# RADIO FREQUENCY CIRCUITS FOR WIRELESS RECEIVER FRONT-ENDS

A Dissertation

by

CHUNYU XIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Electrical Engineering

RADIO FREQUENCY CIRCUITS FOR WIRELESS RECEIVER FRONT-ENDS

A Dissertation

by

CHUNYU XIN

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

| | |
|---|---|
| Edgar Sánchez-Sinencio (Chair of Committee) | Jóse Silva-Martínez (Member) |
| Aydin Ilker Karsilayan (Member) | Don R. Halverson (Member) |
| Eva Sevick-Muraca (Member) | Chanan Singh (Head of Department) |

August 2004

Major Subject: Electrical Engineering

# ABSTRACT

Radio Frequency Circuits for Wireless Receiver Front-ends. (August 2004)

Chunyu Xin, B.S., Nankai University, China;

M.S., Nankai University, China

Chair of Advisory Committee: Dr. Edgar Sánchez-Sinencio

The beginning of the 21st century sees great development and demands on wireless communication technologies. Wireless technologies, either based on a cable replacement or on a networked environment, penetrate our daily life more rapidly than ever. Low operational power, low cost, small form factor, and function diversity are the crucial requirements for a successful wireless product. The receiver's front-end circuits play an important role in faithfully recovering the information transmitted through the wireless channel.

Bluetooth is a short-range cable replacement wireless technology. A Bluetooth receiver architecture was proposed and designed using a pure CMOS process. The front-end of the receiver consists of a low noise amplifier (LNA) and mixer. The intermediate frequency was chosen to be 2MHz to save battery power and alleviate the low frequency noise problem. A conventional LNA architecture was used for reliability. The mixer is a modified Gilbert-cell using the current bleeding technique to further reduce the low frequency noise. The front-end draws 10 mA current from a 3 V power supply, has a 8.5 dB noise figure, and a voltage gain of 25 dB and -9 dBm IIP3.

A front-end for dual-mode receiver is also designed to explore the capability of a multi-standard application. The two standards are IEEE 802.11b and Bluetooth. They work together making the wireless experience more exciting. The front-end is

designed using BiCMOS technology and incorporating a direct conversion receiver architecture. A number of circuit techniques are used in the front-end design to achieve optimal results. It consumes 13.6 mA from a 2.5 V power supply with a 5.5 dB noise figure, 33 dB voltage gain and -13 dBm IIP3.

Besides the system level contributions, intensive studies were carried out on the development of quality LNA circuits. Based on the multi-gated LNA structure, a CMOS LNA structure using bipolar transistors to provide linearization is proposed. This LNA configuration can achieve comparable linearity to its CMOS multi-gated counterpart and work at a higher frequency with less power consumption. A LNA using an on-chip transformer source degeneration is proposed to realize input impedance matching. The possibility of a dual-band cellular application is studied. Finally, a study on ultra-wide band (UWB) LNA implementation is performed to explore the possibility and capability of CMOS technology on the latest UWB standard for multimedia applications.

To my parents, my beloved wife Jinding ⋯

# ACKNOWLEDGMENTS

Upon finishing my graduate study at Texas A&M University, I would like first to give Dr. Edgar Sánchez-Sinencio my most sincere appreciation. As my advisor, he has been the sole support for my research and study in the past several years. I have benefited from his academic foresight and technical insight in research. His great personality made my research experience more joyful and encouraging. I am particularly grateful to him for his advice in the academic realm and also for my professional career.

I would like to thank Dr. Jóse Silva-Martínez for his advice on my course of study and project questions. He was always available for questions and comments. I am also grateful to Dr. Sherif Embabi. I was his teaching assistant for one semester and really learned much from helping him with the class. His attitude toward work and study gave me great encouragement. I also give thanks to Dr. Aydin Ilker Karsilayan. He not only gave advice on my studies, but also supported me and the whole analog and mixed signal center group by maintaining a great UNIX computing environment. Some of the simple yet effective scripts he developed really made life a lot easier.

Most of my research time was spent on two major research projects, the Bluetooth and Chameleon projects. The contribution and cooperation of the team members made the learning and research study a great experience. Here I would like to thank all of the team members, including Wenjun Sheng, Ahmed Emira, Bo Xia, Sung Tae Moon, Ari Yakov Valero-Lopez, Alberto Valdes-Garcia, Ahmed Mohieldin and David Hernandez-Garduño for the great team environment we created together. In these two big projects, we learned from each other, tackled problems together and cheered on each other. I felt the power of team work. I think this valuable experience will

benefit my future professional career greatly.

Thanks should also be given to my committee members, Dr. Don R. Halverson and Dr. Eva Sevick-Muraca, for their time and valuable comments on my research.

I would like to thank the secretary of our group, Ella Gallagher. Her enthusiasm, warm sense of humor and positive attitude is a valuable asset for the group. I am also indebted to all the other professors and students in the analog and mixed signal center for their kind help and technical discussions.

Special thanks go to my parents for their love. Finally, I want to thank my dear wife, Jingding Dou. I would not have been able to achieve this goal without her love and support.

# TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION . . . . . . . . . . . . . . . . . . . . . . . . . . 1

A. Research Motivation . . . . . . . . . . . . . . . . . . . . . 3
B. Dissertation Overview . . . . . . . . . . . . . . . . . . . . 3

II LOW NOISE AMPLIFIER DESIGN OVERVIEW . . . . . . . . . 5

A. Basics on S-parameters . . . . . . . . . . . . . . . . . . . . 5
B. Amplifier's Gain and Stability . . . . . . . . . . . . . . . . 9
C. Noise Performance . . . . . . . . . . . . . . . . . . . . . . 14
1. Two-Port Network Noise Model . . . . . . . . . . . . . 15
2. MOS Transistor Two-Port Noise Parameters . . . . . 21
3. Impact of LNA Gain and Noise Factor on System Sensitivity . . . . . . . . . . . . . . . . . . . . . . . . 24
D. Large Signal Behavior . . . . . . . . . . . . . . . . . . . . 25
1. 1-dB Compression Point . . . . . . . . . . . . . . . . . 27
2. Intercept Point . . . . . . . . . . . . . . . . . . . . . . 28
3. Dynamic Range . . . . . . . . . . . . . . . . . . . . . 29
4. Wide-Band Non-Linearity . . . . . . . . . . . . . . . . 32
E. LNA Topologies in CMOS Technology . . . . . . . . . . . 36
F. Design Procedure of a Source Degenerated CMOS LNA . . 40

III MIXER DESIGN OVERVIEW AND AN IMPLEMENTATION FOR LOW-IF BLUETOOTH RECEIVER . . . . . . . . 53

A. Mixer Mathematical Model . . . . . . . . . . . . . . . . . 53
B. Mixer Metrics . . . . . . . . . . . . . . . . . . . . . . . . 55
1. Conversion Gain or Loss . . . . . . . . . . . . . . . . 56
2. Noise Figure . . . . . . . . . . . . . . . . . . . . . . . 56
3. Port-to-Port Isolation . . . . . . . . . . . . . . . . . . 59
4. Linearity Measurement . . . . . . . . . . . . . . . . . 60
C. Circuit Topologies of Mixers . . . . . . . . . . . . . . . . 60
1. Diode Mixers . . . . . . . . . . . . . . . . . . . . . . . 61
2. Passive Mixer in CMOS Technology . . . . . . . . . . 62
3. Gilbert-cell Mixer . . . . . . . . . . . . . . . . . . . . 64
4. Sub-Sampling Mixer . . . . . . . . . . . . . . . . . . . 65

CHAPTER Page


5. Harmonic Mixer . . . . . . . . . . . . . . . . . . . . . 66
D. Mixer Design for a Low-IF Bluetooth Receiver . . . . . . . 67
1. Low-IF Bluetooth Receiver Architecture . . . . . . . . 69
2. Implementation of the Down-Conversion Mixer . . . . 69
3. Layout Considerations and Simulation Results . . . . 74
4. Experimental Results of the Mixer Within the Receiver 77

IV BLUETOOTH/WI-FI DUAL-STANDARD RECEIVER RF FRONT-END . . . . . . . . . . . . . . . . . . . . . . . . . 80

A. Direct Conversion Bluetooth/Wi-Fi Dual-Standard Receiver 81
B. RF Front-End Design Considerations . . . . . . . . . . . 83
C. Circuits Implementations . . . . . . . . . . . . . . . . . . 86
1. LNA Implementation . . . . . . . . . . . . . . . . . . 87
2. Mixer Implementation . . . . . . . . . . . . . . . . . . 97
3. PTAT Biasing Circuit . . . . . . . . . . . . . . . . . . 100
D. Layout and Experimental Results . . . . . . . . . . . . . 103

V LNA LINEARIZATION TECHNIQUES . . . . . . . . . . . . . 111

A. Non-Linearity Analysis . . . . . . . . . . . . . . . . . . . 112
1. Non-Linear System Representations . . . . . . . . . . 112
2. Non-Linearity of Fully Differential Circuits . . . . . . 115
3. IM3 Due to 5th-Order and 2nd-Order Non-Linearity . 116
4. Non-Linearity Due to Output Impedance . . . . . . . 118
5. Third-Order Distortion of Inductive Source-Degenerated LNA . . . . . . . . . . . . . . . . . . . . . . . . . 122
B. Theoretical Analysis of Multi-Gated Linearization Technique 126
C. Proposed Linearization Scheme Using BJTs in CMOS Process . . . . . . . . . . . . . . . . . . . . . . . . . . . 133
1. Hybrid LNA: BJT as Auxiliary Transistor . . . . . . . 133
2. Volterra Analysis of Resistive-Degenerated BJT . . . . 135
3. Input Matching and Noise Contributions . . . . . . . . 138
4. Sensitivity to Bias Condition, Process Corners and Temperature . . . . . . . . . . . . . . . . . . . . . . . 141
5. The Differential Configuration . . . . . . . . . . . . . 145
D. Measurement Results and Comparisons . . . . . . . . . . . 149

VI A MUTUAL-COUPLED DEGENERATED LNA AND ITS EXTENSION TO CONCURRENT DUAL-BAND OPERATION 155

CHAPTER Page


A. Principle of Impedance Match Using Mutual Inductance . 156
1. Input Impedance . . . . . . . . . . . . . . . . . . . . . . 157
2. Interstage Impedance . . . . . . . . . . . . . . . . . . . 160
3. Effective Transconductance . . . . . . . . . . . . . . . . 162
4. Noise Analysis . . . . . . . . . . . . . . . . . . . . . . 163
B. Chip Measurement Results of the Mutual-Coupled Degenerated LNA . . . . . . . . . . . . . . . . . . . . . . . . 167
C. A Dual-Band Inductive Coupled LNA . . . . . . . . . . . . 168
D. Simulation Results of the Dual-band LNA . . . . . . . . . 171

VII FRONT-END CIRCUITS FOR WIDE-BAND APPLICATION . 175

A. Introduction to Ultra-Wide Band System . . . . . . . . . . 175
B. Distributed RF Front-End Circuits . . . . . . . . . . . . . 178
1. Transmission Line Properties and Characterization . . 178
2. Transmission Lines on Silicon Substrate . . . . . . . . 183
3. Distributed Amplifier as LNA . . . . . . . . . . . . . . 187
4. Analysis of Distributed Mixer . . . . . . . . . . . . . . 198
C. Wide-Band Impedance Match Using Lumped Components 201
1. Impedance Matching Procedure Using Lumped Components . . . . . . . . . . . . . . . . . . . . . . . . . 202
2. Wide-Band LNA with Lumped-Matched Network . . . 205

VIII SUMMARY AND CONCLUSION . . . . . . . . . . . . . . . . 207

REFERENCES . . . . . . . . . . . . . . . . . . . . . . . . . . . . 210

APPENDIX A . . . . . . . . . . . . . . . . . . . . . . . . . . . . 219

APPENDIX B . . . . . . . . . . . . . . . . . . . . . . . . . . . . 222

VITA . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 236

# LIST OF TABLES

TABLE Page

I Important equations for RF amplifier's gain, stability and noise . . . 26

II Equations related to non-linearity . . . . . . . . . . . . . . . . . . . . 35

III Popular CMOS LNA architectures . . . . . . . . . . . . . . . . . . . . 41

IV Calculated and simulated design parameters . . . . . . . . . . . . . . . 51

V Targeted specifications and simulated performance . . . . . . . . . . . 51

VI Mixer specifications for low-IF Bluetooth receiver . . . . . . . . . . . 69

VII Bluetooth mixer simulation results . . . . . . . . . . . . . . . . . . . 77

VIII Bluetooth and Wi-Fi standards key features . . . . . . . . . . . . . . 81

IX Block specifications for the BT/Wi-Fi RF front-end . . . . . . . . . . 87

X Chameleon LNA simulation results . . . . . . . . . . . . . . . . . . . . 96

XI Chameleon Mixer simulation results . . . . . . . . . . . . . . . . . . . 99

XII RF front-end nominal biasing condition . . . . . . . . . . . . . . . . . 102

XIII RF front-end measurement results . . . . . . . . . . . . . . . . . . . . 110

XIV Device noise contribution ratios of single-ended hybrid LNA († signal generator's internal resistance) . . . . . . . . . . . . . . . . . . 141

XV $f_T$ of a $2 \times 2$ NPN transistor . . . . . . . . . . . . . . . . . . . . . . . 142

XVI Noise contribution ratios of differential configuration († signal generator's internal resistance) . . . . . . . . . . . . . . . . . . . . . . . . 149

XVII Comparison of the proposed linearization implementation with the state-of-the-art linear LNA's in the literature . . . . . . . . . . . 153

TABLE Page

XVIII Reported LNA performance for GSM applications . . . . . . . . . . . 169

XIX Reported LNA performance in for cellular applications († indicates the value is for the whole front-end) . . . . . . . . . . . . . . . . 172

XX Basic properties of lossless and low loss T-line . . . . . . . . . . . . 182

XXI Simulation results of lumped-matched wide-band LNA . . . . . . . . . 206

XXII Volterra series versus Taylor series . . . . . . . . . . . . . . . . . . . . 231

XXIII Volterra kernels of degenerated BJT . . . . . . . . . . . . . . . . . . . 235

# LIST OF FIGURES

FIGURE Page

1 A receiver RF front-end . . . . . . . . . . . . . . . . . . . . . . . . . . 1

2 Two-port network showing incident waves ($a_1$, $a_2$) and reflected waves ($b_1$, $b_2$) used in S-parameter definiations . . . . . . . . . . . . . 7

3 Single-stage RF amplifier block diagram . . . . . . . . . . . . . . . . . 9

4 Noisy two-port network representations: (a) z-parameters, (b) y-parameters and (c) ABCD-parameters . . . . . . . . . . . . . . . . . . 15

5 Noise factor calculation . . . . . . . . . . . . . . . . . . . . . . . . . . 17

6 MOS transistor thermal noise model (a) gate and drain noise sources (b) input-referred noise sources . . . . . . . . . . . . . . . . 21

7 (a) In-band blockers generate in-channel intermodulation term (b) The two-tone test spectrum . . . . . . . . . . . . . . . . . . . . . . . . 28

8 1-dB compression point . . . . . . . . . . . . . . . . . . . . . . . . . . 29

9 Meaning of IP2 and IP3 . . . . . . . . . . . . . . . . . . . . . . . . . . 30

10 (a) SFDR defination (b) Relationship between SFDR and IIP3 . . . 31

11 Compression-free dynamic range . . . . . . . . . . . . . . . . . . . . . 32

12 Triple-order beats . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

13 Popular single-ended CMOS LNA topologies (a) Resistive termination (b) Common gate (c) Shunt-series feedback (d) Inductive source-degeneration (e) Current-reuse (f) $C_{gd}$ neutralization . . . . . 37

14 Inductive source degenerated LNA . . . . . . . . . . . . . . . . . . . . 42

15 Flow chart of LNA design procedure . . . . . . . . . . . . . . . . . . . 43

FIGURE Page


16 $\alpha$, $V_{gs} - V_{th}$, $g_m/W$, and $g_{do}/W$ versus drain current density . . . . . 44

17 $C_{gs}/W$ and $f_T$ versus gate over drive voltage $V_{gs} - V_{th}$ . . . . . . . . 45

18 Noise factor scaling coefficient versus quality and current density for $0.18\mu m$ NMOS device (a) 3-D plot (b) 2-D plot . . . . . . . . . . 47

19 IIP3 versus gate overdrive and device size . . . . . . . . . . . . . . . . 48

20 Simulation plots of the designed LNA (a) Voltage gain, S11 and noise figure (b) IIP3 and $P_{1dB}$ . . . . . . . . . . . . . . . . . . . . . . 52

