solved using the channelized receiver architectures or
advanced modulation techniques. In contrast, the efforts in UWB LNA linearization
should be focused on the effect of strong narrowband interferers on the whole frequency
band, or at least a significant portion of the overall bandwidth.

A memoryless, time-variant nonlinear amplifier can be described as truncated power
series which ignore the contribution of higher order nonlinearity:
\[ i(v) \approx c_0 + c_1 v + c_2 v^2 + c_3 v^3 \]  

(4.2)

Suppose a strong narrowband interferer \( A_1 \cos \omega t \) is present at the input of the amplifier, its effect on an arbitrary signal tone \( A_2 \cos \omega t \) can be studied by looking into the output of the amplifier. If the relative bandwidth is smaller than 50%, the harmonics and 2\textsuperscript{nd} order IM products fall out of band. Thus only fundamentals and 3\textsuperscript{rd} order IM products are considered.

\[ \omega_1 : (c_1 A_1 + \frac{3}{4} c_3 A_1^3 + \frac{3}{2} c_3 A_1 A_1^2) \cos \omega t \]  

(4.3)

\[ \omega_2 : (c_1 A_2 + \frac{3}{4} c_3 A_2^3 + \frac{3}{2} c_3 A_2 A_2^2) \cos \omega t \]  

(4.4)

\[ 2\omega_1 - \omega_2 : \frac{3 c_3 A_1^2 A_2}{4} \cos(2\omega_1 - \omega_2) t \]  

(4.5)

\[ 2\omega_2 - \omega_1 : \frac{3 c_3 A_2^2 A_1}{4} \cos(2\omega_2 - \omega_1) t \]  

(4.6)

First, the magnitude of the fundamental at \( \omega_2 \) is a function of the magnitude of the interferer \( A_1 \), as can be seen in equation (4.4). Assuming \( A_1 \gg A_2 \), the gain of the fundamental at \( \omega_2 \) can be approximated as \( (c_1 + \frac{3}{2} c_3 A_1^2) \). Since \( c_3 \) is negative for typical MOSFETs in strong inversion, the gain at \( \omega_2 \) is a decreasing function of \( A_1 \). This desensitization process occurs in the whole frequency band and thus needs to be taken into consideration. In extreme case, the small signal gain drops to zero and the amplifier is completely blocked.
Second, if the power level of the strong narrowband interferer varies with time, the amplitude of the fundamental at $\omega_2$ is modulated by the amplitude of the interferer, which is referred to as cross modulation. Similarly, this cross modulation affects all the signals in band.

Third, the effect of 3rd order intermodulation is considered. Figure 4.4 illustrates the relative magnitude of the fundamentals and the 3rd order IM products. Comparing equation (4.5) and (4.6), it can be observed that the two 3rd order IM products are not equal in magnitude. The IM3 on the interferer side is stronger than the IM3 on the signal side, with a magnitude ratio of $A_1/A_2$. Considering that the radio emission level of UWB signal is limited to -41.3 dBm/MHz maximum by FCC part 15 regulation, this magnitude ratio can be rather large. Thus the IM3 on the signal side is negligibly weak and only the IM3 on the interferer side needs to be considered.

Figure 4.4: Fundamentals and the 3rd order IM products
Now consider the intermodulation effect of strong narrowband interferer on the whole frequency band. A strong narrowband interferer essentially mirrors the whole spectrum around itself, attenuates the power level by \((4c_i)/(3c_iA_i^2)\) and adds to the original spectrum, as shown in Figure 4.5. Thus the power of the interferer is spread into the whole bandwidth through 3\(^{rd}\) order intermodulation and increases the noise floor.

![Figure 4.5: Effect of 3\(^{rd}\) IM on UWB signal](image)

The above discussion leads to the conclusion that desensitization (blocking), cross modulation and intermodulation are three major mechanisms that the strong narrowband interferer corrupts signal in a wide frequency range. Thus efforts should be made to improve the linearity of the UWB LNA and minimize these effects.
4.3 Measure of Linearity for UWB Receiver Front-end

IIP₃ has long been adopted as a measure of linearity for amplifiers due to its significance to narrowband system and established measurement techniques. However, in the context of UWB system, the justification of IIP₃ as an index of linearity needs to be reexamined.

In the three mechanisms discussed above that reveal the major effects of strong narrowband interferer to UWB system, the 3rd order distortion of the amplifier c₃ plays an important role. Minimization of c₃ alleviates all three effects and reduces the corruptive effect of interferers on UWB signal. Since IIP₃ is an indicator of the 3rd order distortion, it remains a good measure of linearity for UWB system. On the other hand, P₁dB indicates the ability of the amplifier to handle large signals without severe distortion. It is complementary to IIP₃ and is another important measure of linearity of front-end circuits. In blocks that work with large signal (e.g. power amplifier), P₁dB becomes the major figure of merit.

From equation (4.2), it is straightforward to derive the expression for the amplitude of the input signal at 3rd order intercept point and the 1dB compression point [23]:

\[ A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{c_1}{c_3} \right|} \]  \hspace{1cm} (4.7)

\[ A_{1dB} = \sqrt{0.143 \left| \frac{c_1}{c_3} \right|} \]  \hspace{1cm} (4.8)
Thus by taking the ratio of equation (4.7) and (4.8), it can be concluded that $A_{1\text{dB}}$ is 9.6 dB lower the $A_{\text{IP}_3}$. This conclusion is frequently cited in literature. However, the measurement results very often deviate from this relationship. The reason of this is that such conclusion depends on a fundamental assumption which is seldom mentioned explicitly: it is assumed that the power coefficients keep constant with increasing input power level. In CMOS RFIC circuit in today’s dominant process, this assumption is rarely satisfied. Since the power coefficients are derived from the Taylor series to approximate the transfer function at the vicinity of the operating point, their values depends on not only the DC operating point, but also the voltage swing magnitude of the input signal. As a result, the power coefficients are functions of input power level and their sensitivity is determined by the DC operating point of the amplifier. For instance, the amplifier biased at optimum gate bias voltage for high linearity has the power coefficients that are very sensitive to bias point and power level, which leads to a much lower $P_{1\text{dB}}$ than that predicted by the $\text{IIP}_3$ value. On the other hand, the amplifier biased at high gate bias voltage has the power coefficients that are less sensitive to bias point and power level, in which case the $P_{1\text{dB}}$ can be better predicted by the $\text{IIP}_3$ value.
CHAPTER 5

LINEARITY IMPROVEMENT THEORY AND TECHNIQUES

5.1 Previous Works

5.1.1 Linearization through Optimum Gate Biasing

Several approaches that improve linearity of LNA have been reported recently. The approach in [1] optimizes the linearity of LNA by biasing the LNA at the optimum biasing point. The optimum biasing at DC is the $V_{gs}$ at which the 3rd order distortion of the MOSFET equals zero, as shown in Figure 5.1 and Figure 5.2. It is proved that the optimum biasing point at RF exhibits a small offset from its DC value. One major limitation of this method is that the IIP3 roll off very fast from its peak value when $V_{gs}$ deviates from the optimum value. Significant IIP3 improvement heavily depends on very accurate biasing (within a couple of mV from the optimum bias voltage), which is very difficult to achieve with process and temperature variation. A narrowband LNA with 10.5
dBm IIP3, 1.8 dB NF, and 14.6 dB power gain designed in 0.25 µm CMOS process was demonstrated in \[1\]. However, the gate bias is provided from off-chip which avoids the biggest problem in this approach, thus a more practical implementation is yet to be developed.

![Figure 5.1](image1.png)

Figure 5.1 1\textsuperscript{st}, 2\textsuperscript{nd} and 3\textsuperscript{rd} order power series coefficients vs. \(V_{gs}\) \[1\]

![Figure 5.2](image2.png)

Figure 5.2 DC optimum biasing point \[1\]

The accurate biasing can be realized on chip using negative feedback \[1\]. However, the biasing voltage thus obtained is very sensitive to the mismatch and temperature/process variations. Further, the effect of such method is limited at high frequency (GHz range)
since the optimum bias voltage changes with frequency due to the stronger feedback through inductor and capacitors.

### 5.1.2 The Derivative Superposition Method

The derivative superposition method is first proposed in [24] and is thoroughly elaborated in [2]. This technique greatly alleviates such stringent biasing requirement to achieve high linearity. A circuit schematic that illustrate this technique is shown in Figure 5.3.

![Composite CMOS for 3rd order distortion cancellation](image)

**Figure 5.3 Composite CMOS for 3rd order distortion cancellation [3]**

![3rd order coefficients of individual MOSFET and composition MOSFET](image)

**Figure 5.4 3rd order coefficients of individual MOSFET and composition MOSFET [3]**
The idea of this technique is to bias two MOSFETs at different $V_{gs}$ and combine their drain current. The drain current of the obtained composite MOSFET has a third derivative that equals the sum of that of the individual MOSFETs. By adjusting the sizes and biasing voltages of two MOSFETs, a relatively flat third derivative at zero crossing can be obtained, as shown in Figure 5.4. Thus the biasing requirement for high IIP$_3$ can be greatly relaxed.

5.1.3 The Modified Derivative Superposition Method

The linearization technique described above works effectively at low frequency. At RF frequency, however, the IIP$_3$ improvement is very limited due to the source degeneration inductance. The source degeneration inductance forms feedback path from drain current to $V_{gs}$ and thus results in a contribution of 2$^{nd}$ order non-linearity of drain current to IIP$_3$. The IIP$_3$ at RF frequency is derived in [3]:

$$IIP_3 = \frac{4g_1^2\omega^2 L C_{GS}}{3|\varepsilon(\Delta\omega, 2\omega)|}$$

$$\varepsilon(\Delta\omega, 2\omega) = g_3 - \frac{2g_2^2/(3g_1)}{1 + \frac{g_2}{2\omega L g_1} + \frac{j2\omega}{\omega_T} + \frac{Z_L(2\omega)}{\omega_T L}}$$

Thus the IIP$_3$ can not be optimized simply by setting the third derivative $g_3$ to zero. The effect of second derivative $g_2$ also needs to be considered. In addition, care should be used to minimize the imaginary part of the $\varepsilon$, which will otherwise set the upper limit of the achievable IIP$_3$. Figure 5.5 shows the effect of the source degenerate inductance on
the linearity of the amplifier. It can be seen that the $I_{IP3}$ peak diminishes at high frequency.

Figure 5.5 $I_{IP3}$ of LNA with different frequency and source degeneration inductance [3]

Smaller source degeneration inductance alleviates the problem described above but makes the input and noise match very difficult. Thus a modified DS method is proposed in [3] to boost $I_{IP3}$ at RF frequency. The circuit is shown Figure 5.6.

Figure 5.6 Modified DS method circuit [3]
The idea of this technique is to tune the circuit such that the contribution of the 2nd order to the IM3 has the same magnitude and opposite phase with the 3rd order contribution. This is made possible with an additional inductor at the source of the MB. With this method, a very high IIP3 (22 dBm) is achieved in 0.25 μm CMOS process [3]. However, it is narrowband in nature because it relies on accurate magnitude and phase match, which is not practical in UWB applications.

### 5.1.4 Feed Forward Linearization

Feed forward technique has long been used for the linearization of power amplifiers. For a nonlinear amplifier, the output signal can be divided into two components: the amplified signal assuming ideally linear amplifier, and the error signal which represents the deviation of the real output from the ideal one. The idea of feed forward linearization is to compute this error signal and subtract it from the output after proper scaling.

![Feed forward linearization topology](image)
A typical feed forward linearization topology is shown in the Figure 5.7 [23]. The input signal $V_{in}$ is first amplified by the main nonlinear amplifier and the signal at M can be written as $V_M = A_v V_{in} + V_D$, where $V_D$ is the error signal. Then $V_M$ is divided by $A_v$ and subtracted by the input signal $V_{in}$. Thus $V_N = V_{in} + V_D/A_v$, and $V_P = V_N - V_{in} = V_D/A_v$. The separated error signal is then scaled by the error amplifier and subtracted from $V_M$. Thus $V_{out} = V_M - V_P A_v = A_v V_{in}$. The distortion caused by the nonlinearity of the main amplifier is removed. Since the two amplifiers cause significant phase shift at high frequency, delay line $\Delta_1$ and $\Delta_2$ are used to match the phase for signals from different path.

A 18 dBm IIP3 LNA is designed using feed forward cancellation in 0.35 $\mu$m CMOS in [25]. The design achieves high linearity while maintains good noise figure and gain. The drawback of this technique is that it requires off-chip signal splitting components.

Other than providing high linearity, the added benefit of feed forward technique is the inherent stability associated with this topology. The major limitation of this technique is that the achievable IIP3 is very sensitive to mismatches. The inevitable device mismatches result in phase mismatch and gain mismatch between different paths. The suppression of the magnitude of the IM products in $V_{out}$ can be calculated as [23]:

$$ E = \sqrt{1 - 2 \left( 1 + \frac{\Delta A}{A} \right) \cos \Delta \phi + \left( 1 + \frac{\Delta A}{A} \right)^2} \quad \text{(5.3)} $$

Thus a poor matching can severely degrades the linearity performance.
The LNA discussed above is not designed for wideband application. Thus the dependence of the linearity on the frequency is not discussed. Due to the sensitiveness of this technique to the mismatches, the high linearity can only be achieved in a narrow frequency band and is not suitable for UWB application.

5.1.5 Active Post Distortion

An active post distortion technique to boost IIP3 of LNA is proposed in [26]. This method achieved 8 dBm IIP3 and 16.2 dB power gain in 0.25 μm CMOS process. A simplified schematic of this method is shown in the Figure 5.8.