21 Mixers perform frequency translation in communication system . . . 54

22 Mixer mathematical model . . . . . . . . . . . . . . . . . . . . . . . . 54

23 Mixing process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

24 Double-side band conversion . . . . . . . . . . . . . . . . . . . . . . . 57

25 Single-side band conversion . . . . . . . . . . . . . . . . . . . . . . . 58

26 Single-side band conversion with image rejection . . . . . . . . . . . 58

27 LO-to-RF leakage . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

28 Single-diode mixer . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61

29 Single-balanced diode mixer . . . . . . . . . . . . . . . . . . . . . . . 62

30 Double-balanced diode mixer . . . . . . . . . . . . . . . . . . . . . . 62

31 CMOS passive double-balanced mixer . . . . . . . . . . . . . . . . . 63

32 Gilbert-cell mixer (a) transistor implementation (b) working principle and (c) single-ended version . . . . . . . . . . . . . . . . . . . 64

33 Sub-sampling mixer . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

34 Harmonic mixer . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67

35 The proposed low-IF Bluetooth receiver . . . . . . . . . . . . . . . . 70

FIGURE Page


36 Bluetooth receiver down-conversion mixer . . . . . . . . . . . . . . . . 71

37 Mixer design flow chart . . . . . . . . . . . . . . . . . . . . . . . . 75

38 Layout of the down-conversion mixers . . . . . . . . . . . . . . . . . . 76

39 Die photo of Bluetooth receiver . . . . . . . . . . . . . . . . . . . . . 78

40 Input return loss of the Bluetooth front-end . . . . . . . . . . . . . . 79

41 Measured IIP3 of the Bluetooth receiver . . . . . . . . . . . . . . . . 79

42 Bluetooth/Wi-Fi receiver . . . . . . . . . . . . . . . . . . . . . . . 83

43 Differential LNA common mode stability . . . . . . . . . . . . . . . . . 85

44 RF front-end block diagram of BT/Wi-Fi Receiver . . . . . . . . . . 86

45 LNA for BT/Wi-Fi receiver . . . . . . . . . . . . . . . . . . . . . . 88

46 Noise optimization . . . . . . . . . . . . . . . . . . . . . . . . . . 92

47 LNA layout . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94

48 Coupling factor between two shifted spirals . . . . . . . . . . . . . . 95

49 Mixer for BT/Wi-Fi receiver . . . . . . . . . . . . . . . . . . . . . 97

50 Waveforms at switching pair common emitters . . . . . . . . . . . . 98

51 Mixer layout . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 99

52 Bias circuit for the RF front-end . . . . . . . . . . . . . . . . . . . 101

53 Die photo of the front-end . . . . . . . . . . . . . . . . . . . . . . . 103

54 Substrate noise isolation by deep trenches and guard rings . . . . . . 103

55 RF front-end test board . . . . . . . . . . . . . . . . . . . . . . . . 104

56 Testing setup for input match . . . . . . . . . . . . . . . . . . . . . 105

57 Input matching for high gain mode . . . . . . . . . . . . . . . . . . . 105

FIGURE Page

58 Input matching for low gain mode . . . . . . . . . . . . . . . . . . . . 106
59 Testing setup for IIP3 and IIP2 measurement . . . . . . . . . . . . . 107
60 IIP3 plot for 2-tone test at 12MHz and 25MHz offset . . . . . . . . . 107
61 IIP2 plot for 2-tone test at 12.2MHz and 12.8MHz offset . . . . . . . 108
62 I/Q mismatch measurement . . . . . . . . . . . . . . . . . . . . . . . 108
63 I-Q mismatch performance . . . . . . . . . . . . . . . . . . . . . . . 109
64 Testing setup for noise figure and conversion gain . . . . . . . . . . . 110
65 Frequency components in two-tone test . . . . . . . . . . . . . . . . . 114
66 Second-order distortion contributing to IM3 terms . . . . . . . . . . 118
67 Load non-linearity large signal model . . . . . . . . . . . . . . . . . . 120
68 Inductive degenerated CMOS LNA . . . . . . . . . . . . . . . . . . . 123
69 MOS transistor 2nd-order and 3rd-order distortion terms . . . . . . . 129
70 Multi-gated linearization using two NMOS transistors . . . . . . . . 130
71 NMOS multi-gated transistor linearity v.s. auxiliary bias voltage . . 130
72 Multi-gated linearization using NMOS and PMOS . . . . . . . . . . 131
73 Complementary multi-gated transistor linearity plot . . . . . . . . . 131
74 Bipolar as auxiliary transistor for 3rd-order linearization . . . . . . . 134
75 Emitter-degenerated BJT large signal model . . . . . . . . . . . . . . 135
76 Resistive-degenerated BJT 3rd-order coefficient at DC and 3GHz . . 138
77 3rd-order terms of the BJT, NMOS and their combination . . . . . . 139
78 Small signal circuit for input impedance calculation . . . . . . . . . . 140
79 Effect of bipolar transistor on input matching . . . . . . . . . . . . . 140

FIGURE Page

80 IIP3 of NMOS-NPN combination vs. bias conditions . . . . . . . . . . 143

81 IIP3 against process corners . . . . . . . . . . . . . . . . . . . . . . . 143

82 IIP3 temperature behavior . . . . . . . . . . . . . . . . . . . . . . . . 144

83 NPN transistor optimal biasing profile . . . . . . . . . . . . . . . . . 144

84 Mixer design flow chart . . . . . . . . . . . . . . . . . . . . . . . . . 146

85 Linearized differential LNA using bipolar differential pair . . . . . . . 147

86 (a) 2nd-order input range expansion and (b) 3rd-order cancellation of the proposed differential LNA . . . . . . . . . . . . . . . . . . 148

87 Die photomicrographs of hybrid linearized LNA's (a) single-ended (b) differential . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 150

88 IIP3 of the single-ended bipolar linearized LNA . . . . . . . . . . . . 151

89 Linearized single-ended LNA S-parameters . . . . . . . . . . . . . . . 151

90 IIP3 of differential LNA with and without bipolar cancellation pair activated . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 152

91 Linearized differential LNA S-parameters . . . . . . . . . . . . . . . . 152

92 Mobile station receiver band . . . . . . . . . . . . . . . . . . . . . . . 156

93 Matching utilizing mutual inductance . . . . . . . . . . . . . . . . . . 157

94 Input impedance small signal circuit . . . . . . . . . . . . . . . . . . 158

95 Inter-stage impedance . . . . . . . . . . . . . . . . . . . . . . . . . . 160

96 Overall transconductance . . . . . . . . . . . . . . . . . . . . . . . . 162

97 Equivalent input noise sources of the inductive coupled LNA . . . . . 163

98 The proposed mutual-coupled degenerated LNA . . . . . . . . . . . . 165

99 Design flow of the mutual-coupled LNA . . . . . . . . . . . . . . . . 166

FIGURE Page


100 Die microphotograph of the mutual-coupled degenerated LNA . . . . 167

101 Measured small signal performance of the mutual-coupled degenerated LNA . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 168

102 Measured IIP3 of the mutual-coupled degenerated LNA . . . . . . . 169

103 Dual-band inductive coupled LNA (bias not shown) . . . . . . . . . . 171

104 Simulated small signal performance of the dual-band LNA (a) S11 and S21 (b) noise figure and minimum noise figure . . . . . . . . . . 173

105 Simulated IIP3 of the dual-band LNA (a) GSM-900 (b) DCS-1800 . . 174

106 UWB signal spectrum . . . . . . . . . . . . . . . . . . . . . . . . . . 176

107 Direct conversion UWB receiver . . . . . . . . . . . . . . . . . . . . . 177

108 Distributed model of transmission line . . . . . . . . . . . . . . . . . 179

109 Coplanar stripline formed by metal 6 (dimensional drawing) . . . . . 184

110 Coplanar stripline rendered by HFSS . . . . . . . . . . . . . . . . . . 185

111 Coplanar stripline operation mode (a) Even mode (b) Odd mode . . 185

112 Micro-stripline HFSS render . . . . . . . . . . . . . . . . . . . . . . . 186

113 Field distribution of micro-stripline . . . . . . . . . . . . . . . . . . . 187

114 $Z_c$ of (a) coplanar stripline and (b) micro-stripline at 30 GHz . . . . 188

115 Schematic representation of a distributed LNA . . . . . . . . . . . . 189

116 Unit section of a distributed LNA . . . . . . . . . . . . . . . . . . . . 190

117 Layout of the five-section distributed LNA . . . . . . . . . . . . . . 192

118 S-parameters of the five-section distributed LNA . . . . . . . . . . . 192

119 Noise figure of the five-section distributed LNA . . . . . . . . . . . . 193

120 Noise model of a distributed LNA unit section . . . . . . . . . . . . 194

FIGURE Page

121 A distributed mixer block diagram . . . . . . . . . . . . . . . . . . . . 199

122 Wide band matching procedure using lumped components . . . . . . 203

123 Wide-band LNA with lumped-matched network . . . . . . . . . . . 205

124 Wide-band LNA S-parameters and noise figure . . . . . . . . . . . . 206

125 Circuit model of a non-linear time-invariant system . . . . . . . . . . 225

126 Second and third order response block diagrams . . . . . . . . . . . . 226

127 A MOS differential pair for Volterra analysis . . . . . . . . . . . . . . 227

# CHAPTER I

# INTRODUCTION

It is generally known that a typical receiver consists of two major parts: i) the RF front-end, which performs small signal amplification and frequency down conversion, and ii) a base-band portion, which performs demodulation and generates the required digital control to the system. Although there is no specific definition for the RF front-end, it usually includes a low noise amplifier (LNA), image rejection filter (IRF), down conversion mixer, local oscillator (LO), frequency synthesizer, intermediate frequency (IF) filter, and other IF amplifiers. Another understanding of the RF front-end is that it only ends at the mixer. Fig. 1 is the block diagram of a super-heterodyne receiver RF front-end. The functions of each block is elaborated below.

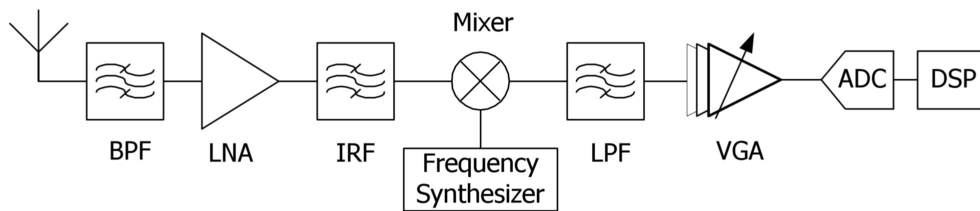



Fig. 1. A receiver RF front-end

The LNA is a very important block of the whole receiver. It interfaces directly with the antenna (usually there is a passive RF band-pass filter between the antenna and the LNA, the function of this filter is to provide band selection). The LNA must be able to provide enough power amplification and introduce small additional

---

The style and format follow *IEEE Journal of Solid-State Circuits.*

amounts of noise to the system. It also should be linear enough to tolerate high power level interferences coming from the wireless channel.

For a heterodyne receiver, the image rejection filter follows the LNA to remove the image frequency component from the band. The image of the signal resides on the other side of the local oscillator frequency, if not removed or attenuated somehow, it will fold into the IF band with the signal, and degrade the signal to noise and distortion ratio. For a zero-IF receiver architecture, there is no image problem, so the image rejection filter is not needed. For a low-IF receiver, which usually has a relatively low intermediate frequency (several MHz), the image is rejected after the mixer by using complex filtering.

It is usually difficult and uneconomical to process the received signal directly at RF frequencies (several hundred MHz to several GHz), so the mixer is another important block in the receiver. It converts the signal from a high frequency to a low frequency to make the signal process easier and more effective. Because the mixer operates among three frequencies (RF, LO and IF), it must be able to provide enough isolation between them, deliver enough conversion gain, and keep reasonable dynamic range.

The local oscillator and frequency synthesizer provide a stable and clean programmable LO signal to the mixer for channel selection. The IF filters and amplifiers are used to further filter out the unwanted signal, and provide the proper signal level to the base-band analog-to-digital converter (ADC). The IF amplifier is usually a variable gain amplifier (VGA), adjusting its gain according to the received signal strength such that the ADC always sees the optimum input signal level.

## A. Research Motivation

The rapid development in the wireless technology introduces new design issues and challenges, such as the low power consumption, high speed, low cost, small form factor and multi-standard programmability. This work is focused on the design of front-end circuits: LNA and mixer for both narrow band and broad band applications. The LNA and mixer are the two blocks in the signal path operating at the highest signal frequency. They see all the interferences and noise coming from the wireless channel. The quality of the front-end circuits directly affect the performance of the whole system. The challenge of LNA design is to implement it by meeting the voltage gain, noise, linearity, silicon area and power consumption at the same time. The mixer design in low-IF and direction conversion architecture requires mainly low noise and high linearity. Also the non-linear switching behavior of mixer's current commutating pair make the design of mixer more difficult. The emphasis will be put on the system integration of the LNA and mixer in Bluetooth and IEEE 802.11b receivers, LNA linearization technique using bipolar transistor, LNA input matching network design using on-chip transformer and possible wide-band LNA implementations. The next section gives a more detailed overview of the organization of the dissertation.

## B. Dissertation Overview

The whole dissertation is organized as follows. Chapter I discusses the research motivations, and the outline skeleton of the work. In Chapter II and Chapter III, an overview of the LNA and mixer design issues and techniques will be given. Based on the current bleeding/injection technique, the later sections of Chapter III introduces a mixer implementation for a low-IF Bluetooth receiver. Chapter IV is dedicated to the RF front-end implementation of a direct conversion Bluetooth/WiFi dual-mode

receiver. The LNA and mixer design trade-offs and considerations are elaborated in this chapter. Chapter V and Chapter VI give novel LNA design techniques for linearity and multi-band operation respectively. Ultra-wide band implementations of front-end blocks is tackled in Chapter VII. Chapter VIII concludes the work of this dissertation.

# CHAPTER II

# LOW NOISE AMPLIFIER DESIGN OVERVIEW

Low noise amplifier (LNA) is the first gain stage encountered in a receiver environment either wired or wireless. It must meet several specifications at the same time, which make its design really challenging. Signal coming from the receiver antenna is very small, usually from -100 dBm (3.2 $\mu V$) to -70 dBm (0.1 $mV$)[1], therefore signal amplification is needed for the following stage (mixer) to handle. This sets the requirement of a certain gain to the LNA. The received signal should have a certain signal-to-noise ratio (SNR) in order to be reliably detected, therefore, noise added by the circuit should be reduced as much as possible, which will set the noise requirement of the LNA. Large signal or blocker can occur at the input of LNA. The circuits should be sufficiently linear in order to have a reasonable signal reception. For portable and mobile applications, reasonable power consumption is another constraint.

## A. Basics on S-parameters

In design and analysis, scattering parameters, which are commonly referred to as S-parameters, are widely used in microwave and RF circuits. S-parameters use a parameter set that relates to the traveling waves that are scattered or reflected when an n-port network is inserted into a transmission line. S-parameter analysis is basi-

[1]The unit dBm is a power unit referred to 1 mW in dB scale. If a signal's power is P Watt, then in dBm, it is $10\log(\mathrm{P}/1\mathrm{mV})$ dBm. When related to a voltage, there is a reference impedance R involved. This impedance is usually $50\,\Omega$. When we say 0.1 mV is corresponding to -70 dBm, what we mean is that the power dissipated into a $50\,\Omega$ resistor by a sinusoidal voltage signal which has a peak value of 0.1 mV is $10^{-7}$ mW, and corresponds to -70 dBm.

cally a modeling method to characterize an n-port linear network. There are other methods to characterize the network, such as H-parameters, Y-parameters and Z-parameters. All of them fall into the same modeling category for a network together with S-parameters. They are behavioral modeling methods. The network or device is treated as a black box, only the interaction between the ports and outer environment is modeled.

For low frequencies, H-, Y-, or Z-parameters are more widely used. They use port voltage and current as variables. But when the frequency moves higher and higher, some problems arise. A bottleneck requirement for H, Y or Z measurement is to apply short and/or open circuit condition at each port. This can be hard to do, especially at RF frequencies where lead inductance and capacitance make short and open circuits difficult to obtain. Active devices, such as transistors and tunnel diode, very often can not be connected in stable short or open circuit conditions. S-parameters, on the other hand, are usually measured with the device imbedded between a $50\,\Omega$ load and source, and there is very little chance for oscillations to occur. Another important advantage of S-parameters stems from the fact that traveling waves, unlike terminal voltages and currents, do not vary in magnitude at points along a lossless transmission line. This means that scattering parameters can be measured on a device located at some distance from the measurement transducers, provided that the measuring device and the transducers are connected by low-loss transmission lines.

The linear equations describing the behavior of the two-port network in Fig. 2 using S-parameters are

$$b_1 = s_{11}a_1 + s_{12}a_2 \tag{2.1}$$

$$b_2 = s_{21}a_1 + s_{22}a_2 \tag{2.2}$$

where $b_1$, $b_2$, $a_1$ and $a_2$ are traveling waves and have dimensions of $\sqrt{Watts}$. The

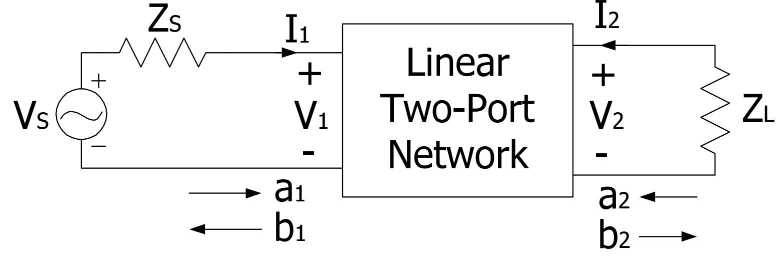



Fig. 2. Two-port network showing incident waves ($a_1$, $a_2$) and reflected waves ($b_1$, $b_2$) used in S-parameter definiations

S-parameters $s_{11}$, $s_{22}$, $s_{21}$ and $s_{12}$ are defined by:

$$s_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0} \tag{2.3}$$

$$s_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0} \tag{2.4}$$

$$s_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0} \tag{2.5}$$

$$s_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0} \tag{2.6}$$

For most measurements and calculations, it is convenient to assume the port reference impedances are positive and real. The two ports can have different reference impedances, but the same reference impedance $Z_0$ will the used for all the ports here.

The independent variables $a_1$ and $a_2$ can be related to port voltages ($V_1$, $V_2$) and currents ($I_1$, $I_2$) as follows:

$$a_1 = \frac{V_1 + I_1 Z_0}{2\sqrt{Z_0}} = \frac{V_{i1}}{\sqrt{Z_0}} \tag{2.7}$$

$$a_2 = \frac{V_2 + I_2 Z_0}{2\sqrt{Z_0}} = \frac{V_{i2}}{\sqrt{Z_0}} \tag{2.8}$$

where $V_{i1} = \frac{V_1+I_1Z_0}{2}$ and $V_{i2} = \frac{V_2+I_2Z_0}{2}$ are the voltage waves incident on port 1 and port 2 respectively. Therefore, $|a_1|^2$ is the power incident on the input of the network,

and is also the power available from a source impedance $Z_0$. $|a_2|^2$ is the power incident on the output of the network, and is also the power reflected from the load.

Similarly, the dependent variables $b_1$ and $b_2$ can be related to port voltages and currents as follows:

$$b_1 = \frac{V_1 - I_1 Z_0}{2\sqrt{Z_0}} = \frac{V_{r1}}{\sqrt{Z_0}} \tag{2.9}$$

$$b_2 = \frac{V_2 - I_2 Z_0}{2\sqrt{Z_0}} = \frac{V_{r2}}{\sqrt{Z_0}} \tag{2.10}$$

where $V_{r1} = \frac{V_1 - I_1 Z_0}{2}$ and $V_{r2} = \frac{V_2 - I_2 Z_0}{2}$ are the voltage waves reflected from port 1 and port 2 respectively. Therefore, $|b_1|^2$ is the power reflected from the input port of the network, or the power available from a $Z_0$ source minus the power delivered to the input of the network. $|b_2|^2$ is the power reflected from the output port of the network, or the power incident on the load, which is also the power that would be delivered to a $Z_0$ load.

From the above explanation of $a_1$, $a_2$ and $b_1$, $b_2$, the four S-parameters are simply related to power gain and mismatch loss:

$$|s_{11}|^2 = \frac{\text{Power reflected from the network input}}{\text{Power incident on the network input}} \tag{2.11}$$

$$|s_{22}|^2 = \frac{\text{Power reflected from the network output}}{\text{Power incident on the network output}} \tag{2.12}$$

$$\begin{aligned} |s_{21}|^2 &= \frac{\text{Power delivered to } z_0 \text{ load}}{\text{Power available from } z_0 \text{ source}} \\ &= \text{Transducer power gain with } Z_0 \text{ load and source} \end{aligned} \tag{2.13}$$

$$|s_{12}|^2 = \text{Reverse transducer power gain with } Z_0 \text{ load and source} \tag{2.14}$$

It needs to be noticed that all of these two-port parameters are equivalent, because they describe the same network property. However, for all the practical purposes, S-parameters are easier for the LNA design and characterization, while under certain scenarios, Y-, H-, or Z-parameters maybe more straightforward. Observe that

not all the networks (e.g. ideal transformer, a serials impedance or a shunt admittance) have Y- or Z-parameters, but they will have S-parameter characterization.

B. Amplifier's Gain and Stability

Gain performance for a RF amplifier is determined by the RF transistor itself and the input/output matching network. Fig. 3 shows a single-stage amplifier block diagram. The amplifier is characterized by its S-parameters and terminated by arbitrary source and load impedance $Z_s$ and $Z_L$. $s_{11}$ and $s_{22}$ are the input and output reflection coefficients with $Z_0$ source and load termination. For an arbitrary impedance termination, the input and output reflection coefficient $\Gamma_{in}$ and $\Gamma_{out}$ for a two-port network [1] can be found to be

$$\Gamma_{in} = \frac{b_1}{a_1} = s_{11} + \frac{s_{12}s_{21}\Gamma_L}{1 - s_{22}\Gamma_L} \tag{2.15}$$

$$\Gamma_{out} = \frac{b_2}{a_2} = s_{22} + \frac{s_{12}s_{21}\Gamma_s}{1 - s_{11}\Gamma_s} \tag{2.16}$$

where $\Gamma_s = \frac{Z_s - Z_0}{Z_s + Z_0}$ and $\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0}$ are the source and load reflection coefficient respectively.

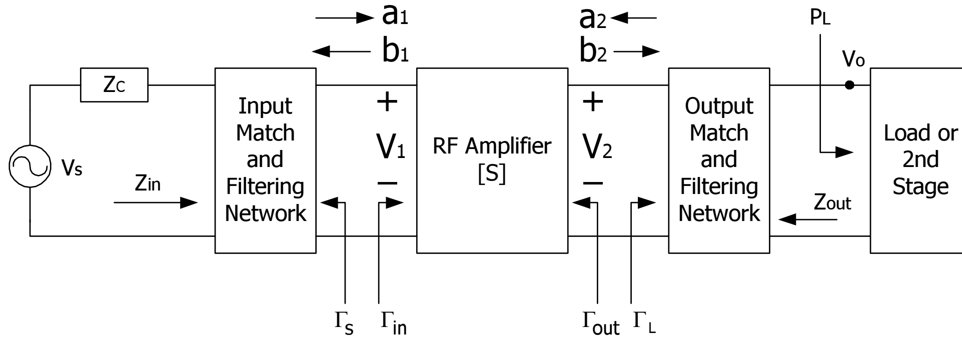



Fig. 3. Single-stage RF amplifier block diagram

If the input and output are simultaneously complex conjugate matched, i.e. $\Gamma_{in} =$

$\Gamma_s^*$ and $\Gamma_{out} = \Gamma_L^*$, the amplifier has maximum power transfer. Usually it is not easy to achieve the simultaneous complex conjugate matching condition. A special case is for an unilateral device, $s_{12}$ is practically zero, then $\Gamma_{in} = s_{11}$ and $\Gamma_{out} = s_{22}$. The input and output are decoupled from each other, matching can be done at the input and output separately.

Several gain definitions exist for an amplifier. Power gain ($G$) characterizes the actual power amplification of an amplifier. It is defined as the power delivered to the load divided by the power input to the network. Available power gain ($G_A$) shows the maximum possible power amplification of the amplifier. For IC implementations, LNA input is interfaced off-chip and usually matched to specific impedance ($50\,\Omega$ or $75\,\Omega$). Its output is not necessarily matched if directly driving on-chip blocks such as mixers. This situation is usually characterized by voltage gain or transducer power gain by knowing the load impedance level.

The voltage gain ($A_V$) is defined as the voltage at the output port divided by the voltage at the input port of the amplifier. Relating to the s-parameters of the amplifier

$$A_V = \frac{V_2}{V_1} = \frac{s_{21}\,(1+\Gamma_L)}{(1-s_{22}\Gamma_L)\,(1+\Gamma_{in})} \tag{2.17}$$

The transducer power ($G_T$) [1] is defined as the power delivered to the load divided by the power available from the source:

$$G_T = \frac{P_L}{P_{AVS}} \tag{2.18}$$

where $P_L$ equals the power incident on load minus the power reflected from load:

$$P_L = |b_2|^2\left(1-|\Gamma_L|^2\right) \tag{2.19}$$

and $P_{AVS}$ is

$$P_{AVS} = \frac{|b_s|^2}{1 - |\Gamma_s|^2} \tag{2.20}$$

where $b_s = \frac{V_s\sqrt{Z_0}}{Z_S+Z_0}$. Therefore

$$G_T = \left|\frac{b_2}{b_s}\right|^2 \left(1 - |\Gamma_s|^2\right)\left(1 - |\Gamma_L|^2\right) \tag{2.21}$$

Using the signal flow chart, the ratio $\frac{b_2}{b_s}$ can be found to be

$$\frac{b_2}{b_s} = \frac{s_{21}}{(1 - s_{11}\Gamma_s)(1 - s_{22}\Gamma_L) - s_{12}s_{21}\Gamma_s\Gamma_L} \tag{2.22}$$

Finally, the transducer power gain expression is

$$G_T = \frac{|s_{21}|^2 \left(1 - |\Gamma_s|^2\right)\left(1 - |\Gamma_L|^2\right)}{|(1 - s_{11}\Gamma_s)(1 - s_{22}\Gamma_L) - s_{12}s_{21}\Gamma_s\Gamma_L|^2} \tag{2.23}$$

By using cascode or neutralization technique, a network can be treated as unilateral, i.e., $s_{12}$ is small and effectively zero, (2.23) will be reduced to

$$\begin{aligned} G_{Tu} = G_T|_{s_{12}=0} &= \tfrac{1-|\Gamma_s|^2}{|1-s_{11}\Gamma_s|^2}\,|s_{21}|^2\,\tfrac{1-|\Gamma_L|^2}{|1-s_{22}\Gamma_L|^2} \\ &= G_S\,|s_{21}|^2\,G_L \end{aligned} \tag{2.24}$$

where $G_S$ and $G_L$ are the source and load mismatch factor respectively:

$$G_S = \frac{1 - |\Gamma_s|^2}{|1 - s_{11}\Gamma_s|^2} \tag{2.25}$$

$$G_L = \frac{1 - |\Gamma_L|^2}{|1 - s_{22}\Gamma_L|^2} \tag{2.26}$$

Once the device and its bias condition is established, $s_{21}$ remains unchanged. $G_S$ is only related to the input parameters $s_{11}$ and $\Gamma_s$. $G_S$ shows the degree of mismatch between the source impedance and the input impedance. Similarly, $M_L$ is only related to the output parameters $s_{22}$ and $\Gamma_L$ and shows the matching condition at the output.

When $\Gamma_s$ equals $s_{11}$'s complex conjugate $s_{11}^*$, $G_S$ reaches its maximum value of

$G_{S,\max} = \frac{1}{1-|s_{11}|^2}$. This can be verified by writing $s_{11} = a + jb$ and $\Gamma_s = x + jy$, substituting them into (2.25) and solve equations $\frac{\partial G_S}{\partial x} = 0$ and $\frac{\partial G_S}{\partial y} = 0$ for x and y. The same result can be obtain for $\Gamma_L$: when $\Gamma_L = s_{22}^*$, $G_L$ reaches its maximum value of $G_{L,\max} = \frac{1}{1-|s_{22}|^2}$.

It is also easy to see that when $|\Gamma_s| = 1$, $G_S$ has a minimum value of zero and when$|\Gamma_L| = 1$, $G_L$ has a minimum value of zero. For other values of $G_S$ or $G_L$, the corresponding $\Gamma_s$ or $\Gamma_L$ lie on a circle in the Smith Chart. Different values of $G_S$ or $G_L$ have different circles. These circles are usually referred to as constant gain circles. They have their centers located on the line drawn from the center of the Smith Chart to the point of $s_{11}^*$ or $s_{22}^*$. Specifically, the center of the circles resides at

$$c_i = \frac{g_i s_{ii}^*}{1 - |s_{ii}|^2 (1 - g_i)} \tag{2.27}$$

and the radius of the circles is

$$r_i = \frac{\sqrt{1 - g_i}\left(1 - |s_{ii}|^2\right)}{1 - |s_{ii}|^2 (1 - g_i)} \tag{2.28}$$

where $g_i = G_i\left(1 - |s_{ii}|^2\right) = \frac{G_i}{G_{i,\max}}$ is the normalized gain value for the gain circle $G_i$. The subscript $i$ represents $S$ for input and $L$ for output, $ii$ represents 11 for input and 22 for output. The detailed proof of the constant gain circle is given in Appendix A.

The stability of RF amplifiers is also a very important metric. Only when the amplifier is stable, the other metrics such as gain, noise figure are meaningful. Suppose the input impedance at the input port of the amplifier is $Z_i = R_i + jX_i$, then the input reflection coefficient module is

$$|\Gamma_{in}| = \left|\frac{Z_i - Z_0}{Z_i + Z_0}\right| = \sqrt{\frac{(R_i - Z_0)^2 + X_i^2}{(R_i + Z_0)^2 + X_i^2}} \tag{2.29}$$

From (2.29) it is easy to see that if the real part of the input impedance is negative,

i.e. if $R_i < 0$, then $|\Gamma_{in}| > 1$. If this negative impedance compensates the loss coming from the input termination network, oscillation can occur. The amplifier is potentially unstable. The same argument holds for the output. If for all the passive terminations at the input and output, the following conditions hold, then the amplifier is unconditionally stable. Otherwise, it is potentially unstable or conditionally stable.

$$|\Gamma_{in}| < 1; \ |\Gamma_{out}| < 1 \tag{2.30}$$

A meaning for (2.30) is that the real parts of input and output impedance of the amplifier are resistive, or, in Smith Chart, $\Gamma_{in}$ and $\Gamma_{out}$ will never go out of the unity circle.

It can be shown that the conditions for unconditionally stable [1] in term of s-parameters are

$$\begin{aligned} |s_{11}| &< 1 \\ |s_{22}| &< 1 \\ K &> 1 \end{aligned} \tag{2.31}$$

where K is the stability factor given by

$$K = \frac{1 - |s_{11}|^2 - |s_{22}|^2 + |s_{11}s_{22} - s_{12}s_{21}|^2}{2\,|s_{12}s_{21}|} \tag{2.32}$$

For example, an amplifier has the following S-parameters at 1250 MHz:

$$s_{11} = 0.38\angle-115^\circ,\ s_{12} = 0.06\angle 14^\circ,\ s_{21} = 6.0\angle 104^\circ,\ s_{22} = 0.50\angle-52^\circ \tag{2.33}$$

It can be calculated from the above data that $|s_{11}| = 0.38 < 1$, $|s_{22}| = 0.5 < 1$, and $K = 1.02 > 1$. So the amplifier is unconditionally stable. While at 750 MHz, the same amplifier has the following S-parameter:

$$s_{11} = 0.56\angle-78^\circ,\ s_{12} = 0.05\angle 33^\circ,\ s_{21} = 8.6\angle 122^\circ,\ s_{22} = 0.66\angle-42^\circ \tag{2.34}$$

In this case, $|s_{11}| = 0.56 < 1$, $|s_{22}| = 0.66 < 1$, but $K = 0.60 < 1$, So it is potentially unstable.

In order to stabilize an active device, one simple way is to add a series resistance or a shunt conductance to the unstable port. For example, the input port of the amplifier at 750 MHz mentioned above can be stabilized by series a 16.5 Ω resistor or shunt a 17.8 Ω resistor. Its output port can be stabilized with a 40 Ω series resistor or a 161 Ω shunt resistor. In practise, due to the coupling between the input and output ports of the amplifier, it is usually sufficient to just stabilize one of the ports, and one should avoid resistive loading of the input port, because it will cause additional noise to be amplified. Therefore, one will generally try to stabilize the output port. The prices paid for stabilizing using resistive loading are poor impedance matching, reduced power gain, and larger noise figure.

### C. Noise Performance

The noise performance of a RF amplifier is represented by its noise factor or noise figure. Noise factor shows the degradation of signal's signal-to-noise-ratio (SNR) due to the amplifier, it is defined as the SNR at the input of the network divided by the SNR at the output of the network:

$$F = \frac{SNR_{in}}{SNR_{out}} \tag{2.35}$$

where $SNR_{in}$ and $SNR_{out}$ are the SNR at the input and output of the amplifier respectively. The represent the signal's quality in terms of noise before and after the network. Noise figure is just the logarithm form of noise factor, its unit is dB:

$$NF(\text{dB}) = 10 \log F \tag{2.36}$$

## 1. Two-Port Network Noise Model

It will be pretty involved if one tries to use a transistor's equivalent noise circuit to analysis the whole amplifier or network. Using a two-port network noise model can simplify the calculation of its noise factor and also gain insight on minimize noise figure [2].