![Figure 5.8 APD high linearity LNA [26]](image)

The idea of this method is to use active device to modify the output of the amplifier in order to achieve a linear transfer function. It can be proved that the 3rd order coefficient $g_3$ can be calculated as:
Where \( \alpha \) and \( \beta \) are the ratio of transconductance between \( M_1 \) and \( M_2 \) and \( M_1 \) and \( M_3 \), respectively. By careful selection of \( \alpha \) and \( \beta \) values, \( g_3 \) can be minimized and IIP3 can be improved. This technique is similar to derivative superposition method in a sense that the total output current consists of the drain current of main and auxiliary device whose nonlinearity tends to cancel each other. There are, however, two major differences between the two techniques. First, the 2\textsuperscript{nd} order nonlinearity of the main and auxiliary devices generates 3\textsuperscript{rd} order IM current component in the drain current of the auxiliary device, which makes the \( g_3 \) of the amplifier a strong function of \( g_2 \) of individual devices. Second, the input of the auxiliary device comes from the output of the main device instead of the input signal, which can cause noticeable phase shift which depends on frequency. With this configuration, the phase match has to be considered in order for this technique to be effective at high frequency. Thus it can be very difficult for this technique to achieve ultra-wideband linearization.

5.2 Linearity Improvement Methods and Tradeoffs

The linearization techniques discussed above have proved the effectiveness of those techniques in narrowband applications such as cellular RF front-end. However, these techniques are all base on precise distortion cancellation at certain frequencies. The
obtained high IIP$_3$ is very sensitive to circuit mismatch as well as the frequency offset. Thus these techniques can not be directly applied to UWB applications.

Currently, the UWB LNAs are usually designed without special treatment on the linearity. The current state of art UWB LNAs exhibit IIP$_3$ values of about -5 dBm, which are not sufficient to guarantee high receiver performance in noisy environment.

The difficulties in designing high linearity UWB LNA lie in the different requirements for the optimization of input matching, noise figure, linearity, power gain and stability within the power constraint. In addition, such simultaneous optimization has to be maintained over a large bandwidth required for the UWB communication.

In many UWB LNA designs, source follower is used as the output stage. Such implementation provides convenient output impedance match over broad frequency range. However, there are many problems associated with the source follower stage. Since the output stage is usually biased at much higher voltage level in order to optimize the linearity, its drain current can be comparable to the first stage, which results in significant additional power consumption. The MOSFET source follower typically has a loss of 2 dB and even worse at GHz frequency. As a result, the overall gain of the LNA will be decreased, which will in turn degrade the noise figure. In addition, the noise contributed by the second stage MOSFET will degrade the noise figure even further. Most importantly, since the later stage has a more significant effect on linearity, the non-
linearity of the output stage MOSFET can reduce the IIP$_3$ of the overall LNA by 5-10 dBm. Thus to avoid using active device for output impedance match is very critical in designing high linearity LNA. Passive impedance matching network has been successfully applied to UWB LNA for output impedance match [13]. Wideband matching network can be designed to provide a flat gain over the bandwidth.

Careful selection of quality factor of LNA also helps in improving the linearity of the LNA. The quality factor of LNA is defined as $Q_s = \frac{1}{\omega C_{gs} R_s}$, which is the Q of the series RLC network at the input. Since the signal is amplified by $Q_s$ before applied to the gate, high $Q_s$ will increase the overall gain but decrease the linearity of the overall LNA. Thus the gain can be traded off for linearity by reducing the quality factor of LNA.

The shrinking feature size of CMOS process has a positive effect on the linearity of the LNA. For short channel devices, high electrical field results in velocity saturation of carriers. Thus the transconductance of the MOSFET approaches a constant value at high bias voltage, which results in a much more linear behavior than its long channel counterpart. By increasing the bias voltage $V_{gs}$, the linearity of the LNA can be greatly improved. However, this will make it more difficult to achieve simultaneous input and noise match. The power consumption will also increase accordingly.

Several important tradeoffs exist in UWB LNA design between linearity and other desired performances. In order to achieve high linearity, high bias voltage is required,
which in turn results in larger drain current and higher power consumption. The power can be reduced by decreasing the width of the MOSFET. However, a too small MOSFET width makes it very difficult to achieve the input and noise match simultaneously, which degrades either gain or noise figure. Tradeoff between gain and linearity can also be seen in determining the quality factor of the LNA. By tuning the quality factor $Q_s$ to a lower value, the gain decreases and $IIP_3$ increases by the same amount (in dB). Thus excess gain can be traded off for linearity while making the choice of quality factor.

5.3 Short Channel MOSFET Device Modeling

In order to gain deeper insights on the nonlinear behavior of the CMOS amplifier, the MOSFET device model is first studied. For long channel MOSFET in saturation region, the device behavior is well described by the classic equation:

$$I_{ds} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{gs} - V_f)^2$$  \hfill (5.5)

With aggressive scaling of CMOS process, deep submicron devices with feature size of around 100 nm are widely used nowadays. The small dimension of the device leads to very high electric field even at moderate voltage (~1V), which in turn leads to a variety of high field effects that differentiate short channel devices from long channel devices. Velocity saturation, as one of these effects, has the fundamental significance on short channel device behavior. As the electric field increases to a certain level (~$10^6$ V/m), the drift velocity of the electron no longer increase linearly with electric field due to the
scattering by phonons. Thus the electron drift velocity saturates at a maximum value of about $10^5$ V/m. A short channel device model which taking into account the velocity saturation was developed in [27]:

$$I_{ds} = \frac{1}{2} \mu_n C_{ax} \frac{W}{L} \frac{(V_{gs} - V_t)^2 (E_{sat})^2}{(V_{gs} - V_t) + (E_{sat})}$$

(5.6)

Where $E_{sat}$ is the electric field strength at which the carrier drift velocity dropped to half of the linear extrapolated value.

### 5.4 Linearity Modeling and Analysis at Low Frequency

The small signal current of the nonlinear amplifier can be expended into the power series in terms of small signal input voltage around biasing point:

$$i(V_{gs}) = g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3$$

(5.7)

The higher order terms are truncated since their effects are usually negligible. The IIP3 of the amplifier is directly related to the first and third order coefficients of the power series as:

$$A_{IP3} = \sqrt[3]{\frac{4}{3} \frac{g_1}{g_3}}$$

(5.8)

It can be seen from the equation above that when $g_3$ is zero, the IIP3 is infinity. In reality, IIP3 will form a peak at this point instead of going to infinity. The reason of this is two fold: first, the effect of higher order distortion which has been neglected will results in a

58
finite IIP3; second, when the signal level is getting larger, the coefficients of the power series deviates from their small signal values.

The coefficients are functions of $V_{gs}$ and can be expressed as:

$$g_1(V_{gs}) = \frac{\partial I_D}{\partial V_{gs}}$$  \hspace{1cm} (5.9)

$$g_2(V_{gs}) = \frac{1}{2} \frac{\partial^2 I_D}{\partial V_{gs}^2}$$  \hspace{1cm} (5.10)

$$g_3(V_{gs}) = \frac{1}{6} \frac{\partial^3 I_D}{\partial V_{gs}^3}$$  \hspace{1cm} (5.11)

Thus once the DC characteristic of the transistor is known, $g_1$, $g_2$, $g_3$ and $A_{IIP3}$ as functions of $V_{gs}$ can be obtained.

### 5.4.1 Linearity Modeling in Strong Inversion Region

Substitute equation (5.6) into (5.9), (5.10) and (5.11), $g_1$, $g_2$ and $g_3$ can be expressed as a function of $V_{gs}$:

$$g_1(V_{gs}) = \frac{1}{2} \mu_s C_{ox} W E_{sat} \left\{ 1 - \frac{(LE_{sat})^2}{(V_{gs} - V_t) + (LE_{sat})^2} \right\}$$  \hspace{1cm} (5.12)

$$g_2(V_{gs}) = \frac{1}{2} \mu_s C_{ox} W L^2 E_{sat}^3 \frac{1}{(V_{gs} - V_t) + (LE_{sat})^3}$$  \hspace{1cm} (5.13)

$$g_3(V_{gs}) = -\frac{1}{2} \mu_s C_{ox} W L^2 E_{sat}^3 \frac{1}{(V_{gs} - V_t) + (LE_{sat})^3}$$  \hspace{1cm} (5.14)
Substitute equation (5.12) and (5.14) into equation (5.8), the expression of IIP3 in strong inversion region can be written as:

\[
A_{IP3}(V_{gs}) = \sqrt{\frac{4}{3}} \left[ 1 - \frac{(LE_{sat})^2}{(V_{gs} - V_t) + (LE_{sat})^2} \right] \left( \frac{(V_{gs} - V_t) + (LE_{sat})}{LE_{sat}} \right)^{3/2} \tag{5.15}
\]

It can be seen from the above equation that \( A_{IP3} \) is a monotonically increasing function with \( V_{gs} \) in strong inversion region. At high overdrive voltage where \( V_{gs} - V_t > LE_{sat} \),

\[
1 - \frac{(LE_{sat})^2}{(V_{gs} - V_t) + (LE_{sat})^2} \approx 1
\]

which leads to:

\[
A_{IP3}(V_{gs}) = \sqrt{\frac{4}{3}} \left( \frac{(V_{gs} - V_t) + (LE_{sat})}{LE_{sat}} \right)^{3/2} \tag{5.16}
\]

Thus increasing the gate bias voltage of the linearity critical device is a convenient way of improving linearity.

### 5.4.2 Linearity Modeling in Weak Inversion Region

Equation (5.6) predicts that the drain current \( I_{ds} \) diminishes when the overdrive voltage approaches zero. However, even at gate bias slightly lower than threshold voltage, there still exists a non-negligible amount of drain current. This is because the electrons start to build up before strong inversion is formed. As long as a small voltage is applied between drain and source, the drift of these electrons will form the subthreshold current, the magnitude of which can be expressed as [6]:

...
\[ I_{ds} = I_{ON} \exp \left( \frac{q(V_{gs} - V_t)}{nkT} - 1 \right) \]  

(5.17)

Where

\[ n = 1 + \frac{C_D}{C_{ox}} \]

\[ C_D \] is the depletion capacitance defined by \[ C_D = \frac{\varepsilon_{si}}{x_d} \]

\[ I_{ON} \] is the drain current at the boundary of the strong inversion and weak inversion, \( V_{ON} \).

\[ V_{ON} = V_t + \frac{nkT}{q} \]  

(5.18)

By equalizing the right hand side of equation (5.6) and (5.14), the value of \( I_{ON} \) to assure the continuity of function at the boundary can be derived as:

\[ I_{ON} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left( \frac{nkT}{q} \right)^2 \left( L + L_{sat} \right) \]  

(5.19)

Substitute equation (5.17) into (5.9), (5.10) and (5.11), \( g_1, g_2 \) and \( g_3 \) can be expressed as a function of \( V_{gs} \):

\[ g_1(V_{gs}) = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left( \frac{nkT}{q} \right) \left( L + L_{sat} \right) \exp \left( \frac{q(V_{gs} - V_t)}{nkT} - 1 \right) \]  

(5.20)

\[ g_2(V_{gs}) = \frac{1}{4} \mu_n C_{ox} \frac{W}{L} \left( \frac{L_{sat}}{q} \right) \exp \left( \frac{q(V_{gs} - V_t)}{nkT} - 1 \right) \]  

(5.21)
\[ g_3(V_{gs}) = \frac{1}{12} \mu_n C_{ox} \frac{W}{L} \left( \frac{nkT}{q} \right)^{-1} (LE_{sat}) \exp\left( \frac{q(V_{gs} - V_t)}{nkT} - 1 \right) \]  

(5.22)

Substitute equation (5.20) and (5.22) into equation (5.8), the expression of IIP3 in weak inversion region can be written as:

\[ A_{IIP3}(V_{gs}) = 2\sqrt{2} \frac{nkT}{q} \]  

(5.23)

At room temperature, this equation leads to about 75 mV or -12.5 dBm. Interestingly, the \( A_{IIP3} \) depends on temperature only for a given device.

### 5.4.3 Interface between Two Models

Thus far the analytical solution of \( g_1, g_2, g_3 \) and \( A_{IIP3} \) has been derived for strong inversion case and the weak inversion case. However, though the two drain current equations can be made continuous at \( V_{gs} = V_t \), they are not continuously derivable. Thus the equations of \( g_1, g_2 \) and \( g_3 \) are continuous at \( V_{gs} = V_t \). Further, both equations lose their accuracy when \( V_{gs} \) approaches \( V_t \). As a result the analytical model thus derived are valid only when \( V_{gs} > V_t + 0.5\cdot V_{transition} \) or \( V_{gs} < V_t - 0.5\cdot V_{transition} \), where \( V_{transition} \) is defined as the voltage range over which the transition from weak inversion to strong inversion is complete.
In this case, it is desirable to have MOSFET device model that unifies the weak inversion and strong inversion seamlessly. A continuous model of MOSFET has been developed by C.C. Enz and etc [28]:

\[
I_{ds} = 2n\mu_C W \frac{kT}{q} \left[ \ln \left( 1 + \exp \left( \frac{V_p - V_S}{2kT/q} \right) \right) \right]^2 - \left( \ln \left( 1 + \exp \left( \frac{V_p - V_D}{2kT/q} \right) \right) \right]^2
\]  \hspace{1cm} (5.24)

Where \( V_p = \frac{(V_g - V_t)}{n} \)

This model is conventionally called EKV model, which is named by the initials of three major contributors. It provides a unified continuous derivable equation for all regions of operation (depletion, strong and weak inversion). The analytical solution of \( g_1 \), \( g_2 \) and \( g_3 \) can be derived using the EKV model. However, they are too complicated to provide any insights on the nonlinear behavior of the MOSFET device. Numerical solution can also be easily achieved if need arises. Since \( g_3 \) changes very fast at \( V_{gs} \) close to \( V_t \), it is not realistic to achieve reliable linearity improvements using this region of operation. For all practical purposes, all the devices need to be biased in either strong inversion region or weak inversion region in derivative superposition method in order to obtain predictable \( g_3 \) value that is weakly \( V_{gs} \) dependent. For these reasons, thorough treatment in the transition region from weak inversion to strong inversion is not given in this work.
5.4.4 Simulation of DC Characteristics of RF NFET

In order to gain an in-depth understanding of the nonlinear behavior of RF NFET, DC characteristics are simulated using Cadence Spectre simulator. A simple cascode test circuit is built using two RF NFETs with $W = 2.5 \, \mu m$ and $L = 120 \, nm$, as shown in Figure 5.9.