An effective way to analyze noise in a given circuit is to assume that the circuit is noiseless and model its internal noise by external noise sources at the input and output ports. These noise sources must have the same noise power appearing at the circuit's terminals as the original noisy circuit.

The noiseless network can be represented by its Z-, Y- or ABCD-parameters as shown in Fig. 4. In the following discussions, it is assumed that port 1 is the input port and port 2 is the output port.

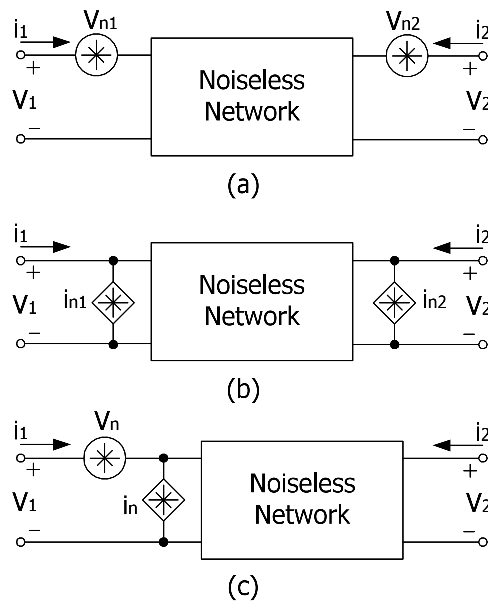



Fig. 4. Noisy two-port network representations: (a) z-parameters, (b) y-parameters and (c) ABCD-parameters

Using the Z-parameters in Fig. 4(a), the voltage-current relationship among ports can be written as

$$v_1 = z_{11}i_1 + z_{12}i_2 + v_{n1} \tag{2.37}$$

$$v_2 = z_{21}i_1 + z_{22}i_2 + v_{n2} \tag{2.38}$$

The equivalent noise source $v_{n1}$ and $v_{n2}$ can be measured from the open-circuited (o.c) measurements as

$$v_{n1} = v_1 \left.\right|_{i_1=i_2=0} \tag{2.39}$$

$$v_{n2} = v_2 \left.\right|_{i_1=i_2=0} \tag{2.40}$$

Using the Y-parameters in Fig. 4(b), the voltage-current relationship among ports can be written as

$$i_1 = y_{11}v_1 + y_{12}v_2 + i_{n1} \tag{2.41}$$

$$i_2 = y_{21}v_1 + y_{22}v_2 + i_{n2} \tag{2.42}$$

The equivalent noise source $i_{n1}$ and $i_{n2}$ can be obtained from the short-circuited (s.c.) measurements as

$$i_{n1} = i_1 \left.\right|_{v_1=v_2=0} \tag{2.43}$$

$$i_{n2} = i_2 \left.\right|_{v_1=v_2=0} \tag{2.44}$$

Referring the noise sources to the input port is convenient for noise analysis. This leads to the ABCD-parameter representation in Fig. 4(c),

$$v_1 = Av_2 + B\left(-i_2\right) + v_n \tag{2.45}$$

$$i_1 = Cv_2 + D\left(-i_2\right) + i_n \tag{2.46}$$

The noise $v_n$ and $i_n$ cannot be measured using the o.c. or s.c. measurement technique, but they can be found as function of $v_{n1}$ and $v_{n2}$ or as function of $i_{n1}$ and $i_{n2}$ through

the following expressions:

$$v_n = v_{n1} - v_{n2}\left(\frac{z_{11}}{z_{21}}\right) \tag{2.47}$$

$$i_n = -\frac{v_{n2}}{z_{21}} \tag{2.48}$$

or

$$v_n = -\frac{i_{n2}}{y_{21}} \tag{2.49}$$

$$i_n = i_{n1} - i_{n2}\left(\frac{y_{11}}{y_{21}}\right) \tag{2.50}$$

Therefore, the noise factor of a two-port network can be calculated using the noise representation in Fig. 4(c). Consider a general network with a noise current source connected to its input port as shown in Fig. 5. It is assumed that the noise of the source $i_s$ and the noise of the network, $i_n$ and $v_n$, are uncorrelated but $v_n$ and $i_n$ may be correlated.

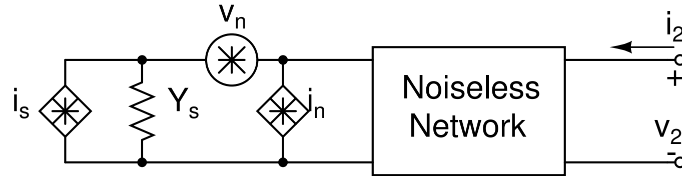



Fig. 5. Noise factor calculation

The total output noise power is proportional to the mean square value of the short-circuited current ($\overline{i_{s.c.}^2}$) at the input port of the noiseless network. The noise due to the source is only proportional to the mean square value of the current $\overline{i_s^2}$. The noise factor is thus given by:

$$F = \frac{\overline{i_{s.c.}^2}}{\overline{i_s^2}} = 1 + \frac{\overline{(i_n + v_n Y_s)^2}}{\overline{i_s^2}} \tag{2.51}$$

where $\overline{i_{s.c.}^2} = \overline{i_s^2} + \overline{(i_n + v_n Y_s)^2}$ and $Y_s = G_s + jB_s$ is the source admittance.

Typically there is correlation between $v_n$ and $i_n$. The gate and drain noise in MOS transistor is a good example. In this case, $i_n$ can be expressed as

$$i_n = i_{nu} + i_{nc} \tag{2.52}$$

where $i_{nu}$ is the part of $i_n$ which is ***u***ncorrelated with $v_n$, and $i_{nc}$ is the part of $i_n$ which is ***c***orrelated with $v_n$. $i_{nc}$ can be related to $v_n$ through the correlation admittance $Y_c$, and

$$i_{nc} = Y_c v_n \tag{2.53}$$

Note that $Y_c$ is not a physical admittance, it is called admittance only because it has the admittance dimension.

The noise factor can be rewritten using the above established equations as:

$$F = 1 + \frac{\overline{\left[i_{nu} + (Y_c + Y_s)\, v_n\right]^2}}{\overline{i_s^2}} = 1 + \frac{\overline{i_{nu}^2} + \left|Y_c + Y_s\right|^2 \overline{v_n^2}}{\overline{i_s^2}} \tag{2.54}$$

The noise source $i_s$ can be expressed in terms of the source conductance $G_s$,

$$\overline{i_s^2} = 4kTG_s\Delta f \tag{2.55}$$

The noise voltage $v_n$ can be expressed in terms of an equivalent (fictitious) noise resistance $R_n$,

$$\overline{v_n^2} = 4kTR_n\Delta f \tag{2.56}$$

The uncorrelated noise current $i_{nu}$ can be expressed in terms of an equivalent (fictitious) noise conductance $G_u$,

$$\overline{i_{nu}^2} = 4kTG_u\Delta f \tag{2.57}$$

Now the noise factor can be written in terms of noise parameters $R_n$, $G_s$ and $G_u$,

$$F = 1 + \frac{G_u}{G_s} + \frac{R_n}{G_s}\left[(G_s + G_c)^2 + (B_s + B_c)^2\right] \tag{2.58}$$

where $G_c$ and $B_c$ are the real and imaginary part of $Y_c$ respectively.

Note that $G_s$ and $B_s$ can be changed independently. So the noise factor can be first minimized by choosing

$$B_s = -B_c \equiv B_{opt} \tag{2.59}$$

Now, the second square term in the square brackets of (2.58) becomes zero, and F is reduced to

$$F = 1 + \frac{G_u}{G_s} + \frac{R_n}{G_s}\left(G_s + G_c\right)^2 \tag{2.60}$$

It can be further minimized by using a proper value for $G_s$, which can be found by differentiating the expression of F in respect to $G_s$ and equating the result to zero:

$$\frac{d}{dG_s}\left[1 + \frac{G_u}{G_s} + \frac{R_n}{G_s}\left(G_s + G_c\right)^2\right] = 0$$

Solving it for $G_s$, one can obtain,

$$G_s = \sqrt{G_c^2 + \frac{G_u}{R_n}} \equiv G_{opt} \tag{2.61}$$

The above $G_s$ and $B_s$ describe the optimum source admittance which would minimize the noise factor, and will be referred to as $G_{opt}$ and $B_{opt}$ respectively.

The minimum noise factor can be written as

$$F_{min} = F\mid_{Y_s=Y_{opt}} = 1 + 2R_n\left(G_{opt} + G_c\right) \tag{2.62}$$

The noise factor now can be written as

$$F = F_{min} + \frac{R_n}{G_s}\left|Y_s - Y_{opt}\right|^2 \tag{2.63}$$

By using normalized impedance and admittance, the noise factor can be expressed as

$$F = F_{min} + \frac{r_n}{g_s}\left|y_s - y_{opt}\right|^2 \tag{2.64}$$

where $r_n = \frac{R_n}{Z_o}$, $g_s = G_s Z_o$, $y_s = Y_s Z_o$ and $y_{opt} = Y_{opt} Z_o$.

From the definition of reflection coefficient, we have

$$y_s = \frac{1 - \Gamma_s}{1 + \Gamma_s}$$

$$y_{opt} = \frac{1 - \Gamma_{opt}}{1 + \Gamma_{opt}}$$

and

$$g_s = \frac{1}{2}(y_s + y_s^*) = \frac{1 - |\Gamma_s|^2}{1 + |\Gamma_s|^2 + (\Gamma_s + \Gamma_s^*)} = \frac{1 - |\Gamma_s|^2}{|1 + \Gamma_s|^2}$$

So the noise factor can be expressed in terms of reflection coefficients $\Gamma_s$ and $\Gamma_{opt}$,

$$F = F_{min} + \frac{4r_n |\Gamma_s - \Gamma_{opt}|^2}{\left(1 - |\Gamma_s|^2\right)|1 + \Gamma_{opt}|^2} \tag{2.65}$$

Now the concept of the constant noise circle will be introduced. For this purpose, rewrite (2.65) as

$$\frac{|\Gamma_s - \Gamma_{opt}|^2}{1 - |\Gamma_s|^2} = N \tag{2.66}$$

where $N = \frac{F - F_{min}}{4r_n} |1 + \Gamma_{opt}|^2$. Since $F \geq F_{min}$, then $N \geq 0$. For constant $F$, $N$ is also constant. It can be shown that the source reflection coefficient $\Gamma_s$ for constant noise factor $F$ is a circle in the Smith Chart $\Gamma_s$-plane (see Appendix A). Its radius and center location are given by

$$r_F = \frac{N}{1 + N}\sqrt{1 + \frac{1}{N}\left(1 - |\Gamma_{opt}|^2\right)} \tag{2.67}$$

and

$$c_F = \frac{\Gamma_{opt}}{1 + N} \tag{2.68}$$

respectively.

If $F = F_{min}$, then $N = 0$ and $c_F = \Gamma_{opt}$, $r_F = 0$, then circle reduces to a point which corresponds to the minimum noise factor. When $F$ increases, $N$ also increases,

moving the center $c_F$ toward the origin of the $\Gamma_s$-plane, and enlarge the radius of the circle. If $F \to \infty$, then $N \to \infty$, $c_F \to 0$ and $r_F \to 1$, this corresponds to the unit circle. It can be concluded that the outer circles have a larger noise factor than the inner ones.

## 2. MOS Transistor Two-Port Noise Parameters

The most popular CMOS LNA usually consists of just one stage utilizing one MOS transistor. Design insight can be gained by studying the two-port noise parameters $R_n$, $G_u$, $G_c$ and $B_c$ of a MOS transistor. From (2.62) and (2.63) , it is observed that the noise factor can be calculated by knowing these four noise parameters.

For the first order approximation, the drain and induced gate current noise dominate the noise performance of a MOS transistor. Fig. 6(a) shows the physical origin of a MOS transistor's noise sources.

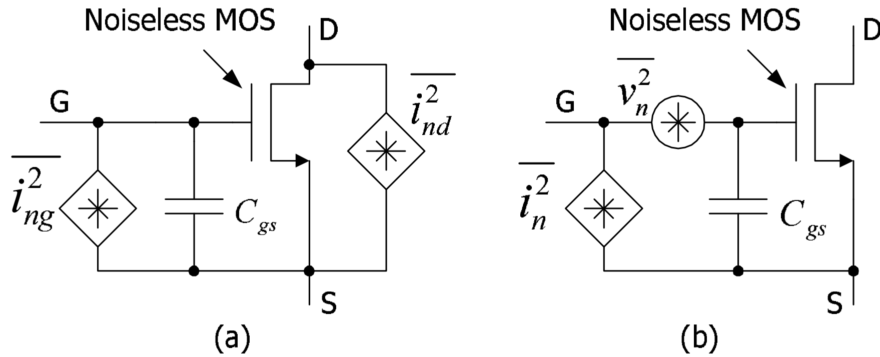



Fig. 6. MOS transistor thermal noise model (a) gate and drain noise sources (b) input-referred noise sources

The mean square value of the drain current noise is

$$\overline{i_{nd}^2} = 4kT\gamma g_{do}\Delta f \tag{2.69}$$

where $g_{do}$ is the drain source conductance at zero drain-source biasing. $\gamma$ is about 2/3 for long channel devices in the saturation region. For short channel devices, $\gamma$ is typically 2-3 or even larger [3]. This increased value of $\gamma$ is due to carrier heating by large electric fields developed across drain and source. Thus keeping the drain-source bias voltage as low as possible will reduce the value of $\gamma$.

The gate noise current mean square value is

$$\overline{i_{ng}^2} = 4kT\delta g_g \Delta f \tag{2.70}$$

where $\delta$ is the gate noise coefficient and is about twice the value of $\gamma$ for long channel devices and the parameter $g_g$ is

$$g_g = \frac{\omega^2 C_{gs}^2}{5g_{do}} \tag{2.71}$$

Both drain and gate noise are originated from the thermal agitation of channel charge, so they are correlated. The correlation coefficient [3] is defined by

$$c = \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2} \cdot \overline{i_{nd}^2}}} \tag{2.72}$$

For long channel devices its value is $j0.395$. The pure imaginary value implies that the correlation is due to capacitive coupling from channel to gate. So the gate noise current can be expressed as the sum of component $i_{ngc}$ which is fully correlated with drain current noise and component $i_{ngu}$ which is completely uncorrelated with drain current noise.

Using the method described in the previous section, refer all the noise sources in Fig. 6(a) to the input port (gate-source) of the MOS device as in Fig. 6(b). The noise voltage is

$$\overline{v_n^2} = \frac{4kT\gamma g_{do}\Delta f}{g_m^2} \tag{2.73}$$

and the noise current is

$$i_n = (i_{nc1} + i_{ngc}) + i_{ngu} \tag{2.74}$$

where $i_{nc1} = j\omega C_{gs} v_n$, $C_{gs}$ is the gate-source capacitance. $i_{nc1}$ is completely correlated with noise voltage $v_n$. It is easy to see that

$$i_{nc} = i_{nc1} + i_{ngc} \tag{2.75}$$

and

$$i_{nu} = i_{ngu} \tag{2.76}$$

Whereby if the gate noise is ignored, noise voltage and noise current will be fully correlated.

The four noise parameters [3] can be found to be

$$R_n = \frac{\gamma g_{do}}{g_m^2} = \frac{\gamma}{\alpha}\frac{1}{g_m} \tag{2.77}$$

$$Y_c = G_c + jB_c = j\omega C_{gs}\left(1 + \alpha\left|c\right|\sqrt{\frac{\delta}{5\gamma}}\right) \tag{2.78}$$

$$G_u = \frac{\delta\omega^2 C_{gs}^2\left(1 - \left|c\right|^2\right)}{5g_{do}} \tag{2.79}$$

where $\alpha = \frac{g_m}{g_{do}}$. It is seen that $G_c$, the real part of $Y_c$, is essentially zero. Thus the real and imaginary part of the optimum source admittance are

$$G_{opt} = \sqrt{\frac{G_u}{R_n}} = \alpha\omega C_{gs}\sqrt{\frac{\delta}{5\gamma}\left(1 - \left|c\right|^2\right)} \tag{2.80}$$

and

$$B_{opt} = -B_c = -\omega C_{gs}\left(1 + \alpha\left|c\right|\sqrt{\frac{\delta}{5\gamma}}\right) \tag{2.81}$$

respectively. The minimum noise factor is

$$F_{min} = 1 + 2R_nG_u = 1 + \frac{2}{\sqrt{5}}\frac{\omega}{\omega_T}\sqrt{\gamma\delta\left(1 - \left|c\right|^2\right)} \tag{2.82}$$

where $\omega_T = \frac{g_m}{C_{gs}}$.