![MOSFET DC test circuit for linearity analysis](image)

**Figure 5.9 MOSFET DC test circuit for linearity analysis**

While keeping the gate bias of the cascode device constant at 1.5 V, the gate bias of the $g_m$ device is swept from 0.2 V to 0.8 V. The drain current $I_{ds}$ normalized to unit width is plotted versus the gate bias $V_{gs}$, as shown in Figure 5.10. The regions labeled as WI, TR and SI are weak inversion, transition region and strong inversion, respectively. The curve that runs through all three regions is the simulated result. Drain current curves predicted by weak inversion model and strong inversion model are compared with simulated result.
It can be observed that the analytical model conforms very well to simulated result in both regions.

Figure 5.10: Normalized drain current of RF NFET vs. gate bias voltage (W=2.5 µm L=120nm)

For short channel devices, the transfer characteristic deviates from that predicted by long channel MOSFET model due to short channel effect. As can be seen in equation (5.6), at small overdrive voltage \((V_{gs} - V_{th})\), the transfer function is close to a quadratic curve. When overdrive voltage is large, the transfer function approaches to the shape of a straight line. Thus intuitively, larger overdrive voltage leads to higher linearity.
Using equation (5.9), transconductance $g_1$ (conventionally noted as $g_m$) as a function of $V_{gs}$ of the device can be obtained from the simulated DC I-V curve, which is shown in Figure 5.11. Again, the analytical model conforms very well to simulated result in both regions. It can be observed that the slope of $g_1$ decreases with $V_{gs}$ at high overdrive voltage, which is caused by velocity saturation of carriers. From equation (5.8), $A_{HP3}$ is proportional to the square root of the absolute value of $g_1$. It can be observed that in strong inversion region, higher $V_{gs}$ helps to increase $|g_1|$ and thus increase the $A_{HP3}$. 
Using equation (5.10), $g_2$ as a function of $V_{gs}$ of the device can be obtained from the simulated DC I-V curve, which is shown in Figure 5.12. The $g_2$ curve in weak inversion region is well predicted by the analytical model. In strong inversion region, the analytical model deviates slightly from the simulated result. The overall trend remains well predicted. $g_2$ is related to second order intermodulation product. It can be observed that in strong inversion region, higher $V_{gs}$ helps to reduce $g_2$ which leads to lower $A_{IP2}$. In addition, within the transition region, $g_2$ assumes the highest value, which indicates strong 2$^{nd}$ order nonlinearity.
Using equation (5.11), \( g_3 \) as a function of \( V_{gs} \) of the device can be obtained from the simulated DC I-V curve, which is shown in Figure 5.13. The \( g_3 \) curve in weak inversion region and strong inversion region is well predicted by the analytical model with \( V_{gs} \) relatively far away from transition region. In transition region (and the region nearby), the \( g_3 \) function is distorted due to the transition of two different current mechanism, which is not modeled in this study.

Figure 5.13: Normalized 3\textsuperscript{rd} order power series coefficient of RF NFET drain current
(W=2.5 \( \mu \)m L=120nm)
From equation (5.8), $A_{\text{IIP3}}$ is proportional to the square root of the reciprocal of the absolute value of $g_3$. It can be observed that in strong inversion region, higher $V_{gs}$ helps to reduce $|g_3|$ thus increase the $A_{\text{IIP3}}$. This is consistent with the requirement to obtain higher $|g_1|$ in order to achieve higher gain and lower noise.

There is, however, another operating point where $|g_3|$ approaches zero, which is referred to as optimum bias point in [1]. In this particular test case, the optimum bias point occurs at $V_{gs}=0.456$ V. At this operating point, very high $A_{\text{IIP3}}$ can be obtained at low frequency.

![Figure 5.14: Low frequency $A_{\text{IIP3}}$ derived from device DC characteristics ($W=2.5$ $\mu$m $L=120$nm)](image)

Figure 5.14: Low frequency $A_{\text{IIP3}}$ derived from device DC characteristics ($W=2.5$ $\mu$m $L=120$nm)
Using equation (5.8), $A_{IIP3}$ as a function of $V_{gs}$ of the device can be obtained from calculated $g_1$ and $g_3$, which is shown in Figure 5.14. The $A_{IIP3}$ vs. $V_{gs}$ curve predicted by analytical model well conforms to simulated result in both weak inversion and strong inversion regions. A strong peak is observed at 0.456 V. With increasing $V_{gs}$, the $A_{IIP3}$ increases in strong inversion region while remains roughly constant in weak inversion region.

5.4.5 Design Challenges of Optimum Gate Biasing

As discussed in section 5.1.1, the idea of linearization through optimum gate biasing is to simply bias the $V_{gs}$ at the zero crossing of the $g_3$ curve, which corresponds to the peak of the derived $A_{IIP3}$. However, it can be observed in Figure 5.14 that the high IIP3 achieved this way rolls off very fast when the $V_{gs}$ drifts away slightly from optimum value, which means that the achieved IIP3 through this method is very sensitive to any change in biasing voltage. It is claimed in [1] that noticeable linearity improvement using optimum gate bias requires the biasing voltage accuracy to be within ±10mV, which is very difficult to achieve due to inevitable process and temperature variation.

Further, what make the biasing requirement even more prohibitive is that the optimum biasing voltage for maximum $A_{IIP3}$ is subject to process and temperature variation by itself. Figure 5.15 shows the process variation of $g_3$ and derived $A_{IIP3}$. Since $g_3$ is proportional to the 3rd order derivative of the drain current, a small change in the shape of
the \( I_{ds} \) curve can cause a rather large shift of the \( g_3 \) curve. The zero crossings of the \( g_3 \) curves determine the optimum gate bias voltages, which correspond to the peaks in \( A_{IIP3} \) curves. Figure 5.16 shows the Monte Carlo simulation results (100 run) of optimum \( V_{gs} \). It can be observed that the 3 sigma variation of optimum \( V_{gs} \) is \( \pm 55 \) mV, which is rather large compared to the required biasing voltage accuracy of within \( \pm 10\) mV.

![Figure 5.15: Process variation of \( g_3 \) and derived \( A_{IIP3} \) (\( W=2.5 \) \( \mu \)m \( L=120\)nm)](image)

![Figure 5.16: Histogram of optimum \( V_{gs} \) over process variation (\( W=2.5 \) \( \mu \)m \( L=120\)nm)](image)
The variation of $g_3$ and derived $A_{IIIP3}$ over temperature is simulated from 0° to 80°. The result is shown in Figure 5.17. As the environment temperature increases, since the threshold voltage $V_t$ is reduced, the weak inversion to strong inversion transition happened at lower voltage, which causes the $g_3$ and derived $A_{IIIP3}$ curve to shift left. Therefore, in addition to ±55 mV 3 sigma process variation of optimum $V_{gs}$, ±30 mV temperature variation is also observed. Thus the effort to achieve linearity improvement through optimum gate biasing would remain unpractical unless an automatic biasing circuit is available to track the optimum gate bias voltage over process and temperature variations.

Figure 5.17: Temperature variation of $g_3$ and derived $A_{IIIP3}$ (W=2.5 µm L=120nm)

Another fundamental drawback of optimum gate biasing method is that it is inherently a low gain design. Since the zero crossing of the $g_3$ curve occurs at the junction of weak inversion and strong inversion region, the transconductance (normalize to device width) of the MOSFET is very small. As can be seen in Figure 5.11, the transconductance $g_1$ at
optimum gate biasing voltage 0.456V is at the magnitude of 245 $\mu$A/V·µm, compared with 634 $\mu$A/V·µm at $V_{gs} = 0.8$V.

5.4.6 Design Challenges of Derivative Superposition Method

Given the difficulties in implementation of optimum gate method, the derivative superposition method discussed in section 5.1.3 seems to be a rationale choice for wideband linearity improvement. The idea of derivative superposition is to use two devices in parallel, one main MOSFET and one auxiliary MOSFET for $g_3$ cancellation.

As shown schematically in Figure 5.18, the main MOSFET is biased at strong inversion region to obtain a high transconductance and a small negative $g_3$ at the same time. The auxiliary MOSFET is biased at weak inversion region to obtain a positive $g_3$. Since the
two MOSFET are placed in parallel, the $g_3$ curves of the two MOSFETs are superposed to cancel each other around the optimum biasing point. The simulated $g_3$ curves are shown in Figure 5.19. Since the normalized $g_3$ at weak inversion has a rather large positive value, a relatively small device can be used for auxiliary MOSFET. In this case, the width of auxiliary MOSFET is 16% of the main MOSFET and biased at 0.3V lower than the gate voltage of main MOSFET.

![Figure 5.19: Superposition of $g_3$ curves (W=2.5 µm L=120nm)](image)

Since the obtained $g_3$ curve of the composite MOSFET is flatter near the zero crossing than in optimum gate biasing case, the biasing requirement is relatively relaxed. Significant linearity improvement can be achieved within ±30 mV from the optimum
biasing voltage. Most importantly, the main MOSFET has rather large transconductance, which can provide required gain with decent noise performance. These advantages make this method much more practical in many cases.

Several problems, however, still exist with derivative superposition method. First of all, though bias requirement is relaxed, given the process, voltage and temperature variation, it is still quite challenging to bias the circuit such that the $g_3$ of two MOSFETs cancel the majority part and leave only a minimal residual. Further, in many applications linearity at large signal level is important. Thus even if optimum bias is available, the large swing of the signal can drive the $V_{gs}$ out of the effective range of linearization, which can severely degrades the effectiveness of this linearity improvement method.

### 5.5 Proposed Extended Effective Range Derivative Superposition Method

The major challenge faced by the linearity improvement method discussed thus far is how to obtain a larger $V_{gs}$ range of effective linearity improvement, or in another word, how to obtain a larger $V_{gs}$ range in which $g_3$ assumes very small value. For the method discussed above, since the $g_3$ curve in strong inversion is rather flat, the effective linearity improvement range is mainly limited by the width of the peak of the $g_3$ curve. For a given device, the shape of $g_3$ curve is fixed. The question that follows is: what else can be done at the circuit design level to create a plateau of $g_3$ curve for auxiliary device?
5.5.1 Derivative Superposition with Multiple $V_{gs}$ Offsets

One obvious idea is to use the even lower $V_{gs}$ to obtain a flatter $g_3$ curve. This, however, will lead to quite large auxiliary MOSFET (comparable to main MOSFET) in order to effectively cancel the $g_3$ of the main device, which results in larger input capacitance and higher noise.

Another method is to bias multiple auxiliary MOSFETs at different $V_{gs}$. As shown in Figure 5.20, the bias voltage of auxiliary MOSFETs are skewed slightly such that the superposition of the $g_3$ curves of all auxiliary MOSFETs can form a shape that approximates the negative image of the main MOSFET $g_3$ curve near the optimum bias point. With the superposition of all devices, the overall $g_3$ curve can assume a very small value over an extended $V_{gs}$ range.

![Figure 5.20: Derivative superposition using multiple auxiliary MOSFETs](image)
The simulated individual and superposed $g_3$ curves are shown in Figure 5.21. In this experiment, the $V_{gs}$ of the auxiliary MOSFETs are offset by 0.275V, 0.3V, and 0.35V. The optimum device width ratio is 1:0.16:0.4:0.2 in order to maximize the $V_{gs}$ range of effective linearity improvement. In this particular example, the effective linearity improvement region is increased to 140 mV, compared with 60 mV with typical derivative superposition method.

![Figure 5.21: Normalized $g_3$ curves of MOSFETs with different $V_{gs}$ offset (W=2.5 µm)](image)

There are, however, still some issues associated with this technique. First, multiple auxiliary MOSFETs mean larger input capacitance and higher noise. Second, since the $g_3$ peak roll off rapidly and goes to negative value as $V_{gs}$ increases, the negative part of the
g3 curve of the auxiliary MOSFET with smallest Vgs offset (0.275V in this case) adds to that of the main MOSFET and reduce the effectiveness of this technique, which sets the upper limit of the achievable Vgs range of linearity improvement. Finally, it is a rather challenging task to generate multiple bias voltages that match each other and follow the temperature dependence of the optimum biasing of different MOSFETs.

5.5.2 The Effect of Gate Length on g3 Curve

With the issues discussed above, a nature question that follows is whether there is another freedom in circuit design level to alter the shape of g3 curve. Fortunately, such a freedom does exist, which is the gate length of the MOSFET.

As the device dimensions of the CMOS process scales progressively, the threshold voltage Vt generally reduces accordingly. This is because the edge effect manifests itself more as the channel length shrinks, which leads to reduced depletion charge underneath the gate. Since the threshold voltage Vt is determined by:

\[ V_t = V_{FB} + 2\Phi_F - \frac{Q_d}{C_{ox}} \]  

(5.25)

In NFET case, Qd is negative. Thus a lower depletion charge reduces the threshold voltage Vt.

Though the trend of reducing Vt with shrinking channel length generally holds across the processes, it is not necessarily valid in a given process. In order to reduce the DIBL
(Drain Induced Barrier Lowering) which causes the variation of threshold voltage with
drain voltage, “halo” is commonly used in modern processes. “halo” represents the
substrate region near the source and drain with increased substrate doping concentration.
As the channel length shrinks, the effect of “halo” manifests itself more and essentially
increases the average doping concentration of the channel, which in turn increases the
threshold voltage of the device. Since this trend is to the opposite of the short channel
effect, this phenomenon is called reverse short channel effect [29]. The dependence of the
threshold voltage on channel length is shown in Figure 5.22. The threshold voltage drops
rapidly as the gate length increases from 120 nm to 1 μm, while keeps relatively constant
for gate length larger than 1 μm due to the vanishing weight of “halo” region in the
overall channel area.

![Graph showing the effect of channel length on threshold voltage.](image)

**Figure 5.22:** The effect of channel length on threshold voltage.
With a focus in the gate length below 1 µm, the range in which most RF circuit design falls in, the effect of gate length on the $g_3$ curve is studied. Figure 5.23 shows the simulated $g_3$ curves of MOSFETs with gate length of 120 nm, 240 nm, 360 nm and 480 nm, respectively. Clearly, the shape of $g_3$ curve changes significantly with gate length. As the gate length increases, the threshold voltage decreases due to the reverse short channel effect, which makes the $g_3$ curve shift to the left. The lower magnitude of $g_3$ curve with larger gate length is due to the reduced W/L ratio. Since W/L is a scalable variable, the magnitude of $g_3$ curve can be scaled according to the required $g_3$ cancellation.