In order to achieve as small of a noise factor as possible, one can design an input matching network to make $Y_s$ as close to $Y_{opt}$ as possible, which makes $F$ approach $F_{min}$. At the same time, in the IC design, one has the freedom to choose the size and biasing condition of transistors. By making $F_{min}$ small will also reduce the overall noise figure. From (2.82), one can conclude that increasing $\omega_T$ and/or the correlation coefficient $c$ will reduce the minimum noise factor. For a chosen process, the correlation coefficient is fixed by the technology. The choice left for the circuit designer to reduce $F_{min}$ is to have large $\frac{\omega_T}{\omega}$. For example, suppose that $\gamma = 2$ and $\delta = 4$, and $c$ is assumed to be 0.395. If $\frac{\omega_T}{\omega} = 5$, then $F_{min} = 1.7\ dB$, while if $\frac{\omega_T}{\omega} = 15$, $F_{min}$ will be reduce to 0.6 dB.

### 3. Impact of LNA Gain and Noise Factor on System Sensitivity

The importance of gain and noise factor specifications on LNA can be discussed further from the receiver's system sensitivity aspect. The sensitivity $P_s$ represents the smallest input signal power that can be reliably detected by the system [4]

$$P_s = -174dBm + 10\log BW + SNR + 10\log F_{tot} \tag{2.83}$$

The first two terms in (2.83) are usually referred to as noise floor. BW is the system bandwidth and is determined by a specific application. System SNR is determined by the bit-error-rate (BER[2]) requirement of the system. For example, for a Bluetooth

[2]BER is directly related with $\frac{E_b}{N_0}$ of the received signal in an additive white Gaussian noise (AWGN) channel. $E_b$ is the energy per bit. $N_0$ is the thermal noise density. If the date rate is R bits per second and the signal bandwidth is B, then the signal's SNR can be related to $\frac{E_b}{N_0}$ as: $SNR = \frac{R}{B}\frac{E_b}{N_0}$. So for a certain BER specification, lowering data rate or increasing signal bandwidth can reduce required SNR, thus improve sensitivity.

receiver, simulation shows that a 12.3 dB SNR is needed for a BER lower than $10^{-3}$, and for an 802.11b receiver, an 11.4 dB SNR is required to achieve better than $10^{-5}$ BER. $F_{tot}$ is the system total noise factor and is directly affected by the LNA's gain and noise factor. $F_{tot}$ can be calculated by

$$F_{tot} = F_{LNA} + \frac{F_{afterLNA} - 1}{G_{LNA}} \tag{2.84}$$

The above equation shows that the LNA's noise factor $F_{LNA}$ appears directly in the system's noise factor. For high sensitivity, low system noise factor is required, therefore $F_{LNA}$ should be made as small as possible. The second term of (2.84) shows that noise coming from the stages following the LNA will be suppressed by the LNA's gain, hence a high gain LNA is desirable for high sensitivity. For example, if the LNA's noise figure is 1.5 dB, its available power gain $G_{LNA}$ is 10 dB, and the overall noise figure of the circuits following the LNA is 18 dB, then the system's total noise figure can be found using (2.84) to be 8.8 dB. If the LNA's gain is increased to 15 dB, the total noise figure will be reduce to 5.3 dB. System sensitivity will be improved by 3.5 dB. If 8.8 dB noise figure is a fixed system specification, the noise figure requirement on the circuits after the LNA can be relaxed to 22 dB.

On the other hand, a high gain of the first stage, which is the LNA in this case, will put a more stringent linearity requirement on the following stages. Therefore a trade-off must be made between gain, noise, and linearity. Table I summaries important results obtained in the previous sections.

### D. Large Signal Behavior

The RF amplifier is a non-linear system in nature. If the input signal is small enough, the circuit can be modeled using a linear model around its operating point. But if the

Table I. Important equations for RF amplifier's gain, stability and noise

| | |
|---|---|
| Input and output reflection coefficient | $\Gamma_{in} = s_{11} + \frac{s_{12}s_{21}\Gamma_L}{1-s_{22}\Gamma_L}$ <br> $\Gamma_{out} = s_{22} + \frac{s_{12}s_{21}\Gamma_s}{1-s_{11}\Gamma_s}$ |
| Unilateral transducer power gain | $G_{Tu} = \frac{1-\lvert\Gamma_s\rvert^2}{\lvert 1-s_{11}\Gamma_s\rvert^2} \lvert s_{21}\rvert^2 \frac{1-\lvert\Gamma_L\rvert^2}{\lvert 1-s_{22}\Gamma_L\rvert^2}$ |
| Voltage gain | $A_V = \frac{s_{21}(1+\Gamma_L)}{(1-s_{22}\Gamma_L)(1+\Gamma_{in})}$ |
| Constant gain circle center and radius | $c_i = \frac{g_i s_{ii}^*}{1-\lvert s_{ii}\rvert^2(1-g_i)}$ <br> $r_i = \frac{\sqrt{1-g_i}\left(1-\lvert s_{ii}\rvert^2\right)}{1-\lvert s_{ii}\rvert^2(1-g_i)}$ |
| Unconditional stable for passive source and load termination | $\lvert s_{11}\rvert < 1$ <br> $\lvert s_{22}\rvert < 1$ <br> $K > 1$ <br> $K = \frac{1-\lvert s_{11}\rvert^2-\lvert s_{22}\rvert^2+\lvert s_{11}s_{22}-s_{12}s_{21}\rvert^2}{2\lvert s_{12}s_{21}\rvert}$ |
| Two-port network noise factor | $F = F_{min} + \frac{R_n}{G_s}\lvert Y_s - Y_{opt}\rvert^2$ <br> $F_{min} = 1 + 2R_n\left(G_{opt} + G_c\right)$ <br> $Y_{opt} = G_{opt} + jB_{opt} = \sqrt{G_c^2 + \frac{G_u}{R_n}} + j\left(-B_c\right)$ |
| MOS transistor four noise parameters | $R_n = \frac{\gamma}{\alpha}\frac{1}{g_m}$ <br> $G_c \approx 0$ <br> $B_c = \omega C_{gs}\left(1 + \alpha\lvert c\rvert\sqrt{\frac{\delta}{5\gamma}}\right)$ <br> $G_u = \frac{\delta\omega^2 C_{gs}^2\left(1-\lvert c\rvert^2\right)}{5g_{do}}$ |
| MOS transistor minimum noise factor | $F_{min} = 1 + \frac{2}{\sqrt{5}}\frac{\omega}{\omega_T}\sqrt{\gamma\delta\left(1-\lvert c\rvert^2\right)}$ |

signal level is relatively high, due to non-linearity, the amplifier's dynamic operation point will be changed and become a function of the signal level. The LNA's proper operation must be checked by using a large signal input. On the other hand, although the signal itself is small, large interferers may come together with the signal. This situation is shown in Fig. 7(a). The interferers can be coming from the adjacent channel or generated by intentionally jammer. The interference specifications are usually provided by the system standards. The non-linearity performance is characterized by the two-tone test ($f_1$, $f_2$) as depicted in Fig. 7(b). Usually distortion term $2f_1 - f_2$ and $2f_2 - f_1$ fall in-band, they are characterized by the 3rd order non-linearity. For example, the desired signal channel in Fig. 7(a) has a bandwidth of 1 MHz and is centered at 1000 MHz. Two large blockers are located 1 MHz away from the center of the channel and separated by 1 MHz, i.e. $f_1 = 1001$ MHz and $f_2 = 1002$ MHz. Thus the lower side IM3 component will be at $2f_1 - f_2 = 1000$ MHz, which is right upon the center of the channel, degrading the signal's SNR. Detailed non-linear distortion analysis can be found in Chapter V. A large in-band blocker tends to desensitize the circuit, it is measured by the 1-dB compression point. Dynamic range measures the signal handling capacity of a circuit, which is bounded by the IIP3 and system noise floor.

### 1. 1-dB Compression Point

The 1-dB compression point ($P_{1dB}$) is the point (input or output) where the fundamental gain reduced by 1 dB from the ideal small signal gain at a certain frequency (see Fig. 8). Assume a non-linear system can be approximated by Taylor series

$$y(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) + \cdots \tag{2.85}$$

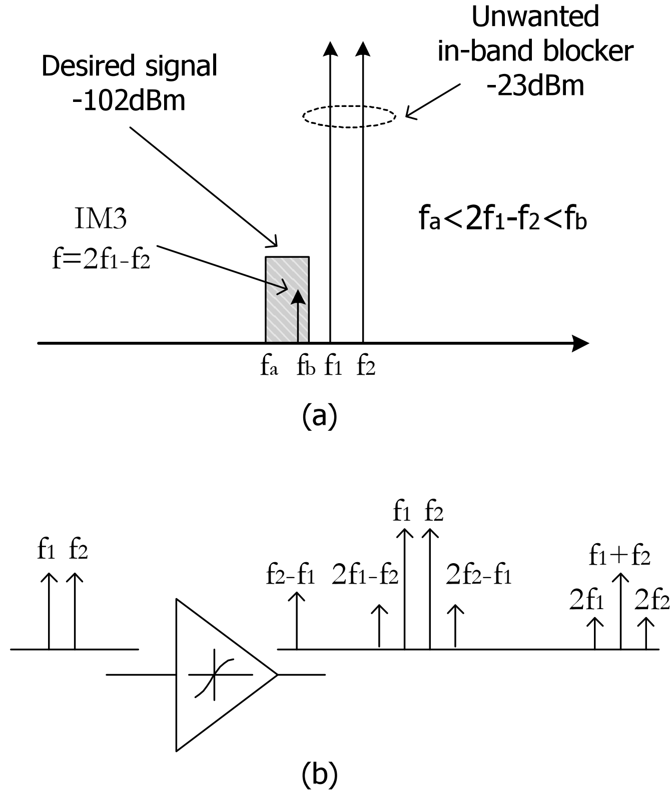



Fig. 7. (a) In-band blockers generate in-channel intermodulation term (b) The two–tone test spectrum

The input-referred 1-dB compression point [4] can be calculated as

$$P_{1dB} = \sqrt{0.145 \left| \frac{\alpha_1}{\alpha_3} \right|} \tag{2.86}$$

where $\alpha_1$ and $\alpha_3$ are the 1st-order and 3rd-order coefficients of the Taylor series expansion of the system'd input/output characteristics as in (2.85).

## 2. Intercept Point

The two tone test is usually used to mimic the real-world scenario in which both a desired signal and a potential interferer feed the input of the amplifier. Due to

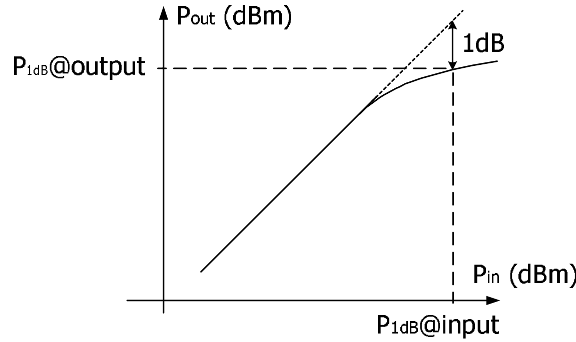



Fig. 8. 1-dB compression point

the non-linearity of the circuit, the 2nd and 3rd order inter-modulation products will appear at the output and they may lie within the pass band thus, degrading the desired output signal.

We usually plot the desired output (fundamental), 2nd order intermodulation output (IM2) and 3rd order intermodulation output (IM3) as a function of the input signal level. The 2nd order intercept point (IP2) is the extrapolated intersection of the fundamental curve and the IM2 curve. The 3rd order intercept point (IP3) is the extrapolated intersection of the fundamental and IM3 curve (Fig. 9).

For the system described by (2.85), the input-referred IP3 (IIP3) [4] is given by

$$IIP3 = \sqrt{\frac{4}{3}\left|\frac{\alpha_1}{\alpha_3}\right|} \tag{2.87}$$

From (2.86) and (2.87), it is easy to show that IIP3 is about 10 dB higher than $P_{1dB}$ for the same system if the third-order non-linearity dominates the linearity behavior.

### 3. Dynamic Range

Dynamic range (DR) is generally defined as the ratio of the maximum input level that the circuit can tolerate without appreciable distortion to the minimum input level at

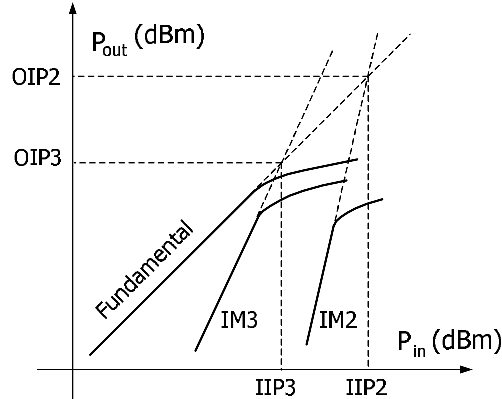



Fig. 9. Meaning of IP2 and IP3

which the circuit provides a reasonable signal quality. Two different definitions are usually used: spurious-free dynamic range (SFDR) and compression-free dynamic range (CFDR).

Fig. 10(a) shows the definition of SFDR. The upper bound of SFDR is based on intermodulation behavior and is defined as the maximum input level in a two-tone test for which the third order intermodulation (IM3) products do not exceed the noise floor. From Fig. 10(b), it can be shown that the input level for which the IM3 products become equal to the noise floor is given by

$$P_{in,max} = \frac{2 \times IIP3 + N_{floor}}{3} \tag{2.88}$$

where $N_{floor} = -174$ dBm $+ NF + 10 \log BW$, is the noise floor. All the quantities are expressed in dBm.

The lower bound of SFDR is limited by the system's sensitivity. If for a certain required signal quality, the minimum SNR is $SNR_{min}$, then the minimum detectable

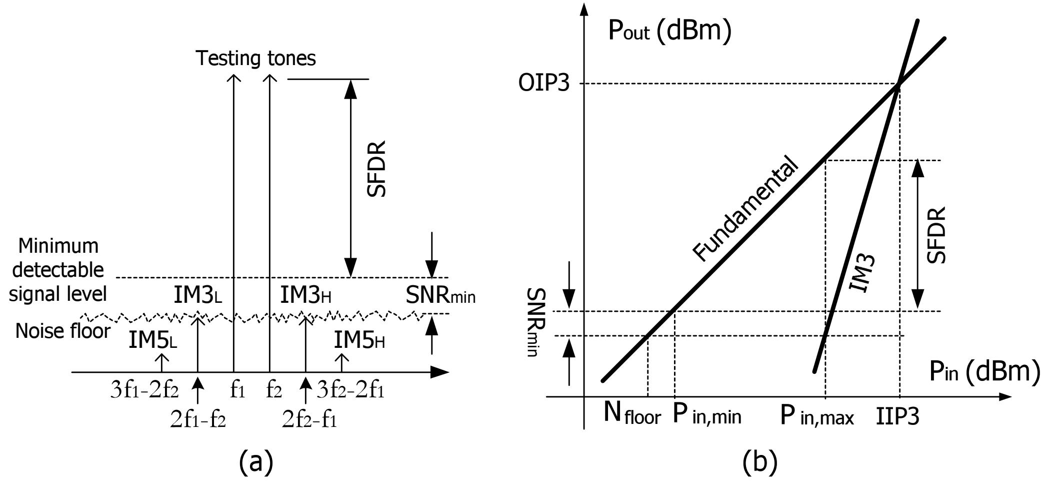



Fig. 10. (a) SFDR defination (b) Relationship between SFDR and IIP3

signal level in dBm at the input is

$$P_{in,min} = SNR_{min} + N_{floor} \tag{2.89}$$

This is also shown in Fig. 10(b). The SFDR is then calculated by the difference between $P_{in,max}$ and $P_{in,min}$:

$$SFDR = \frac{2}{3}\left(IIP3 - NF - 10\log B + 174\text{ dBm}\right) - SNR_{min} \tag{2.90}$$

Compression-free dynamic range (CFDR) is the difference, in dB, between the input-referred 1-dB compression point and the noise floor as in Fig. 11:

$$CFDR = P_{1dB} - N_{floor} \tag{2.91}$$

Here is a numerical example. Suppose an 802.11b receiver has the following system specifications:

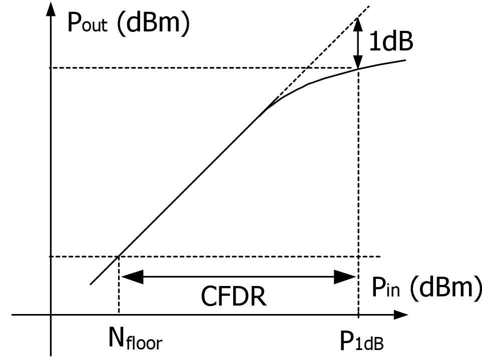



Fig. 11. Compression-free dynamic range

| Parameter | Value | Unit |
| --- | --- | --- |
| $SNR_{min}$ | 11.4 | dB |
| BW | 6.0 | MHz |
| IIP3 | -13 | dBm |
| NF | 12.9 | dB |

The noise floor $N_{floor}$ calculated from the above data is -93 dBm. From (2.88) and (2.89), $P_{in,max}$ and $P_{in,min}$ are -40 dBm and -82 dBm, respectively. Thus SFDR is 42 dB. The 1dB compression point can be estimated from the IIP3 which is usually 10 dB larger than $P_{1dB}$, therefore CFDR is about 90 dB.

4. Wide-Band Non-Linearity

The non-linearities of the wide-band circuits are also measured by their third-order and second-order linearity behavior. But different from narrow-band system, wide-band signal occupies a large amount of bandwidth and the two-tone test for narrow-band system is not sufficient. For example, the cable TV (CATV) system has many equally spaced carries, the linearities are measured by the composite triple-order beat

(CTB) and composite second-order distortion (CSO) [5].

For three equally spaced carries/tones ($x_1$, $x_2$, $x_3$) in a wide-band system, the distortion terms generated by the third order non-linearity are:

$$(x_1 + x_2 + x_3)^3 = H3 + IM3 + TB \tag{2.92}$$

where

$$H3 = x_1^3 + x_2^3 + x_3^3 \tag{2.93}$$

$$IM3 = 3x_1^2 x_2 + 3x_1^2 x_3 + 3x_2^2 x_1 + 3x_3^2 x_1 + 3x_2^2 x_3 + 3x_3^2 x_2 \tag{2.94}$$

$$TB = 6x_1 x_2 x_3 \tag{2.95}$$

The third order harmonic products in $H3$ are at three times of frequency. There are N of them for N tones or carriers. These products are 15.6 dB weaker than the triple-beat term in (2.95) and do not fall near a carrier, thus they usually are ignored. The third order intermodulation products in (2.94) are half of the magnitude of the triple-beats and they are fewer in number. In a system with 20 channels, the contribution of the IM3 terms is less than 0.1 dB and their contribution decreases with increasing number of channels.

The triple-beat (TB) term creates spurious frequencies at $f_1 \pm f_2 \pm f_3 (f_1 < f_2 < f_3)$ and those spurs are 6 dB stronger than the third order intermodulation products in (2.94). Fig. 12 illustrates the triple-order beats generation. The beats are drawn offset from the carriers for showing purpose, they are actually reside on the carriers. In the case of N tones/carriers, the total number of composite distortion products can be approximated by $\frac{N^2}{4}$at the edge of the band and $\frac{3}{8}N^2$at the middle of the band. So triple-beats dominate the third-order distortion performance in a wide-band system.

The triple-beat power in dBm can be related to the intercept point power $P_{IP3}$

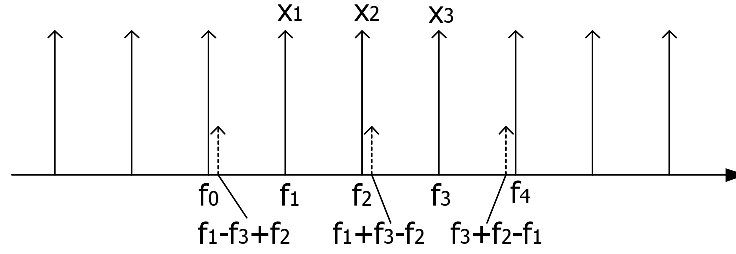



Fig. 12. Triple-order beats

and signal power $P_s$ as:

$$TB = P_{IP3} - 3\left(P_{IP3} - P_s\right) + 6 \tag{2.96}$$

The CTB is usually measured by how many dB the triple-beat down from the signal power. Because the mid-band has the maximum number of beats, so

$$CTB = 2\left(P_{IP3} - P_s\right) - 10\log\left(\frac{3}{8}N^2\right) - 6 \tag{2.97}$$

Composite second order distortion (CSO) is a result of one or two carries experiencing a second order non-linearity. Comparing to a narrow band system, multiple second-order beats exist for a wide-band system having N carriers. For a CATV system having $f_L$as its lowest channel, $f_H$as its highest channel and $d$ is the frequency offset from a multiple of 6 MHz (1.25 MHz), then the number of second-order beats above any given carrier $f$ is calculated by

$$N_B = (N-1)\frac{f - 2f_L + d}{2\left(f_H - f_L\right)} \tag{2.98}$$

and the number of second-order beats below a given carrier $f$ is

$$N_B = (N-1)\left(1 - \frac{f-d}{f_H - f_L}\right) \tag{2.99}$$

Each of the above second-order beats is an IP2 tone, so its power can be expressed in dBm as

$$SO = P_{IP2} - 2\left(P_{IP2} - P_s\right) \tag{2.100}$$

The CSO is specified as how many dBs the beats power down from the signal power and will be given by

$$CSO = P_{IP2} - P_s - 10\log N_B \tag{2.101}$$

Note that all the quantities are measured at the output of the system. A typical value of CTB and CSO for a CATV system would be -97 dBc and -89 dBc.

Table II summaries the equations obtain in this section.

Table II. Equations related to non-linearity

| | |
|---|---|
| Non-linear system characteristics | $y\left(t\right) = \alpha_1 x\left(t\right) + \alpha_2 x^2\left(t\right) + \alpha_3 x^3\left(t\right) + \cdots$ |
| Non-linear circuit 1dB compression point | $P_{1dB} = \sqrt{0.145\left\lvert\frac{\alpha_1}{\alpha_3}\right\rvert}$ |
| Non-linear circuit 3rd order intercept point | $IIP3 = \sqrt{\frac{4}{3}\left\lvert\frac{\alpha_1}{\alpha_3}\right\rvert}$ |
| System dynamic range | $SFDR = \frac{2}{3}\left(IIP3 - NF - 10\log B + 174\text{dBm}\right) - SNR_{min}$ |
| Wide-band system triple-beat distortion | $CTB = 2\left(P_{IP3} - P_s\right) - 10\log\left(\frac{3}{8}N^2\right) - 6$ |
| Wide-band system second order distortion | $CSO = P_{IP2} - P_s - 10\log N_B$ |

## E. LNA Topologies in CMOS Technology

Before continuing the discussion of LNA topologies, we will review the LNA specifications in some wireless standards. For an LNA used in a GSM mobile receiver, it will typically require about 20 dB gain, less than 2 dB noise figure, better than -10 dBm IIP3. Its power consumption should be minimized and return loss should no less than -8 dB. For the LNA used in a Bluetooth low-IF receiver, it will need to have a voltage gain of 18 dB, noise figure of 3.5 dB and IIP3 of -3.5 dBm. The power consumption also needs to be minimized and input return loss should be better than 10 dB. The challenge of LNA design is to fulfill all the specifications with limited power consumption and silicon area.

LNA's performance is more dependent on process technology than on circuit topology. Indeed, LNA usually only involves one or two transistors in its signal path, and there is not much degree of freedom to form different architectures. Still, for a fixed technology, different circuit structures will produce a different performance and design trade-off. Fig. 13 shows several popular LNA structures implementable in a CMOS integrated circuit [3] [6]. Because an LNA's input will directly interface with a RF filter which generally requires certain impedance termination, so input impedance matching is a must requirement for all the LNA's listed in Figs. 13. The LNA structures are distinguished from each other by how the input impedance matching is achieved.

Fig. 13(a) achieves input matching by directly placing a 50 Ω resistor in parallel with the gate of transistor $M_1$. This is the most straightforward method but the noise figure is exceptionally high. A lower bound of its noise factor is

$$F \geq 2 + \frac{4\gamma}{\alpha}\frac{1}{g_m R_s} \tag{2.102}$$

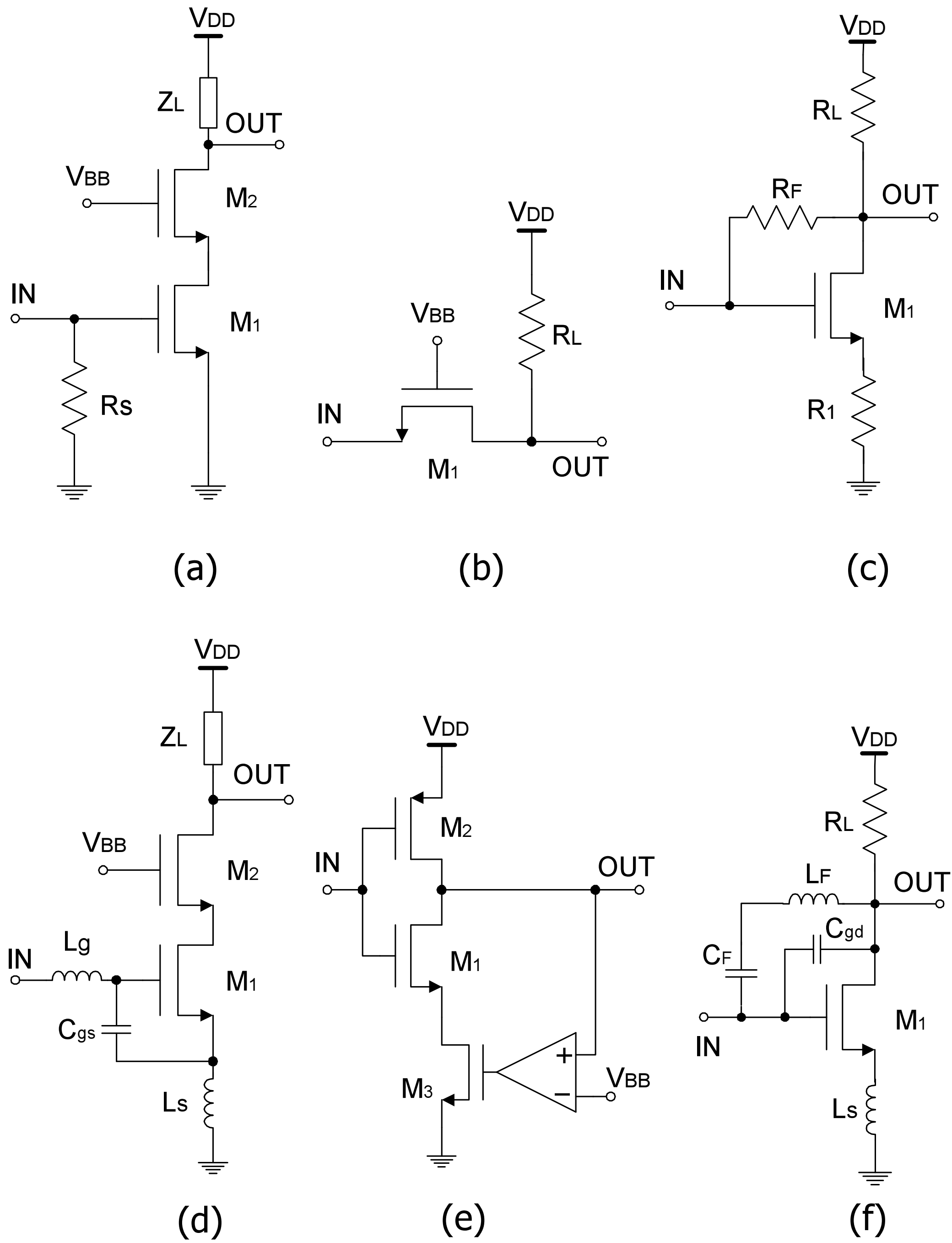



Fig. 13. Popular single-ended CMOS LNA topologies (a) Resistive termination (b) Common gate (c) Shunt-series feedback (d) Inductive source-degeneration (e) Current-reuse (f) $C_{gd}$ neutralization

So the noise figure is readily larger than 6dB. The primary contribution of noise comes from the termination resistor and transistor drain noise. Only in very rare cases, is this LNA structure employed and usually not referred to as an LNA anymore.

The difficulty of input impedance match comes from the high input impedance if a common source configuration is used. Common gate configuration avoids the high gate input impedance as in Fig. 13(b). For the first order approximation, the real part of input impedance is just $\frac{1}{g_m}$. By carefully choosing the size of the transistor and biasing condition, 50 Ω impedance match is readily obtained. Ignoring gate current noise, a lower bound of noise factor for this topology is represented by

$$F \geq 1 + \frac{\gamma}{\alpha} \tag{2.103}$$

This bound is about 2.2 dB and 4.8 dB for a long and short channel device respectively. The induced gate noise will make the noise factor larger, but the drain noise is still the dominant factor.

Fig. 13(c) utilizes negative shunt feedback to modify the input impedance of a common source stage. Its input impedance can be calculated approximately by

$$Z_{in} = \frac{R_F}{1 + A} \tag{2.104}$$

where A is the voltage gain from input to output and is approximately in the order of $\frac{R_L}{R_1}$, assuming $M_1$'s $g_m$ is large enough. The noise figure of this structure is far better than that of 13(a) but it is still too high to use in some applications. A noise factor expression for this LNA without the source resistor $R_1$ is given as [7]

$$F = 1 + R_s \delta g_g + \left(\frac{G_s + G_F}{g_m - G_F}\right)^2 R_s \left(G_L + \gamma g_{do}\right) + \left(\frac{G_s + g_m}{g_m - G_F}\right)^2 R_s G_F \tag{2.105}$$

where $G_s$ is the conductance of the signal generator, $G_F = R_F^{-1}$ and $G_L = R_L^{-1}$. For lower power consumption, drain noise is the major noise contribution, but for high

power consumption, gate noise and the noise due to $R_F$ also become significant.

All of the three architectures introduced above have another similarity. Their input impedance matching can cover a wide frequency band. Proper input impedance is required by the RF filter preceding the LNA and the maximum power transfer. The variations of Fig. 13(a), (b) and (c), are also widely used in wide-band systems [8] [9].

Fig. 13(d) is a very popular narrow-band LNA. It is narrow-band because impedance matching is only established within a very narrow frequency range due to the resonant nature of the reactive matching network. Impedance matching is established by inductive degeneration. Around operation frequency $\omega_o = \frac{1}{\sqrt{C_{gs}(L_g+L_s)}}$, the input impedance only presents a real part $Z_{in} = \frac{g_m}{C_{gs}} L_s$. Detailed analysis can be found in [10]. The noise figure of this inductive degenerated LNA can be readily made below 2 dB or even lower [11] [12].

A closed form noise factor expression can be found to be [7]

$$F = 1+\left(\frac{\omega_o}{\omega_T}\right)^2 R_s\gamma g_{do}+\left[\left(\frac{\omega_T}{\omega_o g_m R_s}\right)^2+1\right] R_s\delta g_g+0.79R_s\left(\frac{\omega_o}{\omega_T}\right)\sqrt{\gamma g_{do}\delta g_g} \quad (2.106)$$

In the above equation, the second term represents the contribution of the drain noise, the third term is the gate noise contribution, and the last term shows the noise contribution due to the correlation between the gate and drain noise. When considering the series resistance of the inductors used in the LNA, an additional term $\frac{R_{L_g}}{R_s}$ should be added to (2.106). The noise from series resistance of the degeneration inductor is negligible. Among all the noise contributors, gate noise is the largest one. This is because the gate noise current sees a high impedance due to the resonance of the input matching network. Therefore, in order to reduce the noise factor, the Q value of the input matching network should be limited.

The basic problem with using CMOS transistor for LNAs is its inherently low transconductance and hence low gain. However, [6] uses a current-reuse technique

to almost double a single stage transconductance without increasing bias current. Fig. 13(e) shows a simplified schematic of this design. The key point is that given the same bias current, the effective transconductance is $g_{m1} + g_{m2}$, while it is simply $g_{m1}$ in the case of no $M_2$ presented. A major drawback of this design is its high input and output impedances, thus requiring external impedance matching networks. This prevents the use of this LNA in fully integrated applications Due to the high gain property, strong Miller effect reduces the reverse isolation of this LNA. In the actual design, two identical stages are cascaded to improve the reverse isolation.

In order to improve the reverse isolation of the LNA, $C_{gd}$ neutralization technique can be used as shown in Fig. 13(f). The LNA's reverse isolation is limited by the drain-source parasitic capacitor $C_{gd}$. An inductor $L_F$ is added in parallel with this capacitor to provide a different feedback polarity to cancel the effect of $C_{gd}$. Care must be taken to ensure that the inductive feedback does not incur any potential stability issues.

Table III summarizes the performance of the five LNA circuits discussed in this section.

### F. Design Procedure of a Source Degenerated CMOS LNA

In this section, an inductive source degenerated LNA will be designed using a 0.18 $\mu m$ CMOS technology to show the LNA design procedure and trade-offs. The simplified LNA schematic is redrawn in Fig 14. We want to have an LNA working at 2.4 GHz ISM band with less than 1.6 dB noise figure (noise factor: 1.45), -8 dBm IIP3, 20 dB (10 V/V) voltage gain and drawing no more than 10 mA current from a 1.8 V power supply.