Figure 5.23: Normalized $g_3$ curves of MOSFETs with different length (W=2.5 µm)
5.5.3 Extended Effective Range Derivative Superposition Method

With the gate length proved to be an extra freedom in controlling the shape of $g_3$, novel method of derivative superposition for linearity improvement can be designed. As shown in Figure 5.24, instead of using several auxiliary MOSFETs with different bias, MOSFETs with different gate lengths are used to create desired $g_3$ curves with the same bias voltage. The gate length of the main MOSFET M1 is 120 nm. The gate lengths of the auxiliary MOSFETs M3, M4, and M5 are 120 nm, 240 nm, and 360 nm, respectively. The simulated individual and superposed $g_3$ curves of M1 and M3-M5 are shown in Figure 5.25. It can be observed that the $g_3$ curves of devices with different gate length peak at different $V_{gs}$ such that superposed $g_3$ curve goes in proximity to the X axis over an extended range from 0.62 V to 0.76 V. Therefore, this method increases the effective linearity improvement region, while avoiding the difficulties of providing multiple bias voltages for auxiliary MOSFETs and associated performance degradation as well.

![Figure 5.24: Proposed Extended Effective Range Derivative Superposition method](image)

Figure 5.24: Proposed Extended Effective Range Derivative Superposition method
Figure 5.25: Normalized $g_3$ curves of MOSFETs with EERDS Method ($W=2.5 \, \mu m$)

Figure 5.26: Calculated $A_{III}$ curves of MOSFETs with EERDS Method ($W=2.5 \, \mu m$)
The calculated $A_{\text{IIP3}}$ curves based on the simulated data using equation (5.8) are shown in Figure 5.26. It can be observed that the $A_{\text{IIP3}}$ curve of extended effective range DS method has a higher magnitude and a wider range than that of conventional DS method and optimum gate biasing method. On all three curves, a lower peak exists at $V_{\text{gs}}$ where the $g_3$ curve of main device cross zero. Its magnitude appears to be very small since the $A_{\text{IIP3}}$ peak that operating point is so narrow compared to the resolution of the simulation such that the optimum point is “missed” by the measurement point. It is obvious that the area enclosed by the $A_{\text{IIP3}}$ curve of EERDS method is much larger than that of typical DS method and optimum bias method. The above observations prove the effectiveness of the extended effective range DS method at DC or low frequency range.

### 5.6 Linearity Modeling and Analysis at High Frequency

At high frequency, however, the capacitances and inductances start to show their effects, which make the assumption of memoryless system invalid. Thus the model derived in section 5.4 can not be directly applied at RF frequency. In typical source degenerated common source LNA with DS method as shown in Figure 5.3, it is proved that the IIP3 of the LNA is a function of $g_1, g_2, \text{degenerate inductance } L, \text{gate source capacitance } C_{\text{gs}}$ and frequency $\omega$ according to equation (5.1) and (5.2) [3]. Thus it is necessary to reexamine the EERDS method at high frequency.
5.6.1 High Frequency Small Signal Analysis of EERDS Method

In order to gain insights of the mechanism through which the feedback degrades the effectiveness of linearization, a formal small signal analysis is performed. The small signal model of a nonlinear amplifier with source inductor is shown in Figure 5.27. In this model, a source impedance of $Z_s$ is assumed (not necessarily conjugate matched) and the drain-gate capacitance is neglected. The nonlinearity of the device is modeled using Volterra series. For amplifier with composite device, $g_1$, $g_2$ and $g_3$ here represent the sum of that of the main and auxiliary devices. A source inductor is added to study the feedback effect of 2$^{nd}$ order intermodulation current on the IIP$_3$ of the amplifier.

![Small signal model of nonlinear amplifier with source inductor](image)

Figure 5.27: Small signal model of nonlinear amplifier with source inductor

First, neglect the higher order terms and solve for first order operating point at arbitrary frequency $\omega$. The $V_{gs}$ can be expressed in term of $V_g$:
\[ v_{gs}(\omega) = v_g(\omega) \frac{1}{j\omega C_{gs}} \frac{\frac{1}{j\omega L_s} + \frac{1}{j\omega C_{gs}} + \omega_s L_s}{} \] (5.26)

Given the first order term of \( V_{gs} \), all the current terms with different order in the Volterra series that result from the first order term of the voltage signal at the gate can be obtained according to the relationships shown in Figure 5.27. Apply a two tone excitation with frequency \( \omega_1 \) and \( \omega_2 \) with a small frequency spacing \( \Delta\omega \), the following current component at different frequencies can be obtained.

\[
i_1(\omega_1) = i_1(\omega_2) = g_{1v_{gs}}(\omega) \] (5.27)

\[
i_2(2\omega_1) = i_2(2\omega_2) = \frac{1}{2} g_{2v_{gs}}^2(\omega) \] (5.28)

\[
i_2(\omega_2 - \omega_1) = i_2(\omega_2 + \omega_1) = g_{2v_{gs}}^2(\omega) \] (5.29)

\[
i_3(3\omega_1) = i_3(3\omega_2) = \frac{1}{4} g_{3v_{gs}}^3(\omega) \] (5.30)

\[
i_3(2\omega_2 - \omega_1) = i_3(2\omega_1 - \omega_2) = \frac{3}{4} g_{3v_{gs}}^3(\omega) \] (5.31)

Now, the feedback effect of the 2\(^{nd}\) order current component is investigated. Considering how the signals are mixed together by 2\(^{nd}\) order nonlinearity, there are three different ways that the 2\(^{nd}\) order component mixes with fundamental to contribute to the 3\(^{rd}\) order intermodulation product. The mechanism and the expression of the 3\(^{rd}\) order intermodulation current generated through these two ways are shown below:
**Mechanism A:** $2\omega_2 \otimes \omega_1 \Rightarrow 2\omega_2 - \omega_1$

In this mechanism, the $2\omega_2$ component of the drain current generates a voltage at $v_s$, which then mix with the $\omega_1$ tone of the input signal though 2nd order nonlinearity and creates $2\omega_2 - \omega_1$ component of the drain current. This current component can be expressed as:

$$i_{3A}(2\omega_2 - \omega_1) = -\frac{L_s}{g_2^2 v_{gs}^3(\omega)} \frac{C_{gs}}{j2\omega_s + \frac{1}{j2\omega C_{gs}} + Z_s} \quad (5.32)$$

**Mechanism B:** $\omega_2 - \omega_1 \otimes \omega_2 \Rightarrow 2\omega_2 - \omega_1$

In this mechanism, the $\omega_2 - \omega_1$ component of the drain current generates a voltage at $v_s$, which then mix with the $\omega_2$ tone of the input signal though 2nd order nonlinearity and creates $2\omega_2 - \omega_1$ component of the drain current. This current component can be expressed as:

$$i_{3c}(2\omega_2 - \omega_1) = -g_2^2 v_{gs}^3(\omega) \frac{L_s}{C_{gs}} \frac{1}{j\Delta \omega L_s + \frac{1}{j\Delta \omega C_{gs}} + Z_s} \quad (5.33)$$

The total 3rd order intermodulation current at frequency $2\omega_2 - \omega_1$ thus can be found by summing up equation (5.31) through (5.33):
\[i_3(2\omega_2 - \omega_1) = v_{gs}^3(\omega) \left[ \frac{3}{4} g_3 - g_2^2 \left( \frac{1}{2} \frac{L_s}{C_{gs}} + \frac{L_s}{j2\omega L_s + \frac{1}{j2\omega C_{gs}} + Z_s} \right) \right] \] (5.34)

At 3rd order intermodulation intercept point, \(i_3(2\omega_2 - \omega_1) = i_1(\omega)\). Combine equation (5.26), (5.27) and (5.34), the expression of \(A_{\text{IIP}_{3}}\) can be obtained.

\[
A_{\text{IIP}_{3}} = \left| -g_1 \left( j\omega L_s + \frac{1}{j\omega C_{gs}} + \omega_f L_s \right) \right|^2 \omega^2 C_{gs}^2 \frac{3}{4} g_3 - g_2^2 \left( \frac{1}{2} \frac{L_s}{C_{gs}} + \frac{L_s}{j2\omega L_s + \frac{1}{j2\omega C_{gs}} + Z_s} \right) \right| \] (5.35)

For 2 tones with small spacing (\(\Delta\omega \ll \omega\)), \(j\Delta\omega L_s + \frac{1}{j\Delta\omega C_{gs}} + Z_s\) approaches infinity.

Thus equation (5.35) can be reduced to

\[
A_{\text{IIP}_{3}} = \left| -g_1 \left( j\omega L_s + \frac{1}{j\omega C_{gs}} + \omega_f L_s \right) \right|^2 \omega^2 C_{gs}^2 \frac{3}{4} g_3 - g_2^2 \left( \frac{1}{2} \frac{L_s}{C_{gs}} \right) \right| \] (5.36)

It can be observed that when source inductor is added, an extra term of \(g_2\) appears in the denominator. Thus simply cancel out the \(g_3\) term does not necessarily result in the
optimization of the $A_{\text{IP3}}$. Since $g_2$ term is a vector whose phase changes with frequency, it is difficult to have $g_3$ term and the $g_2$ term cancel out each other over a wide frequency range. A smaller $g_2$ magnitude, however, always helps to alleviate this problem and improve the $A_{\text{IP3}}$.

### 5.6.2 Linearization through 3rd Order Harmonic Current Cancellation

In order to determine the effectiveness of the EERDS linearization method at high frequency, 2 tone test is applied at 4 GHz on the circuit shown in Figure 5.28.

![Figure 5.28: 3rd order harmonic current cancellation test (no source inductor)](image-url)

Figure 5.28: 3rd order harmonic current cancellation test (no source inductor)
According to the EERDS method discussed in section 5.5, the main device is biased at higher $V_{gs}$ ($V_{b\_main}$) to provide the $g_m$ for the amplifier and the aux1 and aux2 are biased at lower $V_{gs}$ ($V_{b\_aux}$) to cancel the $g_3$ of the main device over an extended range of $V_{gs}$. A cascode stage is used to isolate the input and the output. Shunt peaking is used at the output to provide a relatively constant load impedance over the frequency range of interest. Differential topology is used to be consistent with latter experiment.

The 3rd order intermodulation current of the main device and 2 auxiliary devices (with $L_{aux1}=120$ nm and $L_{aux1}=240$ nm) are measured with respect to the gate bias voltage of main devices with the auxiliary device biased with a fixed voltage offset from the $V_{b\_main}$. The magnitude and phase of the 3 currents and the sum are shown in Figure 5.29 and Figure 5.30, respectively. It can be observed that the 3rd order intermodulation current of device aux1 and device aux2 form two peaks with opposite phase to that of the main device. By tuning the size of device aux1 and device aux2 the major part of the 3rd order intermodulation of the main device can be cancelled out over an extended range of $V_{b\_main}$ with minimal residue, which greatly improves the linearity of the amplifier. This is further proved by the IP3 simulation of the same circuit with result shown in Figure 5.31. Within the designed effective range of $V_{b\_main}$ from 0.64 V to 0.78 V, an improvement of 10 to 15 dB is observed in both IIP3 and OIP3.

If we observe Figure 5.29 closely, a strong correspondence can be found with respect to Figure 5.21, which describes the superposition of $g_3$ curves in EERDS method. This is an
expected phenomenon since the 3\textsuperscript{rd} order intermodulation current is proportional to $g_3$ for a given input signal if the feedback at high frequency is neglected as indicated by equation (5.8). Thus Figure 5.29 is essentially a high frequency parallel of Figure 5.21.

Figure 5.29: 3\textsuperscript{rd} order intermodulation current cancellation using EERDS method - magnitude (no source inductor)
There is, however, a noticeable difference that worth to discuss here. In Figure 5.29, the 3rd order intermodulation current “bounce” back instead of crossing the X axis as g3 curve does in Figure 5.21. This is because the 3rd order intermodulation current is plotted as magnitude which has non-negative value. However, it can be observed that the phase has a sharp transition of 180° at the gate bias voltage where the 3rd order intermodulation current “bounces” back, which means that the current actually becomes negative at that point. Thus if we flip the section of 3rd order intermodulation curve that has opposite phase, the 3rd order intermodulation current plot well resembles the g3 curves in Figure 5.21. As a result, the EERDS method developed at low frequency can be applied to the
common source amplifier without source inductor with minor degradation at high frequency.

![Graph](image)

Figure 5.31: Simulated IP3 at high frequency with EERDS method (no source inductor)

### 5.7 Modified EERDS Method for Amplifier with Source Inductor

In front-end RF circuit, the source inductor is frequently used for different reasons. For LNA that works in GHz range, the source degeneration inductor typically has a value around 0.5 nH. In order to manifest the effect, a much larger source inductance of 3 nH is added to the source of the same composite device and repeat the test. The circuit is shown in Figure 5.32. In the whole comparison scheme differential circuit is used. This is because the source inductor would make a noticeable difference in term of voltage gain in
single-ended amplifier case, which makes the comparison uncontrolled. The differential circuit does not make the linearity degradation effect much different from that of the single-ended circuit, because the 2\textsuperscript{nd} order intermodulation current which causes the degeneration of the linearization is in phase in both sides of the differential circuit and adds to each other instead of cancel out as the fundamental and 3\textsuperscript{rd} order intermodulation current do. Thus the feedback of 2\textsuperscript{nd} order intermodulation currents of differential circuit is same as that of the single-ended one. The only difference that worth to mention is the simulated IIP\textsubscript{3} of the differential circuit needs to be reduced by 6 dB before compared with that of single-ended circuit since the actual signal level across the $V_{gs}$ of each $g_m$ transistor is only half of the input signal.

It can be observed in Figure 5.33 that the total 3\textsuperscript{rd} order intermodulation current increases within the effective bias voltage range and the width of the effective bias voltage range shrinks as well. This degradation is mainly due to the feedback of 2\textsuperscript{nd} order intermodulation current to the source which mixes with the fundamental of the signal through second order nonlinearity of the devices and generates 3\textsuperscript{rd} order intermodulation current.
Comparing Figure 5.33 and Figure 5.29, one major difference is that the 3rd order intermodulation current of aux2 in Figure 5.33 decreases at $V_{b\_main}$ lower than 0.71 V and increases at $V_{b\_main}$ higher than 0.71 V. Considering the phase change of 180° of the 3rd order intermodulation current of aux2 at $V_{b\_main}$ of 0.71 V, the net increase of the 3rd order intermodulation current of aux2 is in phase with, and thus adds to, that of the main device, which results in an net increase of total 3rd order intermodulation current and the degradation of the EERDS linearization. Further, the 3rd order intermodulation current of aux1 is reduced, which leads to less compensation of the nonlinearity of the main device. However, the change in 3rd order intermodulation current of aux1 is smaller compared to that of the aux2. In addition, part of the reason of the slight decrease of 3rd order
intermodulation current of aux1 is that a small portion of the increased total 3\textsuperscript{rd} order intermodulation current back flow into the drain of aux1, which effectively reduce the net 3\textsuperscript{rd} order intermodulation current. The over all effect can be observed in Figure 5.35: both IIP\textsubscript{3} and OIP\textsubscript{3} peaks narrow down which indicates a reduced effective range of gate bias.

Figure 5.33: 3\textsuperscript{rd} order intermodulation current cancellation using EERDS method – magnitude (with 3nH source inductor)
Figure 5.34: 3\textsuperscript{rd} order intermodulation current cancellation using EERDS method -phase (with 3nH source inductor)

Figure 5.35: Simulated IP\textsubscript{3} at high frequency with EERDS method (with 3nH source inductor)
With the above observation, the question that follows is: why the 3rd order intermodulation current of aux2 at $V_{b_{main}}$ close to 0.71 V increases more significantly than that of the main device and aux2? If we compare Figure 5.12 and Figure 5.13, it can be observed that the $g_2$ curve reaches its peak at exactly the same $V_{gs}$ where the $g_3$ curve crosses X axis, which can easily be proved analytically. Since the degeneration of the linearity is caused by frequency mixing associated with $g_2$, the 3rd order intermodulation current of a given device is most susceptible to this effect near the zero crossing of $g_3$ curve, or equivalently, the “bounce” back point of 3rd order intermodulation current. Precisely speaking, the 3rd order intermodulation current of all three devices are equivalently affected by the source inductor, but only for aux2, the gate bias voltage corresponds to the most noticeable increase of 3rd order intermodulation current (i.e. the $V_{gs}$ for maximum $g_2$) falls in the effective range of EERDS method and becomes noticeable, while similar effects on the main device and aux1 can be neglected since they mainly occur outside the gate bias range of interest.

Given that the $g_2$ of the aux2 is the major cause of the linearity degradation at high frequency for amplifier with source inductor, any approaches that try to maximize the effective range of $V_{gs}$ has to address this problem. In this work, a modified EERDS method is proposed. As shown in Figure 5.36, instead of connecting the source of all main and auxiliary devices to the same node X, the source of the aux2 is connected to the ground directly. As a result, the major feedback path of 2nd order intermodulation current
is removed. Since the aux2 contributes negligible $g_m$ compared to the main device, the signal balance at the output is not compromised.

![Diagram](image)

**Figure 5.36:** 3rd order harmonic current cancellation using modified EERDS method (with source inductor)

With the implementation of modified EERDS method, the simulated magnitude and phase of the several 3rd order intermodulation are shown in Figure 5.33 and Figure 5.34, respectively. Since no source inductor is present at the source of aux2, the feedback of $g_2$ component to $V_{gs}$ is greatly suppressed. As a result, the 3rd order intermodulation current of aux2 in at $V_{b_{aux}}$ near 0.71 V increases less significantly. Minimal reduction of the 3rd order intermodulation current of aux1 is observed. The 3rd order nonlinearity of the main device is effectively compensated over an extended effective range. This conclusion is further verified by the IP3 simulation of the same circuit with result shown in Figure 5.39.
Within the designed effective range of $V_{b_{\text{main}}}$ from 0.66 V to 0.77 V, an improvement of 10 to 19 dB is observed in both IIP3 and OIP3. In this case, the device size is not changed for all 3 test cases for a fair comparison. For amplifier with inductive source degeneration, achieved linearity and effective range can be further improved with the fine tuning of the device size.

Figure 5.37: 3rd order intermodulation current cancellation using modified EERDS method –magnitude (with 3nH source inductor)
Figure 5.38: 3rd order intermodulation current cancellation using modified EERDS method -phase (with 3nH source inductor)

Figure 5.39: Simulated IP3 at high frequency with modified EERDS method (with 3nH source inductor)
5.8 Conclusions

At DC, the linearity model of the MOSFET device in both strong inversion and weak inversion region is developed from the short channel device model and subthreshold current model. Simulation results prove that the model well predicts the power series coefficients and the $A_{IIIP3}$ of the MOSFET device. Based on the DC linearity model, Extended Effective Range Derivative Superposition (EERDS) method is developed. Compared with conventional Superposition method, EERDS method increases the effective range of gate bias in which the derivative superposition linearization remains effective.

At high frequency, the feedback of 2nd order intermodulation current mixes with fundamental through 2nd order nonlinearity of the device. A formal analysis of such mechanism in a common source amplifier with inductive source degeneration is carried out using small signal model which taken into account the 2nd and 3rd nonlinearity. An analytical expression of $A_{IIIP3}$ is derived as a function of frequency and source inductance. The source inductance is proved to have a major impact on linearization at high frequency. The mechanism of this effect is identified and discussed in detail. A modified EERDS method is proposed to mitigate the effect of source inductor, which is proved to be effective by the simulation results.
CHAPTER 6

HIGH LINEARITY 3-10 GHZ CMOS UWB LNA WITH ACTIVE BALUN

6.1 Introduction

In CHAPTER 4, the importance of high linearity UWB LNA in UWB system has already been elaborated. Using the Extended Effective Range Derivative Superposition Method (EERDS Method) developed in CHAPTER 5, a 3-10 GHz high linearity UWB LNA with active balun is designed and fabricated in IBM 0.13 μm CMOS process with 1.6 V supply voltage. The circuit includes controllable on-chip bias circuit and a buffer stage to drive the signal off-chip. Both input and output ports are matched to 50 Ω. The chip is packaged using standard 12×12 mm QFN package. Post layout simulation is performed with the layout parasitics being taken into account.
6.2 Circuit Design and Implementation

6.2.1 Constant $g_m$ Bias Circuit

In order to improve the robustness of the design to temperature and supply voltage variations, constant $g_m$ bias circuit was designed. The idea is to provide a bias current that can sustain a relatively constant transconductance of the core device of the circuit, such that the variation of gain (and preferably other performances as well) can be minimized.

![Constant $g_m$ bias circuit](image)

Figure 6.1: Constant $g_m$ bias circuit

Figure 6.1 shows the circuit diagram of the constant $g_m$ bias circuit. In this circuit, the W/L ratio of M1 is 4 times that of M2. The cascode current mirror on the top forces the drain current of both device to be equal. If long channel model is assumed, it can be
easily proved that the $g_m$ of the $M_2$ equals to $1/R_3$. In short channel case, since the drain current deviates from what is predicted by long channel model, the obtained $g_m$ is slightly smaller than $1/R_3$ but still remain relatively constant with temperature and voltage variations. The current thus obtained can be amplified as required using current mirror and serve as bias current of the core circuits.

### 6.2.2 High Linearity 3-10 GHz CMOS UWB LNA Core Circuit

Based on the wideband linearization techniques described in CHAPTER 5, a high linearity UWB LNA was designed with a power supply of 1.6 V. The parasitics of the package, bonding wire and the pads are modeled and their effects are taken into account.

![Figure 6.2: High linearity 3-10 GHz CMOS UWB LNA Core Circuit](image)
The LNA circuit is shown in Figure 6.2. The designed LNA employs common source cascode architecture with inductive source degeneration. A 0.386 nH source inductor is used to generate a real impedance of 50 Ω at the input without the thermal noise of a real resistor. Third order Chebychev bandpass filter is used for wideband impedance match at the input. Cascode stage is used to mitigate the miller effect and reduce the reverse isolation at the same time. A 54 Ω resistor $R_L$ is placed in series with the load inductor $L_L$ in order to reduce the quality factor and thus increase the bandwidth. The drain current is 7.7 mA and the total power consumption is 12.3 mW.

### 6.2.3 Active Balun Circuit

A two stage active balun is used immediately after the LNA to perform the single-ended to differential conversion. The first stage is a source degenerated common source amplifier. In this topology, the signals at drain and source of $M_1$ have a phase difference of roughly 180° and a magnitude ratio determined by $g_m$ of $M_1$ and source/drain impedance. By carefully choosing the value of $L_1$, $L_2$, and $M_1$, a balanced differential signal can be obtained. This topology presents a relatively small input capacitance which helps to improve the gain and noise figure of the LNA. However, it has limited bandwidth beyond which the signal balance degrades. In order to obtain good signal balance in a wide frequency range from 3 GHz to 10 GHz, a second stage is used to condition the signal. This stage use differential common source cascode topology. Common mode inductive degeneration inductor $L_2$ is used to reduce the common mode
gain and improved the signal balance. The common mode rejection of the circuit is proportional to the impedance seen at node X, which reaches its maximum when the inductance \( L_2 \) resonant with the parasitic capacitance seen at node X. In order to maximize the effective bandwidth of the active balance, this resonance frequency needs to be tuned such that it offsets with that of the first stage.

Figure 6.3: High linearity 3-10 GHz CMOS active balun circuit

The input of the first stage of the balun is DC coupled to the output of the LNA. Thus the gate bias is provided by the output of the LNA. The gate of the cascode device is diode biased and connected to the AC ground. The bias voltages \( V_{b1} \) and \( V_{b2} \) are provided from the bias circuit. According to the EERDS method described in section 5.5, the main
device $M_3$ and $M_4$ are biased in strong inversion region with gate voltage $V_{b1}$, while the auxiliary device $M_{2a}, M_{2b}, M_{5a}$, and $M_{5b}$ are biased with a lower voltage $V_{b2}$ which makes them operate at weak inversion region. The size of the main device $M_3$ and $M_4$ is $40\mu m/120nm$. The auxiliary device $M_{2a}, M_{2b}, M_{5a}$, and $M_{5b}$ are sized to $20\mu m/120nm$, $25\mu m/240nm$, $20\mu m/120nm$, and $25\mu m/240nm$, respectively.

### 6.2.4 Output Buffer Circuit

In order for the output signal to be measured, an output buffer is necessary to drive the signal off-chip. This output buffer needs to be able to drive a $50 \, \Omega$ load with approximately unity gain over the frequency range of interest (3-10 GHz).

![High linearity 3-10 GHz CMOS output buffer circuit](image)

Figure 6.4: High linearity 3-10 GHz CMOS output buffer circuit
A fully differential push pull amplifier is used in this design. The schematic is shown in Figure 6.4. M₅ and M₆ are source follower stage to drive low impedance. The gate voltage is simply biased at $V_{DD}$. The M₁, M₂, M₃, and M₄ are configured as common source stage featuring conventional DS method in order to provide a relatively higher linearity. Both input and output are AC coupled and the bias $V_{b1}$ and $V_{b2}$ are provided by the bias circuit.
6.3 Results and Discussion

The designed High linearity 3-10 GHz CMOS LNA was laid out and extracted with the parasitic resistance, capacitance and inductance taken into account. The complete layout of the LNA, active balun and the output buffer is shown in Figure 6.5. The chip area of the layout is 1.04 mm$^2$. Using Cadence SpectreRF simulator, the extracted view of the design was simulated and evaluated.

Figure 6.5: Layout of High linearity 3-10 GHz CMOS LNA with active balun
6.3.1 Impedance Matching

The simulation results of input matching of high linearity 3-10 GHz CMOS LNA are shown in Figure 6.6. The simulated $S_{11}$ is lower than -10 dB over the most part of the whole designed frequency range (3-10 GHz), and slightly higher than -10 dB at the very low end of the frequency range. This excellent input matching guarantees maximal power can be delivered to the input port of the LNA, which in turn helps with gain and noise performance. Since the EERDS method was applied in the balun stage, input matching of the LNA is not affected.

Figure 6.6: High linearity 3-10 GHz CMOS LNA input matching
The simulation results of output matching of high linearity 3-10 GHz CMOS LNA are shown in Figure 6.7. The simulated $S_{22}$ is lower than -10 dB over the whole designed frequency range (3-10 GHz). This excellent output matching guarantees maximal power can be delivered to the load of the LNA, which improves gain and linearity.

![Graph showing S22 dB vs Frequency (GHz)](image)

Figure 6.7: High linearity 3-10 GHz CMOS LNA output matching

### 6.3.2 Gain Response

Figure 6.8 shows the simulation results of the gain response of high linearity 3-10 GHz CMOS LNA. The simulated $S_{21}$ remains flat at the level of 12.5 dB from 3 to 9 GHz and starts to roll off at 9 GHz. The gain drops to about 7 dB at 10 GHz. Higher gain of the first stage helps minimize the noise figure of the LNA. In this design, LNA gain is moderate.
for several reasons. First, the designed LNA needs to drive a 50 Ω load for testing purpose, which limits the maximum gain that can be achieved. In real application where the LNA drives on-chip circuits, the gain bandwidth of the LNA can be significantly boosted. Second, application of EERDS method increases the load seen by the first stage of the LNA, which further reduces the LNA gain.

The two curves in Figure 6.8 represent the $S_{21}$ of LNA with and without EERDS method, respectively. Since the EERDS method mainly impacts the 3\textsuperscript{rd} order intermodulation products while impose minimal effect on the fundamental, negligible effect is observed on the gain response of the LNA.

![Figure 6.8: High linearity 3-10 GHz CMOS LNA gain](image-url)
In order to evaluate the effectiveness of the single-ended to differential conversion, the signal imbalance at the output was simulated. Figure 6.9 and Figure 6.10 show the gain and phase imbalance of the high linearity 3-10 GHz CMOS LNA respectively. It can be observed that the gain imbalance is less than 0.2 dB and the phase imbalance is less than 4 degree over the whole frequency range. The application of EERDS method has minimal effect on the signal imbalance.