The step by step design details from hand calculations to simulation verifications

Table III. Popular CMOS LNA architectures

| | Highlight | Drawback | Noise figure |
|---|---|---|---|
| Resistive termination | Effortless input match | Excessive large NF | $> 6$ dB |
| Common gate | Easy input match moderate NF | Large NF and power | $2.2 \sim 4.8$ dB |
| Series/shunt feedback | Broadband input and output match | Stability issue and isolation | $1.8 \sim 5$ dB |
| Inductive degeneration | Good narrowband match, small NF | Large area | $\sim 2$ dB |
| Current reuse | High gain, Low power | External matching network required | $\sim 2.2$ dB |
| Inductor neutralization | Good reverse isloation | Increased area, stablility concern | $\sim 2$ dB |

are provided below. A flow chart is given in Fig. 15 to further visualize the design flow.

Step 1: Important Design Equations

Because low-noise is the most important requirement for a LNA, the design consideration will start from the LNA's noise factor equation. The major noise contribution comes from $M_1$. Under input impedance match conditions:

$$\omega_T L_s = R_s \tag{2.107}$$

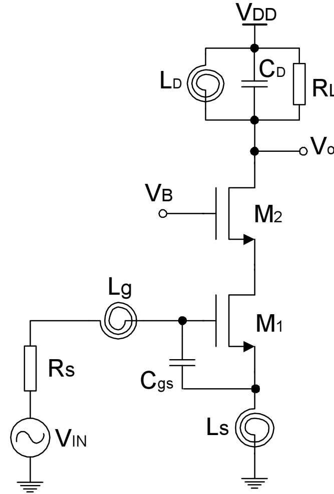



Fig. 14. Inductive source degenerated LNA

and

$$\frac{1}{\sqrt{C_{gs}\left(L_g + L_s\right)}} = \omega_o \tag{2.108}$$

where $C_{gs}$ and $\omega_T$ are $M_1$'s gate-source capacitance and cut-off frequency respectively. $\omega_o$ is the operation frequency. The noise factor of the LNA can be shown to be

$$F = 1 + \kappa_{nf}\left(\frac{\omega_o}{\omega_T}\right) \tag{2.109}$$

where

$$\kappa_{nf} = \frac{\gamma}{\alpha}\frac{1}{2Q}\left[1 - 2\left|c\right|\chi_d + 4\left(Q^2+1\right)\chi_d^2\right] \tag{2.110}$$

$$Q = \frac{1}{2R_s\omega_o C_{gs}} \tag{2.111}$$

$$\chi_d = \alpha\sqrt{\frac{\delta}{5\gamma}} \tag{2.112}$$

$\kappa_{nf}$ is called noise factor scaling coefficient. $R_s$ is the source impedance and usually is $50\,\Omega$ or $75\,\Omega$. $Q$ is the input quality factor and also controls the voltage gain from input to gate-source port.

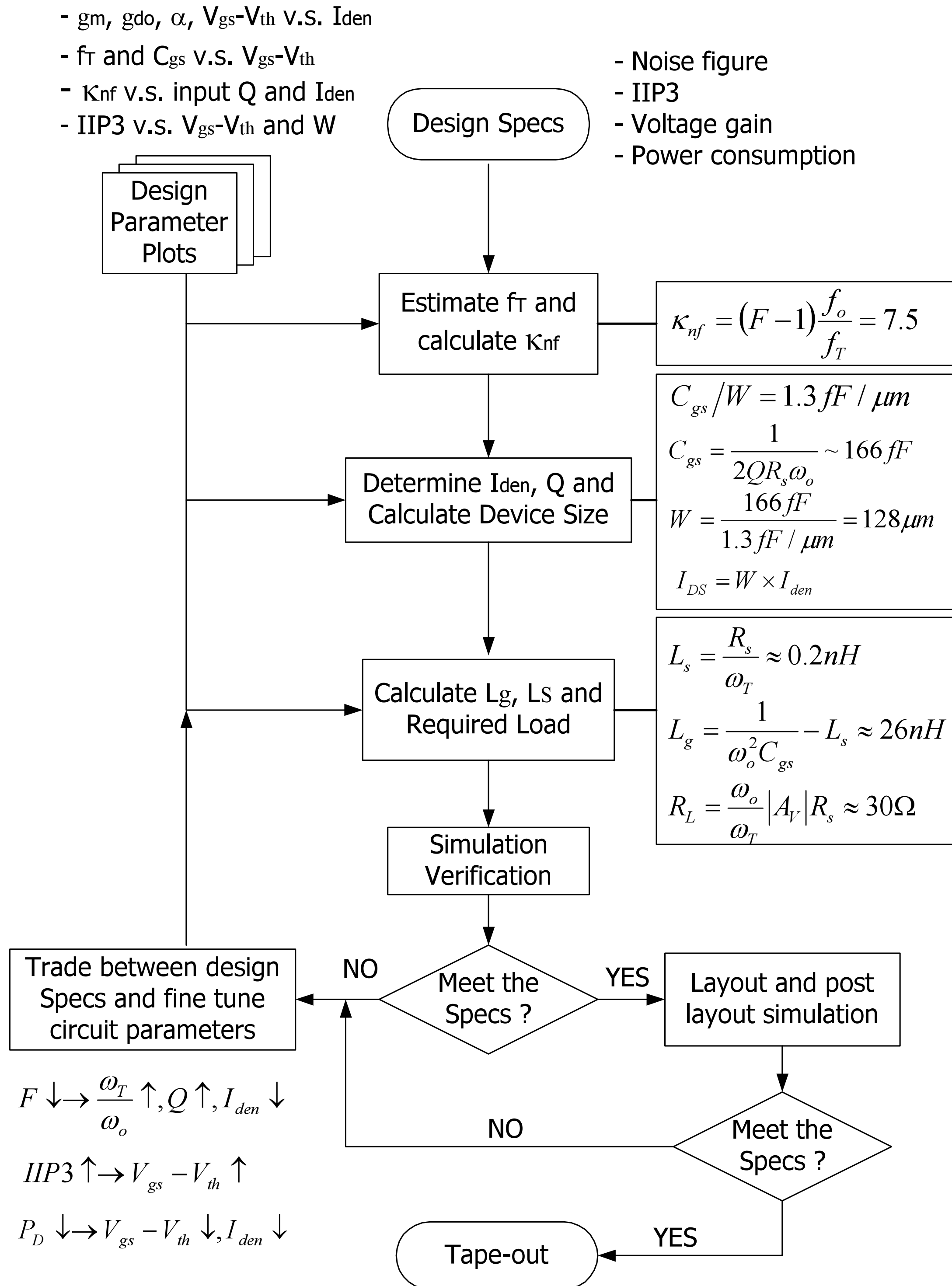



Fig. 15. Flow chart of LNA design procedure

Step 2: Know the Process and Obtain Design Guide Plots

To start, it is necessary to know the behavior of MOS transistor's parameters related with the above equations. The data can be collected through measurement or simulation. The first parameter we want to investigate is $\alpha$. It is defined as the ratio of $g_m$ to $g_{do}$ and changes with biasing condition. Fig. 16 plots the value of $\alpha$ and $V_{gs} - V_{th}$ versus drain current density $I_{ds}/W$. $g_m$ and $g_{do}$ are also shown on the graph in the form of $g_m/W$ and $g_{do}/W$.

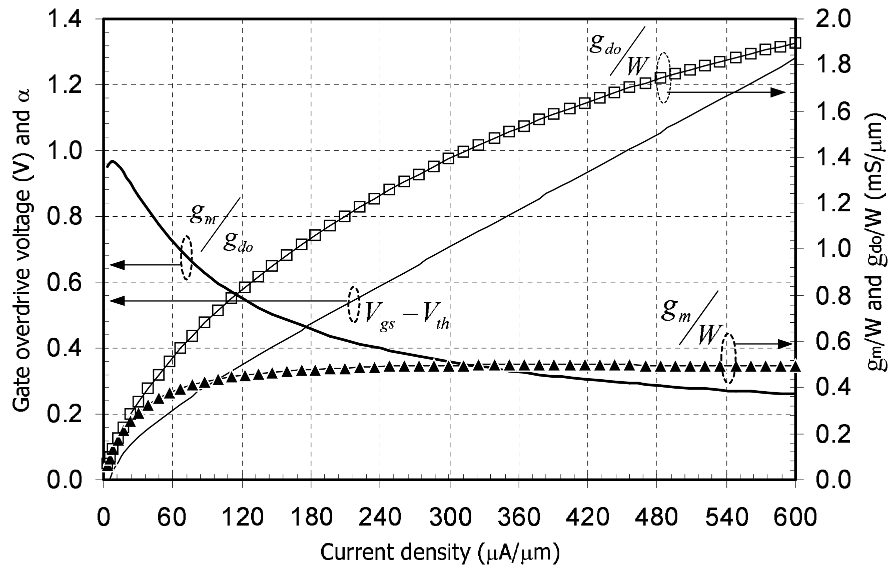



Fig. 16. $\alpha$, $V_{gs} - V_{th}$, $g_m/W$, and $g_{do}/W$ versus drain current density

Here minimum length of transistor is used. When bias current is low, MOS transistor can be modeled using long channel approximation, thus $\alpha$ is near unity. With the increasing of bias current, short channel effect becomes significant rapidly. $\alpha$ deviates from unity. It is also observed that after current density is greater than $70\,\mu A/\mu m$, making it larger does not help to increase $g_m$ much. $\gamma$ is treated as a constant and assumed to be about 3. $\delta/\gamma$ is also assumed to be constant and equal

to 2.

The next two parameters are the gate-source capacitance normalized to gate width, i.e. $C_{gs}/W$, and device cut off frequency $f_T$ when transistor is biased in saturation region. Fig. 17 shows the curves of $C_{gs}/W$ and $f_T$ versus gate over drive voltage $V_{gs} - V_{th}$.

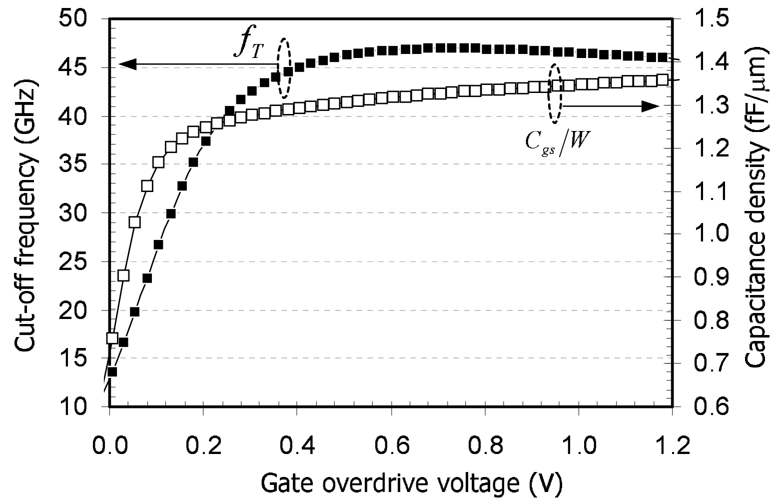



Fig. 17. $C_{gs}/W$ and $f_T$ versus gate over drive voltage $V_{gs} - V_{th}$

Some insights can be gained from these plots. $f_T$ increases with $V_{gs} - V_{th}$ when $V_{gs} - V_{th}$ is small. When $V_{gs} - V_{th}$ becomes larger than 0.3 V, short channel effects such as mobility reduction and velocity saturation make $g_m$ increasing slowly with $V_{gs} - V_{th}$ and finally reaching a saturation value. Therefore the speed of $f_T$ increment with $V_{gs} - V_{th}$ is also reduce. When gate overdrive is small, transistor is biased at weak inversion region, the charge controlled by the gate is sparse thus gate-source capacitance is small. Channel charge increases rapidly with gate overdrive and so does the gate-source capacitance. After $V_{gs} - V_{th}$ becomes larger than 0.2 V, $C_{gs}/W$ increases with gate overdrive very slowly. At high gate overdrive, $g_m$ is almost constant and $C_{gs}$

increases slowly, so $f_T$ begins to degrade gradually after $V_{gs} - V_{th}$ is larger than 0.8 V.

Now the noise factor scaling coefficient can be plotted against $Q$ and drain current density $I_{ds}/W$ as illustrated in Fig. 18, where (a) is the 3-D plot for visual inspection and (b) is the 2-D plot for design lookup purpose. Generally, the larger the current density, the smaller the scaling coefficient; and the larger the Q, the larger the larger the scaling coefficient. For a really large current density there is an optimal value of Q which minimize the scaling coefficient. But the current density maybe too large for practical applications. It is also worth to notice that for a fixed current density, choosing large Q will reduce the total current consumption.

The last plot is not related directly to noise factor but to the linearity performance. Fig. 19 is the IIP3 versus gate overdrive at 2 GHz for different size of devices. Note that IIP3 is more dependent on gate over-drive voltage than device size.

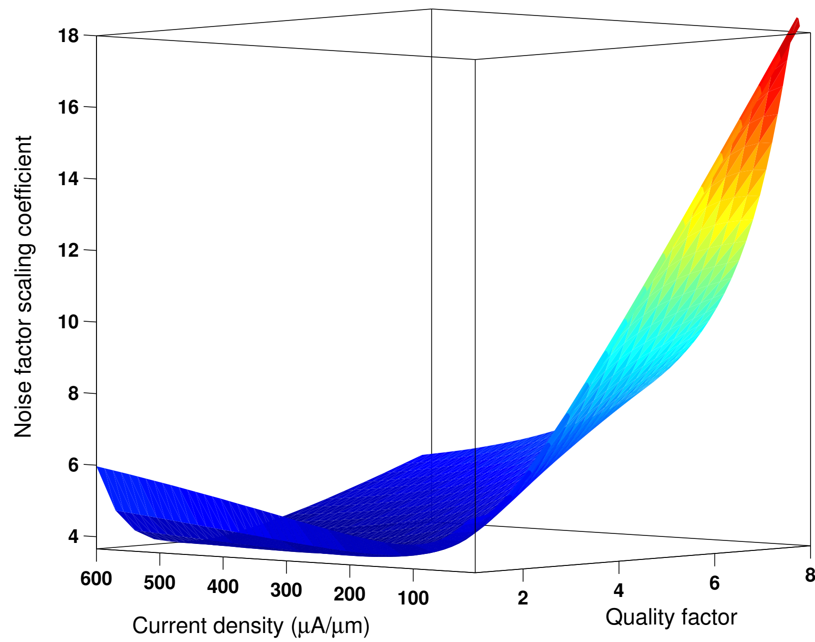
Step 3: Estimate $f_T$ and Calculate $\kappa_{nf}$

Because we have a small current budget, the gate over drive can not be very large. It will probably some where between 0.2 V to 0.4 V. Here we do not intend to fix the gate overdrive voltage but rather have a rough idea of what value of $f_T$ we can use. From Fig. 17, $f_T$ is estimated to be about 40 GHz. For 1.6 dB noise figure, from (2.109) we need $\kappa_{nf}$ to be no larger than 7.5:

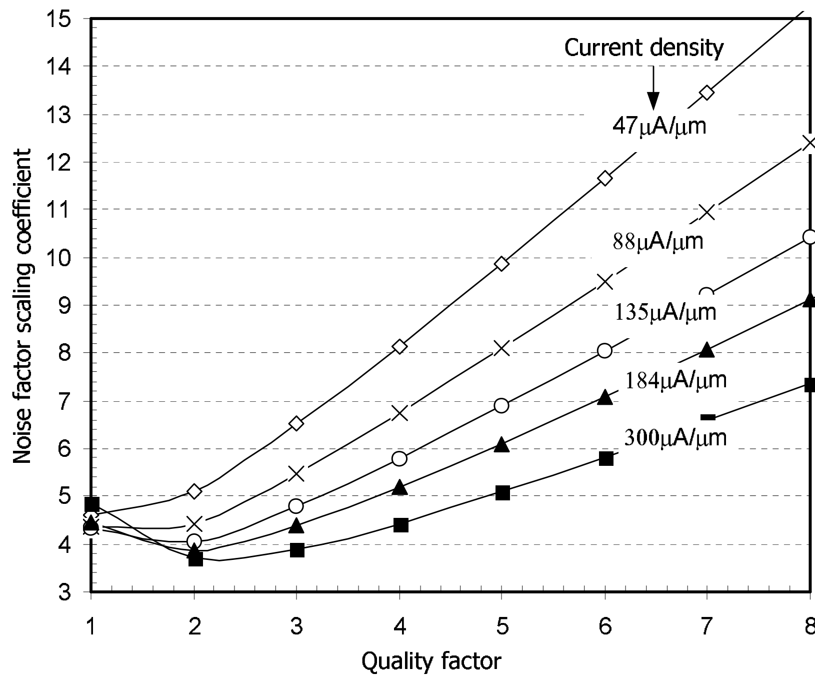
$$\kappa_{nf} = (F - 1)\,\frac{f_T}{f_o} = 7.5 \qquad (2.113)$$