Figure 6.9: High linearity 3-10 GHz CMOS LNA gain imbalance
6.3.3 Noise Figure

Figure 6.11 shows the simulation results of the noise figure of high linearity 3-10 GHz CMOS LNA. The simulated noise figure remains flat at the level of 5 dB from 3 to 9 GHz and starts to increase rapidly at 9 GHz. The noise figure reaches about 7 dB at 10 GHz. In addition to the lower gain due to the reasons described in section 6.3.2, the LNA has to use a noisy buffer to drive the signal off-chip before it is further amplified. Thus, the noise figure is higher than minimum achievable value in this design. In real application where the LNA drives on-chip circuits, the noise figure of the LNA can be significantly improved.
Figure 6.11: High linearity 3-10 GHz CMOS LNA noise figure

The two curves in Figure 6.11 represent the noise figure of LNA with and without EERDS method, respectively. Since the EERDS method is applied in the balun stage, where the signal has already been amplified by the first stage of the LNA. As a result, when referred to the input of the LNA, the noise contribution of the composite nfets is significantly reduced. Thus negligible effect of EERDS method is observed on the noise figure of the LNA.

6.3.4 Linearity

The simulation results of the input referred 3\textsuperscript{rd} order interception point (IIP\textsubscript{3}) of high linearity 3-10 GHz CMOS LNA are shown in Figure 6.12. The simulated IIP\textsubscript{3} generally
increases with the frequency within the design frequency range. This is because the LNA quality factor decreases with increasing frequency, which reduces the signal level across the gate and source node of the input stage of the LNA. As a result, higher power level is required at the input port in order for the linearity limiting transistor to reach the same level of distortion. Again, the output buffer reduces the IIP₃ of this design.

The two curves in Figure 6.12 represent the IIP₃ of LNA with and without EERDS method, respectively. Without EERDS method, the IIP₃ of the LNA ranges from -7.33 dBm to -4.24 dBm depending on at what frequency the two tone test is carried out. With EERDS method, the IIP₃ of the LNA ranges from -4.76 dBm to -1.05 dBm depending on at what frequency the two tone test is carried out. Thus an improvement of 3 dB in IIP₃ is
observed over the whole frequency range (3-10 GHz) with the application of EERDS method.

The simulation results of the output referred 3\textsuperscript{rd} order interception point (OIP\textsubscript{3}) of high linearity 3-10 GHz CMOS LNA are shown in Figure 6.13. The simulated OIP\textsubscript{3} remains relatively constant within the design frequency range. Similarly, the output buffer reduces the OIP\textsubscript{3} of this design.

![Figure 6.13: OIP\textsubscript{3} of high linearity 3-10 GHz CMOS LNA](image)

The two curves in Figure 6.13 represent the OIP\textsubscript{3} of LNA with and without EERDS method, respectively. Without EERDS method, the OIP\textsubscript{3} of the LNA ranges from 9.46 dBm to 11.19 dBm depending on at what frequency the two tone test is carried out. With EERDS method, the OIP\textsubscript{3} of the LNA ranges from 12.27 dBm to 14.93 dBm. Thus an
improvement of 3 dB in OIP$_3$ is observed over the whole frequency range (3-10 GHz) with the application of EERDS method. Thus the simulation data proves that the EERDS method effectively increases the IIP$_3$ and OIP$_3$ of the LNA.

In order for a design to have practical usage, it has to be able to sustain any variations in temperature, supply voltage and process. Figure 6.14 shows that the IIP$_3$ has about 4.5 dB of variation from 20°C to 100°C, which is mainly due to the reduced gain at higher temperature. The OIP$_3$ is almost constant over the temperature range. In Figure 6.15, with supply voltage varying from 1.5 V to 1.7 V, the variation of IIP$_3$ and OIP$_3$ are 1.5 dB and 2 dB, respectively. The Monte Carlo Statistical simulation with process and mismatch variation yields a sigma of 1.31 dB for IIP$_3$ and 0.94 dB for OIP$_3$, as can be observed in Figure 6.16. These results prove that EERDS method is inherently robust to any temperature, supply voltage and process variation, which mainly attributes to its higher tolerance to $V_{gs}$ shift with extended effective range for linearization.
Figure 6.14: High linearity 3-10 GHz CMOS LNA IP₃ temperature variation

Figure 6.15: High linearity 3-10 GHz CMOS LNA IP₃ supply voltage variation
Figure 6.16: High linearity 3-10 GHz CMOS LNA IP3 process variation

Figure 6.17: The effect of tone spacing of two tone test on linearization
From equation (5.35), the IIP$_3$ of the LNA is a function of the tone spacing of the two tone test. When the tone spacing is large enough, the term \( \frac{L_i}{C_{gs}} \left( \frac{1}{j\Delta\omega L_s + j\Delta\omega C_{gs} + Z_s} \right) \) has a non-negligible value and start to show its effect. In order to study how the large tone space affect the measured IIP$_3$, the LNA with active balun is simulated with fundamental at 8GHz and two tone frequency spacing from 1 MHz to 1 GHz. In Figure 6.17, the IIP$_3$ of the designed LNA with and without EERDS method is plotted versus the frequency spacing of the two tone test. It can be observed that without EERDS method, IIP$_3$ basically keep constant with increasing tone spacing. With the application of EERDS method, the measured IIP$_3$ starts to decrease at the tone spacing of 10 MHz and drop by only 1 dB at the tone spacing as large as 1 GHz. Thus even though the EERDS linearization is slightly degraded with large tone spacing, significant linearity improvement still hold.

The input referred 1 dB compression point of the LNA was simulated by sweeping the input power of the LNA and locate the level of input power that causes 1 dB of gain compression. Figure 6.18 and Figure 6.19 show the simulation results of the LNA with and without EERDS method, respectively. The LNA with EERDS method exhibits a $P_{1\text{dB}}$ of -13.95 dBm while the LNA without EERDS method has a $P_{1\text{dB}}$ of -15.4 dBm. Thus the EERDS method improves the $P_{1\text{dB}}$ by 1.5 dB, which is less compared to the improvement on IIP$_3$. This is because when the power level increases, the signal level at the input of
the EERDS composite nfet becomes larger and causes $V_{gs}$ to go out of the effective range, which degrades the effectiveness of the linearity improvement. Considering the LNA being the first stage of the receiver train and typically operates at power level far below the $P_{1dB}$, this will not affect the linearity performance of the LNA.

Figure 6.18: $P_{1dB}$ of high linearity 3-10 GHz CMOS LNA with EERDS method
In both plots, the data of the complete LNA and the LNA without buffer are given. For LNA with EERDS methods, the output buffer costs 1 dB of gain and 1 dB of $P_{1\text{dB}}$. For LNA without EERDS methods, the output buffer costs 1.5 dB of gain and 1.5 dB of $P_{1\text{dB}}$. When LNA is driving on-chip circuits, the output buffer is not necessary and the $P_{1\text{dB}}$ can be improved accordingly.

### 6.3.5 Reverse Isolation and Stability

At multi-GHz frequency, the capacitive coupling between traces, substrate, and devices becomes increasingly significant. If strong coupling presents between any nodes with
high signal level and the input node and creates a positive feedback, the stability of the amplifier would be compromised and in worst case, self-oscillation occurs.

In order to evaluate the isolation between the output and the input, $S_{12}$ of the designed LNA was simulated and the results are shown in Figure 6.20. It can be observed that in both cases (with and without EERDS method), the reverse isolation is lower than -50 dB. Several factors contribute to this high reverse isolation. First, the designed LNA consists of 3 cascaded stages, which increases the reverse isolation. Second, the output of the LNA is differential signal with good balance, which further reduces the possibility of feedback through ground or substrate. Third, separate power supply and ground pins are used for different stages, which minimize the undesired interaction between adjacent stages.

Figure 6.20: $S_{12}$ of high linearity 3-10 GHz CMOS LNA
Further, the Rollet’s stability factor $k_f$ of the high linearity 3-10 GHz CMOS LNA was simulated over the frequency range from 1 MHz to 50 GHz in order to assure the stability of the LNA. As shown in Figure 6.21, the simulated $k_f$ is well above 1 from 1 MHz to 50 GHz, which indicates that the designed LNA is unconditionally stable with sufficient safe margin for process variations and modeling inaccuracy.

**6.3.6 Conclusions**

The LNA input port takes single-ended signal and converts it to differential signal with good signal balance (gain imbalance $< 0.2$ dB, phase imbalance $< 4$ degree), which
benefits the design of the following stages. Both input and output ports are matched to 50 Ω impedance. The LNA and balun consume a total power of 30.3 mW. The designed LNA exhibits a flat gain response of 12.5 dB and a noise figure of 5 dB from 3-9 GHz and increases up to 7 dB at 10 GHz. A high IIP3 of -1 dBm has been achieved. The simulation results prove that the EERDS method effectively improves the IIP3 of the LNA. In addition, the EERDS method is proved to be robust in presence of voltage, process, and temperature variations. The improvement of P1dB is less significant due to the limitation of the effective range of the Vgs for EERDS method.
CHAPTER 7

WIDEBAND / NARROWBAND DUEL MODE CMOS UWB LNA WITH ACTIVE BALUN

7.1 Introduction

The UWB LNA described in CHAPTER 6 provides high linearity which increases the robustness of the UWB transceiver to strong narrowband interferers. Further improvement on the interference robustness can be achieved by adding flexibility in frequency domain. Programmable band-select front-end circuit is developed in this work to enhance the robustness of the UWB transceiver to strong narrowband interferers that is proposed in this work. With a band-select LNA, the transceiver can detect and avoid the strong narrowband interferers, which greatly improve the ability of UWB RF front-end to handle strong narrowband interferers. In this work, a 3-10 GHz CMOS Band-select (wideband/narrowband) UWB LNA circuit with active balun is designed and fabricated in IBM 0.13 μm CMOS with 1.6 V supply voltage.
7.2 Circuit Design and Implementation

7.2.1 Wideband / Narrowband Dual Mode CMOS UWB LNA Core Circuit

The designed 3-10 GHz CMOS Band-select (wideband/narrowband) UWB LNA circuit has an input stage (including the matching network and the common source stage) that is quite similar to the wideband LNA described in CHAPTER 6. The band-select is realized by altering the shunt peaking of the amplifier, as shown in Figure 7.1. In wideband mode, the BS (band-select) signal enable transistor M3, which provide a low impedance path to VDD. As a result, L_{L2} and R_{L2} are bypassed and the impedance of L_{L1} and R_{L1} dominates. L_1 and R_1 are designed to provide an optimum flat gain response over 3-10 GHz. When the LNA switches to narrowband mode, the BS (band-select) signal disable transistor M3, the low impedance path through M3 is cut off and the impedance of L_2 and R_2 add to that of L_1 and R_1. By carefully choosing the value of L_2 and R_2, the frequency range of the LNA can be reduced to 3 to 5 GHz to accommodate the band group 1 operation.

The switch, however, is never perfect at GHz frequency range. In wideband mode, the pFET M3 is switched on to provide a low impedance path. The on resistance of M3 has a magnitude of several ohm which is non-negligible compared to the impedance of L_{L2} and R_{L2}. In order to minimize the on resistance, large width is desired for M3. In narrowband mode, the pFET M3 is switched off. The resistance looking into the drain is almost infinity. However, the gate drain capacitance and the body drain capacitance of the M3 provides a
much lower impedance to the ac ground, which can only be improved by reducing the size of the switch. Thus a tradeoff exists in terms of switch device size when optimizing the performance in both modes. In addition, the effect of the switch device on the passive load network has to be taken into consideration.

Figure 7.1: 3-10 GHz CMOS Band-select (wideband/narrowband) UWB LNA circuit

7.2.2 Active Balun Circuit

The 3-10 GHz CMOS active balun circuit has similar topology with the wideband active balun described in CHAPTER 6. Since this design utilizing band-select to deal with the
interference problem, the extended range derivative superposition method is not applied in the active balun. The composite core devices have been replaced by single MOSFET.

![Diagram of 3-10 GHz CMOS active balun circuit]

Figure 7.2: 3-10 GHz CMOS active balun circuit

### 7.2.3 Output Buffer Circuit

The output buffer circuit has a similar push pull topology compared to the output buffer described in CHAPTER 6. The major difference here is that the composite MOSFETs have been replaced by simple MOSFETs. The bias voltage of the $g_{m}$ device is programmable and is controlled by the comp_enable bit. The buffer is designed to drive a 50 Ω load over a broad frequency band from 3-10 GHz.
7.3 Results and Discussion

The designed 3-10 GHz band-select (narrowband / wideband) CMOS LNA was laid out and extracted with the parasitic resistance, capacitance and inductance taken into account. The complete layout of the LNA, active balun and the output buffer is shown in Figure 7.4. The chip area of the layout is 1.12 mm$^2$. Using Cadence SpectreRF simulator, the extracted view of the design was simulated and evaluated.
Figure 7.4: Layout of 3-10 GHz band-select (narrowband / wideband) CMOS LNA with active balun

### 7.3.1 Impedance Matching

The simulation result of input matching of 3-10 GHz band-select (narrowband / wideband) CMOS LNA is shown in Figure 7.5. The simulated $S_{11}$ is lower than -10 dB over the most part of the whole designed frequency range (3-10 GHz), and slightly higher
than -10 dB at the very low end of the frequency range. This good input matching guarantees maximal power can be delivered to the input port of the LNA, which in turn helps with gain and noise performance. Since the input matching network remains unchanged, similar ultra-wideband input matching is obtained in both low band and high band modes. The minor difference observed is caused by the change of load impedance of the first stage.

The simulation result of output matching of 3-10 GHz band-select (narrowband / wideband) CMOS LNA is shown in Figure 7.6. The simulated $S_{22}$ is lower than -10 dB over the whole designed frequency range (3-10GHz). This good output matching
guarantees maximal power can be delivered to the load of the LNA, which in turn helps with gain and linearity.

![S22 (dB) vs Frequency (GHz) for Narrow band mode and Wide band mode](image)

Figure 7.6: Output impedance matching of 3-10 GHz band-select (narrowband / wideband) CMOS LNA

### 7.3.2 Gain Response

Figure 7.7 shows the simulation results of the gain response of 3-10 GHz band-select (narrowband / wideband) CMOS LNA. In wideband mode, the simulated $S_{21}$ remains flat at the level of 15 dB from 3 to 10 GHz, and starts to roll off rapidly at 10 GHz. In narrowband mode, the simulated $S_{21}$ remains flat at the level of 20 dB from 3 to 5 GHz, and starts to roll off rapidly at 5 GHz. With such band-select capability, the LNA can be
adjusted according to the interferer level and frequency at the time of operation. When no strong interferer is detected, wideband mode can be used to take the advantage of wider bandwidth. When strong interferer is detected at frequency higher than 5 GHz (e.g. 802.11a signal), low band mode can be used to reduce the bandwidth in order to attenuate the high interferer and mitigate the adverse effect. As a result, the robustness of the UWB front-end circuit to high frequency strong interferers can be significantly enhanced.

![Graph showing gain response of 3-10 GHz band-select (narrowband / wideband) CMOS LNA](image)

**Figure 7.7:** Gain response of 3-10 GHz band-select (narrowband / wideband) CMOS LNA

In order to evaluate the effectiveness of the single-ended to differential conversion, the signal imbalance at the output was simulated. Figure 7.8 and Figure 7.9 show the gain and phase imbalance of the 3-10 GHz band-select (narrowband / wideband) CMOS LNA.
respectively. It can be observed that the gain imbalance is less than 0.3 dB and the phase imbalance is less than 10 degree over the whole frequency range. Switching between two bands has minimal effect on the signal imbalance.

Figure 7.8: 3-10 GHz band-select (narrowband / wideband) CMOS LNA gain imbalance
7.3.3 Noise Figure

Figure 7.10 shows the simulation result of the noise figure of 3-10 GHz band-select (narrowband / wideband) CMOS LNA. In wideband mode, the simulated noise figure remains flat at the level of 5 dB from 3 to 8 GHz and reaches 7 dB at 10 GHz. In narrowband mode, the simulated noise figure remains flat at the level of 3 dB from 3 to 5 GHz. In real application where the LNA drives on-chip circuits, the noise figure of the LNA can be significantly improved.
7.3.4 Linearity

The simulation result of the input referred $3^{rd}$ order interception point ($IIP_3$) of 3-10 GHz band-select (narrowband / wideband) CMOS LNA is shown in Figure 7.11. In wideband mode, the $IIP_3$ of the LNA ranges from -7.51 dBm to -4.50 dBm depending on at what frequency the two tone test is carried out. In narrowband mode, the $IIP_3$ of the LNA ranges from -14.6 dBm to -9.25 dBm in 3-5 GHz frequency range. At frequency higher than 5 GHz, the $IIP_3$ increases rapidly and reaches a maximum value of 6.0 dBm at 10 GHz due to the gain roll off. This proves that the LNA in narrowband mode is more immune to high frequency interferers. The corresponding $OIP_3$ is shown in Figure 7.12.
Figure 7.11: IIP$_3$ of 3-10 GHz band-select (narrowband / wideband) CMOS LNA

Figure 7.12: OIP$_3$ of 3-10 GHz Band-select (narrowband / wideband) CMOS LNA
The input referred 1 dB compression point of the 3-10 GHz Band-select (narrowband / wideband) CMOS LNA was simulated by sweeping the input power of the LNA and locate the level of input power that causes 1 dB of gain compression. Figure 7.13 and Figure 7.14 show the simulation results of the LNA in narrowband and wideband mode, respectively. The LNA in narrowband mode has a $P_{1\text{dB}}$ of -21.29 dBm while the LNA in wideband mode has a $P_{1\text{dB}}$ of -15.61 dBm.

In both plots, the data of the complete LNA and the LNA without buffer are given. For LNA in narrowband mode, the output buffer costs 2.6 dB in $P_{1\text{dB}}$. For LNA in wideband mode, the output buffer costs 3.1 dB in $P_{1\text{dB}}$. When LNA is driving on-chip circuits, the output buffer is not necessary and the $P_{1\text{dB}}$ can be improved accordingly.

Figure 7.13: $P_{1\text{dB}}$ of 3-10 GHz Band-select (Narrowband mode) CMOS LNA
7.3.5 Reverse Isolation and Stability

The reverse isolation $S_{12}$ of the designed LNA was simulated and the results are shown in Figure 7.15. It can be observed that in both narrowband and wideband modes, the reverse isolation is lower than -45 dB. The factors that contribute to this high level of reverse isolation have been elaborated in section 6.3.5 and are not repeated here. Further, the Rollet’s stability factor $k_f$ of the 3-10 GHz band-select (narrowband / wideband) CMOS LNA was simulated over the frequency range from 1 MHz to 50 GHz in order to assure the stability of the LNA. As shown in Figure 7.16, the simulated $k_f$ is well above 1 from 1 MHz to 50 GHz, which indicates that the designed LNA is unconditionally stable with sufficient safe margin for process variations and modeling inaccuracy.
Figure 7.15: $S_{12}$ of 3-10 GHz Band-select (narrowband / wideband) CMOS LNA

Figure 7.16: Rollet's stability factor $k_f$ of 3-10 GHz Band-select (narrowband / wideband) CMOS LNA
7.3.6 Conclusions

The designed LNA exhibits excellent performance in both narrowband and wideband mode. The LNA input port takes single-ended signal and converts it to differential signal with good signal balance (gain imbalance < 0.3 dB, phase imbalance <10 degree), which benefits the design of the following stages. Both input and output ports are matched to 50 Ω impedance. In narrowband mode, flat gain of 20 dB, noise figure of 3 dB, IIP₃ of -14 to -9 dBm and P₁dB of -21.9 dBm are achieve over frequency range of 3-5 GHz. In wideband mode, flat gain of 15 dB, noise figure of 5-7 dB, IIP₃ of -5 dBm and P₁dB of -15.6 dBm are achieved over frequency range of 3-10 GHz. The LNA and balun consume a total power of 38.7 mW. The simulation results proved that the band-select capability effectively relaxes the tradeoff between robustness to strong narrowband interferers and high throughput that comes with ultra wide bandwidth.
CHAPTER 8

HIGH BAND / LOW BAND DUEL MODE CMOS UWB LNA
WITH ACTIVE BALUN

8.1 Introduction

The band-select UWB LNA described in CHAPTER 7 offers selection between wideband and narrowband modes which greatly improves the robustness of the receiver to strong narrowband interferers. However, when a strong narrowband interferer exists in low band, switching to the wideband mode increases the IIP3 of the LNA in low band but does not provide any attenuation to the interferer. In order to provide maximal protection of front-end circuit from strong narrowband interferers, a band-select LNA with non-overlapping bands is highly desirable. In this work, a 3-10 GHz CMOS band-select (high band/low band) UWB LNA circuit with active balun is designed and fabricated in IBM 0.13 μm CMOS with 1.6 V supply voltage. The bias circuit and the output buffer are similar to the circuit in CHAPTER 7 and are not discussed in detail here.
8.2 Circuit Design and Implementation

8.2.1 High Band / Low Band Duel Mode CMOS UWB LNA Core Circuit

In order to implement an amplifier with two non-overlapping bands, the required passive load networks are significantly different. As a result, it will be very difficult to achieve such band-select simply by switch in/out some impedance. New band-select method is necessary in order to achieve this goal.

The designed 3-10 GHz CMOS band-select (high band/low band) UWB LNA circuit is shown in Figure 8.1. The input stage (including the matching network and the common source stage) is quite similar to the wideband LNA described in CHAPTER 6. Novel band-select method is developed to provide non-overlapping band-select. In order to eliminate the interaction between two passive load network, two separate cascode devices are used. In high band mode, $V_{b2}$ is biased at 1.2V while $V_{b3}$ is pulled down to ground. M$_3$ is completely cut off and all the drain current is flowing through M$_2$. The bandpass passive load network with a bandwidth from 6 to 10 GHz then convert the ac current to voltage signal and drive the next stage. In low band mode, $V_{b3}$ is biased at 1.2V while $V_{b2}$ is pulled down to ground. Thus M$_2$ is completely cut off and all the drain current is flowing through M$_3$. The passive load network for low band mode then convert the ac current to voltage signal and drive the next stage.
For low band mode, shunt peaking is used to provide a maximal flat gain response from 3 to 5 GHz to accommodate the band group 1 operation. For high band mode, a novel passive load network is used to provide a maximal flat gain response from 6 to 10 GHz and significant attenuation at lower band groups in the mean time. This passive load network is essentially a 2nd order bandpass filter. From another perspective, it can be viewed as a band pass filter consists of $L_{1,3}$ and load capacitor and a band stop filter consists of $C_1$, $L_{L,1}$, and $R_{L,1}$ with center frequency almost overlap with each other. With careful tuning of passive elements, a band pass filter with maximal flat response in passband and fast roll off in the stopband can be achieved.
8.2.2 Active Balun Circuit

Since the band-select (high band/low band) UWB LNA has two separate output ports for high band and low band operation, two input stages are necessary in balun circuit. The input $V_{\text{in,HB}}$ and $V_{\text{in,LB}}$ are used for high band mode and low band mode, respectively. In high band mode, the BS signal is high, which enable nfet $M_2$. Thus nfet $M_1$ is in saturation region and the single-ended high band input signal converted to differential signal which is fed to the following common source stage. Inductive peaking is used in the load in order to provide a flat gain response. In the meanwhile, $\text{BS}_b$ signal goes low and disable the nfet $M_4$. No drain current is flowing through the nfet $M_3$. Thus the low band input stage is disabled in high band mode.

![Variable gain 3-10 GHz CMOS active balun circuit](image)

Figure 8.2: Variable gain 3-10 GHz CMOS active balun circuit
In low band mode, the BS signal is low, which disable nfet M2. No drain current is flowing through the nfet M1. Thus the low band input stage is disabled in high band mode. In the meanwhile, BS_b signal goes high and enable the nfet M4. Thus nfet M3 is in saturation region and the single-ended low band input signal converted to differential signal which is fed to the following common source stage. Due to the lower operation frequency in low band mode, simple resistive load can be used in order to reduce the chip area.

The high band and low band signal paths have to be converged in order to use the same mixer and base band circuitry. In order to minimize the effect of extra capacitive loading, this converge point is chosen to be the source of the cascode stage. The relatively lower impedance looking into the source of the cascode stage makes it less sensitive to capacitive loading. Two groups of common source nfets are used for different modes. M5 – M8 are for high band mode only and M9 – M12 are for low band mode only. In either mode, the ac current is fed through the cascode stage and is converted to voltage signal by inductively peaking passive load at the output.

The gain switching is implemented in two different ways. First, the comp_enable bit controls the bias current of the common source nfet, which determines the gain of the amplifier. In addition, two nfets in each mode (M7 and M8 in high band mode / M11 and M12 in low band mode) can be switched on or off by gain control bit. Thus a total of 4 gain steps are available to increase the dynamic range of the amplifier.
8.3 Results and Discussion

The designed 3-10 GHz band-select (low band / high band) CMOS LNA was laid out and extracted with the parasitic resistance, capacitance and inductance taken into account. The complete layout of the LNA, active balun and the output buffer is shown in Figure 8.3. The chip area of the layout is 1.44 mm$^2$. Using Cadence SpectreRF simulator, the extracted view of the design was simulated and evaluated.

Figure 8.3: Layout of 3-10 GHz band-select (low band / high band) CMOS LNA with active balun
8.3.1 Impedance Matching

The simulation result of input matching of 3-10 GHz band-select (low band / high band) CMOS LNA is shown in Figure 8.4. The simulated $S_{11}$ is lower than -10 dB over the most part of the designed frequency range (3-10 GHz), and slightly exceeds -10 dB at the very low end of the frequency range. Since the input matching network remains unchanged, similar Ultra-wideband input matching is obtained in both low band and high band modes. The minor difference observed is caused by the difference of load impedance of the first stage in high band and low band mode.

![Figure 8.4: Input impedance matching of 3-10 GHz Band-select (low band / high band) CMOS LNA](image-url)
The simulation results of output matching of 3-10 GHz band-select (low band / high band) CMOS LNA are shown in Figure 8.5. The simulated $S_{22}$ is lower than -10 dB over the whole designed frequency range (3-10GHz). This good output matching guarantees maximal power can be delivered to the load of the LNA.

Figure 8.5: Output impedance matching of 3-10 GHz Band-select (low band / high band) CMOS LNA

### 8.3.2 Gain Response

Figure 8.6 shows the simulation results of the gain response of 3-10 GHz band-select (low band / high band) CMOS LNA. In low band mode, the simulated $S_{21}$ remains flat at the level of 15 dB from 3 to 5 GHz, and starts to roll off rapidly at 5 GHz. In high band
mode, the simulated $S_{21}$ remains flat at the level of 15 dB from 6 to 10 GHz, and starts to roll off rapidly at frequency below 6 GHz or higher than 10 GHz. With such band-select capability, the LNA can be adjusted according to the interferer level and frequency at the time of operation. When strong interferer is detected in low band, high band mode can be used to mitigate the interference, and vice versa. As a result, the robustness of the UWB front-end circuit to high frequency strong interferers can be significantly enhanced.

Figure 8.6: Gain response of 3-10 GHz Band-select (low band / high band) CMOS LNA

The simulation results of gain steps obtained with low band and high band mode are shown in Figure 8.7 and Figure 8.8, respectively. In each case, 8 dB of variable gain range is achieved with 4 gain steps, which further enhance the adaptability and robustness of the receiver front-end.
Figure 8.7: Gain response of 3-10 GHz Band-select (low band mode) CMOS LNA at different gain settings.

Figure 8.8: Gain response of 3-10 GHz Band-select (high band mode) CMOS LNA at different gain settings.
In order to evaluate the effectiveness of the single-ended to differential conversion, the signal imbalance at the output was simulated. Figure 8.9 and Figure 8.10 show the gain and phase imbalance of the 3-10 GHz band-select (low band / high band) CMOS LNA respectively. It can be observed that the gain imbalance is less than 0.4 dB and the phase imbalance is less than 10 degree over the corresponding frequency range in each mode. Switching between two bands has minimal effect on the signal imbalance.

Figure 8.9: 3-10 GHz Band-select (low band / high band) CMOS LNA gain imbalance
8.3.3 Noise Figure

Figure 8.11 shows the simulation results of the noise figure of 3-10 GHz band-select (low band / high band) CMOS LNA. In low band mode, the simulated noise figure remains flat at the level of 3.7 dB from 3 to 5 GHz. In high band mode, the simulated noise figure ranges from 6 to 7 dB from 6 GHz to 10 GHz. The relatively higher noise figure is caused by the lossy passive filter for band pass characteristics and the output buffer to drive the signal off-chip. In real application where the LNA drives on-chip circuits, the noise figure of the LNA can be significantly improved.
Figure 8.11: 3-10 GHz Band-select (low band / high band) CMOS LNA noise figure

8.3.4 Linearity

The simulation result of the input referred 3rd order interception point (IIP3) of 3-10 GHz band-select (low band / high band) CMOS LNA is shown in Figure 8.12. In low band mode, the IIP3 of the LNA ranges from -14 dBm to -10 dBm depending on at what frequency the two tone test is carried out. In high band mode, the IIP3 of the LNA remains relatively constant at -5 dBm in 6-10 GHz frequency range. As shown in Figure 8.13, the corresponding OIP3 is 7 dBm and 12.5 dBm for low band and high band mode, respectively.
Figure 8.12: IIP₃ of 3-10 GHz Band-select (low band / high band) CMOS LNA

Figure 8.13: OIP₃ of 3-10 GHz Band-select (low band / high band) CMOS LNA
The input referred 1 dB compression point of the 3-10 GHz band-select (low band / high band) CMOS LNA was simulated by sweeping the input power of the LNA and locate the level of input power that causes 1 dB of gain compression. Figure 8.14 shows the simulation results of the LNA input referred $P_{1\text{dB}}$ in low band and high band mode. The LNA in low band mode has a $P_{1\text{dB}}$ of -21.73 dBm while the LNA in high band mode has a $P_{1\text{dB}}$ of -15.29 dBm.

Figure 8.15 shows the simulation results of input referred $P_{1\text{dB}}$ of the complete LNA and the LNA without buffer in high band mode. The output buffer costs 2.6 dB in $P_{1\text{dB}}$ in high band mode and 2.8 dB in $P_{1\text{dB}}$ in low band mode. When LNA is driving on-chip circuits, the output buffer is not necessary and the $P_{1\text{dB}}$ can be improved accordingly.

![Figure 8.14: $P_{1\text{dB}}$ of 3-10 GHz Band-select (low band / high band) CMOS LNA](image-url)
Figure 8.15: Effect of output buffer on $P_{1\text{dB}}$ (high band mode)

### 8.3.5 Reverse Isolation and Stability

The reverse isolation $S_{12}$ of the designed LNA was simulated and the results are shown in Figure 8.16. It can be observed that in both low band and high band modes, the $S_{12}$ is lower than -45 dB. To further examine the stability of the amplifier, the Rollet’s stability factor $k_f$ of the 3-10 GHz band-select (low band / high band) CMOS LNA was simulated over the frequency range from 1 MHz to 50 GHz in order to assure the stability of the LNA. As shown in Figure 8.17, the simulated $k_f$ is well above 1 from 1 MHz to 50 GHz, which indicates that the designed LNA is unconditionally stable with sufficient safe margin for process variations and modeling inaccuracy.
Figure 8.16: $S_{12}$ of 3-10 GHz Band-select (low band / high band) CMOS LNA

Figure 8.17: Rollet’s stability factor $k_f$ of 3-10 GHz Band-select (low band / high band) CMOS LNA
8.3.6 Conclusions

The designed LNA exhibits excellent gain and noise performance and provide band-select flexibility at the same time. The passband of the LNA is selectable between 3-5 GHz and 6-10 GHz. The LNA input port takes single-ended signal and converts it to differential signal with good signal balance (gain imbalance < 0.4 dB, phase imbalance <10 degree), which benefits the design of the following stages. Both input and output ports are matched to 50 Ω impedance. In low band mode, flat gain of 15 dB, noise figure of 3.7 dB, IIP3 of -14 to -10 dBm and P1dB of -21.73 dBm are achieve over frequency range of 3-5 GHz. In high band mode, flat gain of 15 dB, noise figure of 6-7 dB, IIP3 of -5 dBm and P1dB of -15.29 dBm are achieve over frequency range of 6-10 GHz. In each mode, gain can be adjusted within a range of 8 dB. The LNA and balun consume a total power of 29.1 mW. The simulation results prove that the band-select and tunable gain capability effectively increases the adaptability and robustness of the receiver front-end circuit.
CHAPTER 9

CONCLUSIONS

The rapid development of the UWB technology demands more and more stringent performance requirement on RF front-end circuit. LNA as the first block of the receiver train is particularly crucial to the overall receiver performance. The ideal LNA design is supposed to provide low noise, high and flat gain, high linearity over multi-GHz frequency range while consuming minimum amount of power. In practical design, these performances tradeoff with each other and can not be optimized simultaneously. In this work, the design considerations and tradeoffs on different key performances in UWB front-end circuit design are discussed.

With increasing level of integration of the RF transceiver, on-chip active balun is more frequently used to replace its off-chip counterpart, which simplifies the board level design and saves the cost and board area as well. Several balun topologies and their performance are discussed in the context of UWB application.
In UWB system, the front-end circuits are exposed to strong narrowband interferers over a very wide bandwidth. From a system perspective, the effect of narrowband interferer on the UWB front-end circuit is studied. It is concluded that high linearity LNA helps to alleviate this problem. Particularly, in the UWB receiver adopting the concept of ADC in frequency domain which divides the wide bandwidth into multiple subbands, the linearity of the UWB LNA becomes the bottleneck of the receiver and sets the upper limit of the receiver dynamic range. Thus the linearity of the LNA has more significance in UWB receiver than other typical receivers.

At DC, the linearity model of the MOSFET device in both strong inversion and weak inversion region is developed from the short channel device model and subthreshold current model. Simulation results prove that the model well predicts the power series coefficients and the $A_{IP3}$ of the MOSFET device. This linearity model provides some insights on the nonlinear behavior of the MOSFET device and the possible linearity improvement methods as well. In particular, for derivative superposition method, it shows the best range of $V_{gs}$ that is suitable for main and auxiliary devices. Further, this model allows the power series coefficients to be calculated directly from the device size and the DC bias, thus makes it easier to match the main and auxiliary devices for $g_3$ cancellation.

Based on the DC linearity model, Extended Effective Range Derivative Superposition (EERDS) method is developed. Compared with conventional Superposition method,
EERDS method increases the effective range of gate bias in which the derivative superposition linearization remains effective. This extended effective range has two major advantages. First, it greatly relaxes the requirement for precise gate bias, which makes the design more robust to temperature, supply voltage, mismatch, and process variation. Second, when the signal at the input is large, the effectiveness of conventional derivative superposition method degrades. With EERDS method, the extended effective range can accommodate large signal without causing much extra distortion and improve the large signal linearity significantly.

At high frequency, the mechanism of $3^{rd}$ order intermodulation product generation is much more complex due to the feedback through parasitics and passive devices. A formal analysis of such mechanism in a common source amplifier with inductive source degeneration is carried out using small signal model which takes into account the $2^{nd}$ and $3^{rd}$ nonlinearity. An analytical expression of $A_{\text{IIP3}}$ is derived as a function of frequency and source inductance. The source inductance is proved to have a major impact on linearization at high frequency. In order to solve this problem, the EERDS is modified slightly to mitigate the effect of source inductor. The effectiveness of this technique is proved by the simulation results.

With the application of proposed EERDS method, a 3-10 GHz high linearity UWB LNA with active balun is designed and fabricated in IBM 0.13 μm CMOS process. The designed LNA exhibits a flat gain response of 12.5 dB and a noise figure of 5 dB in the
majority part of 3-10 GHz frequency range (with slight degradation at high frequency end). A high IIP3 of -1 dBm has been achieved. Good signal balance (gain imbalance < 0.2 dB, phase imbalance < 4 degree) is observed at the output. The simulation results prove that the EERDS method effectively improves the IIP3 and P1dB of the LNA and balun. In addition, the EERDS method is proved to be robust in presence of voltage, process and temperature variations. The EERDS method can also be applied to other ultra-wideband RF front-end circuit such as mixer, RF VGA and PA.

In addition to the proposed linearization techniques, programmable band-select front-end circuit is another approach developed in this work to enhance the robustness of the UWB transceiver to strong narrowband interferers. With a band-select LNA, the transceiver can detect and avoid the strong narrowband interferers, which greatly improve the ability of UWB RF front-end to survive strong narrowband interferers. Two 3-10 GHz band-select UWB LNAs, one configurable between narrowband (3-5 GHz) and wideband (3-10 GHz) mode and the other configurable between low band (3-5 GHz) and high band (6-10 GHz) mode, are designed and fabricated in IBM 0.13 μm CMOS process to demonstrate the proposed concept. Both LNAs exhibit excellent gain and noise performance and provide band-select flexibility at the same time. Table 9.1 shows the major performance index of the two band-select LNAs designed in this work. It is proved that the band-select and tunable gain capability effectively increases the adaptability and robustness of the receiver front-end circuit.
Table 9.1: Summary of band-select LNA with active balun performance

A comparison of this work and other reported works is shown in Table 9.2. It can be observed that with EERDS method, the IIP₃ and P₁dB of the 3-10 GHz high linearity CMOS LNA are effectively improved without compromising other key performances. The higher power consumption is due to the extra stage of the active balun circuit. For two band-select LNAs, moderate linearity is observed. However, the band-select capability allows the front-end circuit to detect and avoid the strong narrowband interferers, which provides the receiver more adaptability and thus better robustness.

Table 9.2: Summary of LNA performance and comparison with other works

1*: 3-10 GHz high linearity CMOS LNA with active balun (CHAPTER 6)
2*: 3-10 GHz band-select (narrowband / wide band) CMOS LNA with active balun (CHAPTER 7)
3*: 3-10 GHz band-select (low band / high band) CMOS LNA with active balun (CHAPTER 8)
APPENDIX A

UWB LNA AND BALUN TEST PLAN

General Information

Input signal: Single-ended
Input impedance: 50 ohm
Output signal: Differential
Output impedance: 50 ohm each (p or n)
Power supply: 1.6V
Logic high: 1.6V
Logic low: 0V
Settling time: Allow 500 ns for circuits to settle after any switch of control bits
Power up sequence: Use a ramp time larger than 100 us to power up the chip.
IC Protection: RF circuits don’t have ESD protection, (except control bits). Keep the chip away from ESD hazard.
RF signal can be applied before or after power up.

Avoid voltage higher than 1.8 V.

Signal coupling: Both input and output signal can be dc coupled thus no coupling capacitor is needed.

Signal dc bias: No dc bias is needed for both input and output. The signal path is ac coupled on chip.

Clamp capacitor: An on-board clamp capacitor between vcc and gnd is desired

DC bias: No DC bias is necessary for normal mode operation. In external bias mode when vb_en is high, two DC bias voltages are needed to set the bias of the circuit.

**Design Notation**

**Design A:** 3-10 GHz high linearity CMOS LNA with active balun (CHAPTER 6)

**Design B:** 3-10 GHz band-select (narrowband / wide band) CMOS LNA with active balun (CHAPTER 7)

**Design C:** 3-10 GHz band-select (low band / high band) CMOS LNA with active balun (CHAPTER 8)
Operation Modes

Control bits function:

<table>
<thead>
<tr>
<th>Control Bit</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp_disable</td>
<td>Enable composite nfet</td>
<td>Disable composite nfet</td>
</tr>
<tr>
<td>BS</td>
<td>Low/Narrow band</td>
<td>High/Wide band</td>
</tr>
<tr>
<td>Gain</td>
<td>Low gain</td>
<td>High gain</td>
</tr>
<tr>
<td>Vb_en</td>
<td>Self bias</td>
<td>External bias</td>
</tr>
</tbody>
</table>

Control bits scope:

<table>
<thead>
<tr>
<th>Control Bit</th>
<th>Design A</th>
<th>Design B</th>
<th>Design C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp_disable</td>
<td>applicable</td>
<td>applicable</td>
<td>applicable</td>
</tr>
<tr>
<td>BS</td>
<td>applicable</td>
<td>N/A</td>
<td>applicable</td>
</tr>
<tr>
<td>Gain</td>
<td>N/A</td>
<td>N/A</td>
<td>applicable</td>
</tr>
<tr>
<td>Vb_en</td>
<td>applicable</td>
<td>applicable</td>
<td>applicable</td>
</tr>
</tbody>
</table>

External Bias Mode:

The designed circuits are self-biased and no external bias circuits are needed. However, when vb_en bit is 1, the external bias mode is enabled and the bias current is controlled by two voltage signals: vb_high and vb_low. Vb_low can be roughly set at ~600mV. Vb_high need to be precisely controlled over the voltage range from 0.7V to 1.4V with step precision better than 3 mV. Perform a course sweep from 0.7V to 1.4V with a step of
25 mV first to find the voltage range (~200mV) for best performance and then do a fine sweep in the range thus determined with a step less than 3 mV

**Measurements to Be Performed**

**De-embedding the PCB board**

Measure the loss through the PCB transmission line (including the input and the output) over the frequency range from 3 GHz to 10.6 GHz. This loss is going to be considered when evaluating the measurement data.

**DC Measurement**

Measure current consumption of each design

Test all modes for Design A, B and C

**IP3 2 Tone Test**

Frequency range: 3 GHz~10.6 GHz

Two tone separation: 10 MHz and 100 MHz

Input power level: -30 dBm and -20 dBm

Test all modes for Design A, B and C

External bias mode test required.
**P1dB Compression Point**

Frequency range: 3 GHz–10.6 GHz

Input power level: Sweep from -40 dBm to 10 dBm

Test all modes for Design A, B and C

**Transducer Power Gain and Signal Balance**

Frequency range: 3 GHz–10.6 GHz

Frequency step: 100MHz

Input power level: -50 dBm

Test all modes for Design A, B and C

**S-parameter Measurement**

Measure the S-parameter at the input and the output over the frequency range from 3 GHz to 10.6 GHz.

Test all modes for Design A, B and C

**Noise Figure**

Frequency range: 3 GHz–10.6 GHz

Test all modes for Design A, B and C