Step 4: Determine Current Density, Q factor and Calculate Device Size

For 0.2 V to 0.4 V gate overdrive, from plots in Fig. 16, the current density is somewhere between $60\,\mu A/\mu m$ and $140\,\mu A/\mu m$. From Fig. 19, the device raw IIP3 is from 7 dBm to 14 dBm, so if Q is chosen to be 4, IIP3 will have good change meet the




(a)




(b)

Fig. 18. Noise factor scaling coefficient versus quality and current density for $0.18\mu m$ NMOS device (a) 3-D plot (b) 2-D plot

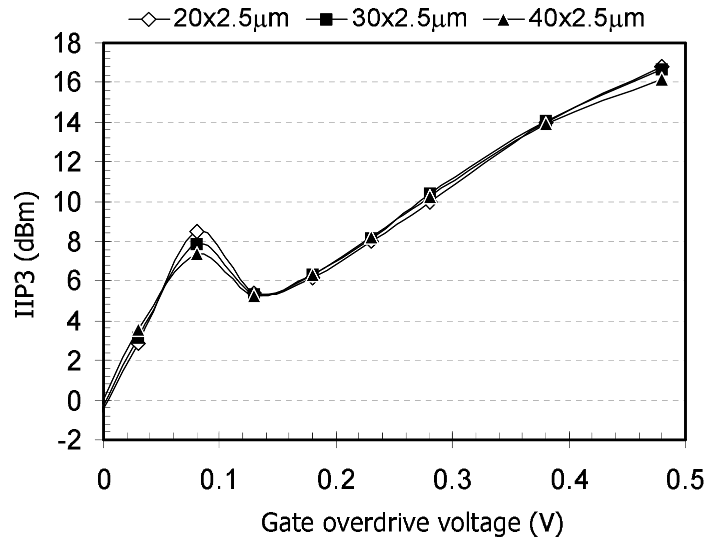



Fig. 19. IIP3 versus gate overdrive and device size

specification. From Fig. 18(b), for Q of 4, the current density is chosen to be about $70\,\mu A/\mu m$. The required gate-source capacitance is calculated to be

$$C_{gs} = \frac{1}{2QR_s\omega_o} \approx 166\,fF \tag{2.114}$$

From Fig. 17, the gate-source capacitance density is about $1.3\,fF/\mu m$, so the required device width will be

$$W = \frac{166\,fF}{1.3\,fF/\mu m} \approx 128\,\mu m \tag{2.115}$$

The device current can be estimated by

$$I_{ds} = W \times 70\,\mu A/\mu m \approx 8.9\,mA \tag{2.116}$$

Under the above conditions, the transistor's gate overdrive voltage is about 0.23 V and its transconductance $g_m$ is about $50\,mA/V$. Before continuing to calculate the

required gate and source inductance, $\omega_T$ is re-calculated by :

$$\omega_T = \frac{g_m}{C_{gs}} \approx 2\pi \times 48\, GHz \qquad (2.117)$$

Considering the effect of gate-drain capacitance $C_{gd}$, the $\omega_T$ is pretty near the value being estimated before.

Step 5: Calculate $L_s$, $L_g$ and Required Load

The source degeneration inductance is obtained by

$$L_s = \frac{R_s}{\omega_T} \approx 0.2\, nH \qquad (2.118)$$

And the gate inductance is

$$L_g = \frac{1}{\omega_o^2 C_{gs}} - L_s \approx 26\, nH \qquad (2.119)$$

The cascoded transistor $M_2$ will use the same size as the main driving transistor and its size can be optimized through simulation if required. The gate bias voltage of $M_2$ is tied to power supply in this example, its value can also be optimized. Generally, it is chosen just large enough to bias $M_1$ in saturation region for all possible input levels. The use of $M_2$ reduces the Miller effect and improves reverse isolation of the LNA.

In order to obtain a desired voltage gain, the load impedance has to be set correctly. Assuming the load is LC tuned and its equivalent impedance around $\omega_o$ is dominated by load resistor $R_L$. The voltage gain of the LNA can be shown to be

$$A_v = jQg_m R_L = j\left(\frac{\omega_T}{\omega_o}\right)\frac{R_L}{R_s} \qquad (2.120)$$

So the value of load resistor should be

$$R_L = \left(\frac{\omega_o}{\omega_T}\right) |A_v| \, R_s \approx 30\,\Omega \tag{2.121}$$

Step 6: Simulation Verification

Now the design calculation is finished. But most of the design parameters need to be fine tuned through circuit simulation. The above procedure shows the design trade-offs between all the design parameters and can be used to guide circuit adjustment in simulation. The initially calculated and simulated final device parameters are listed in Table IV. Table V compares the targeted specifications and simulation results. The hand calculations match the simulation results well except that the gate inductance is much more over estimated by hand calculations. Fig. 20 shows the simulation plots of various parameters of this LNA. Note that it is assumed that high quality on-chip inductors are available in our design, their Q values can be as large as 20 at the working frequency.

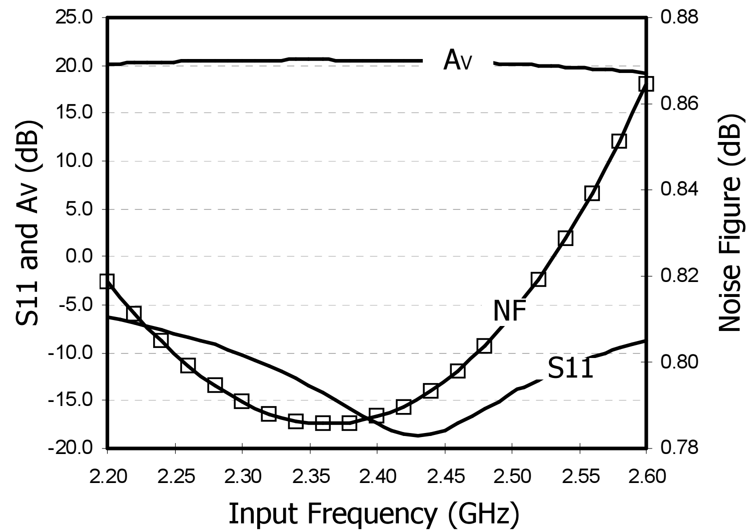
A differential LNA can be designed using the same procedure provide above. A MOS differential pair should be investigated in order to obtain the required design plots. Detailed discussion of differential LNA design issues will be given in Chapter IV.

Table IV. Calculated and simulated design parameters

| Device parameter | Calculation | Simulation |
| --- | --- | --- |
| W | 128 $\mu m$ | 127.5 $\mu m$ |
| $I_{ds}$ | 8.9 mA | 8 mA |
| $g_m$ | 50 mA/V | 50.7 mA/V |
| $C_{gs}$ | 166 fF | 151fF |
| $L_s$ | 0.2 nH | 0.2 nH |
| $L_g$ | 26 nH | 16 nH |
| $R_L$ | 30 Ω | 40 Ω |

Table V. Targeted specifications and simulated performance

| Design parameter | Target | Simulated |
| --- | --- | --- |
| Noise figure | 1.6 dB | 0.8 dB |
| Current drain | <10 mA | 8.0 mA |
| Voltage gain | 20 dB | 21 dB |
| IIP3 | -8 dBm | -6.4 dBm |
| $P_{1dB}$ | – | -20 dBm |
| $S_{11}$ | – | -17 dB |
| $S_{12}$ | – | -25 dB |
| Power Supply | 1.8 V | 1.8 V |



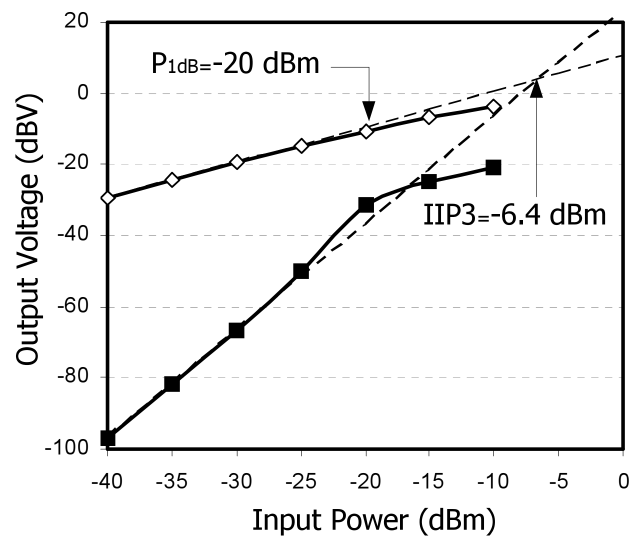




Fig. 20. Simulation plots of the designed LNA (a) Voltage gain, S11 and noise figure (b) IIP3 and $P_{1dB}$

# CHAPTER III

# MIXER DESIGN OVERVIEW AND AN IMPLEMENTATION FOR LOW-IF BLUETOOTH RECEIVER

In the receiving path (Rx) of a communication system, one of the goals is to pick up the useful signal from the presence of noise and interferer, i.e. select interested channel band. While it is hard to design a filter with a 1 MHz band centered at 2.4 GHz (Q$\sim$2400), it is relatively easier to select a 1 MHz band centered at 4 MHz (Q$\sim$4). The down-conversion mixer is the key block that transfers the spectrum from a high frequency band to a low frequency band.

At the same time, in the transmitting path (Tx) of the system, in order to make the transmission of the baseband signal efficient and practical, the low frequency band usually needs to be transferred to a much higher frequency band. Thus, the signal can be radiated out by a compact enough antenna. Up-conversion mixer does this job. It translates the low frequency baseband spectrum to the RF spectrum, then the power amplifier amplifies this high frequency signal to a sufficient amount of power and radiates out by the antenna. Fig. 21 is the simplified block diagram of a low-IF/zero-IF transceiver. It shows the position and roles of the up- and down-conversion mixers.

## A. Mixer Mathematical Model

Most mixer implementations use some kind of multiplication of two signals, the signal to be up- or down-converted (IF or RF) and the signal whose frequency determines the output frequency (LO). So mathematically, an ideal mixer can be treated as a

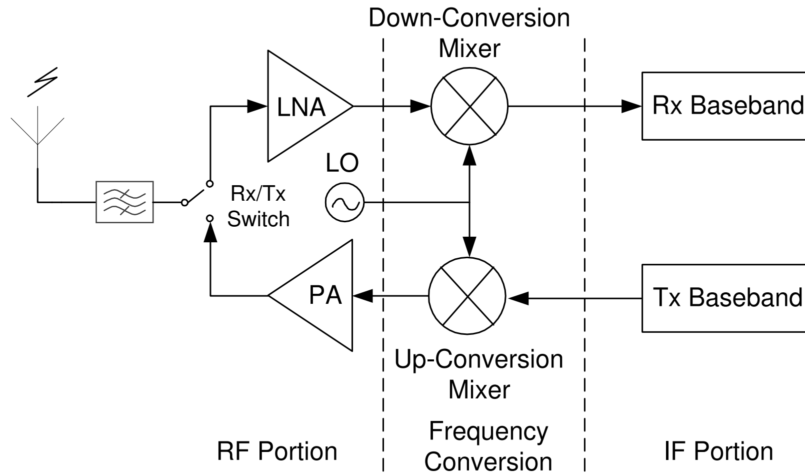



Fig. 21. Mixers perform frequency translation in communication system

multiplier as shown in Fig. 22.

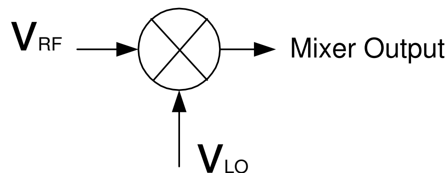



Fig. 22. Mixer mathematical model

Assume the RF input signal is $v_{RF}(t) = A_{RF}\cos(\omega_{RF}t)$ and the LO signal is $v_{LO} = A_{LO}\cos(\omega_{LO}t)$, then the output of an ideal multiplier is

$$v_{IF} = v_{RF}(t) \times v_{LO}(t) = \frac{A_{RF}A_{LO}}{2}\left[\cos(\omega_{RF} - \omega_{LO}) + \cos(\omega_{RF} + \omega_{LO})\right]$$

For an up-conversion mixer, component $\omega_{RF} - \omega_{LO}$ is filtered out. For a down-conversion mixer, component $\omega_{RF} + \omega_{LO}$ is filtered out. If the input signal occupies a frequency band, that band of frequency will be moved to a lower frequency for down-conversion or a higher frequency for up-conversion. More insights can be obtained by

considering this mixing model in frequency domain. Usually the ideal LO signal is a single tone, $v_{LO} = \cos(\omega_{LO} t)$. Here the amplitude of the LO is normalized to unity. The Fourier transform of the LO signal is

$$V_{LO}(\omega) = \pi\delta(\omega - \omega_{LO}) + \pi\delta(\omega + \omega_{LO}) \tag{3.1}$$

The RF signal resides within a certain band around its center $\omega_{RF}$. The amplification in time domain will become convolution in frequency domain, i.e., the RF signal will convolve with two $\delta$-functions in frequency domain. Fig. 23 demonstrates this process. It can be seen that the RF frequency band is translated into side-band, the lower side-band centered around $\omega_{RF} - \omega_{LO}$ and the upper side-band centered around $\omega_{RF} + \omega_{LO}$.

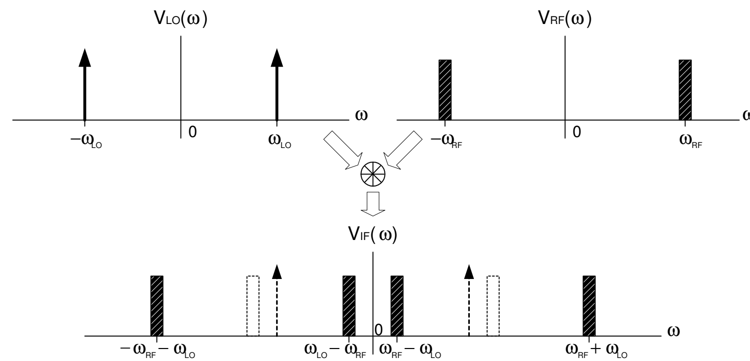



Fig. 23. Mixing process

B. Mixer Metrics

In order to evaluate the performance of mixers, several metrics are defined, they are: conversion gain or loss, noise figure, port isolations, linearity and power consumption, etc.

### 1. Conversion Gain or Loss

Conversion gain or loss is a measure of mixer efficiency, it is defined as the ratio of the desired IF output (voltage or power) to the RF input signal value (voltage or power). More specifically:

$$\text{Voltage conversion gain} = \frac{\text{r.m.s. voltage of the IF signal}}{\text{r.m.s. voltage of the RF signal}}$$

$$\text{Power conversion gain} = \frac{\text{IF power deliverd to the load}}{\text{Power availabe from the RF source}}$$

The power gain definition here is actually transducer power gain. Usually for a discrete implementation of mixer, power gain is specified. For on-chip implementations, people usually specify its voltage conversion gain.

Mixers can be active or passive. Active mixers are capable of providing voltage gain and power gain. At most a passive mixer can only provide voltage or current gain but not power gain. The parametric converter, which is a special kind of passive mixer, can provide power gain.

### 2. Noise Figure

A mixer's noise figure (NF) is the signal-to-noise ratio (SNR) at the input (RF) port divided by the SNR at the output (IF) port. Although the mixer's noise figure definition looks similar to the LNA's noise figure definition, there are some subtle differences between them.

The signal-to-noise ratios at the input and output are referred to different operation frequencies. Mixers perform frequency translation, so there are two different types of noise figures according to two different types of frequency conversion schemes, single-side band (SSB) noise figure and double-side band(DSB) noise figure.

For the DSB noise figure, the desired signal appears at the both sides of the LO

frequency. Fig. 24 gives the situation of the double-side band conversion. Assume the RF signal power for each side is $S_i$, the noise power for each side presented at the RF input is $N_i$, the conversion power gain is $G$, the mixer internal noise referred to its output is $N_{no}$, the noise factor of the double-side band mixer can be calculated as

$$F_{DSB} = \frac{S_i}{N_i} \times \frac{2N_iG + N_{no}}{2S_iG} = 1 + \frac{N_{no}}{2N_iG} \tag{3.2}$$

For the SSB noise figure, the desired signal only appears at one side of the LO

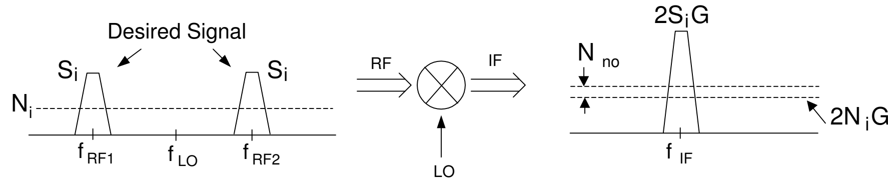



Fig. 24. Double-side band conversion

frequency. Suppose the noise at image frequency has not been removed, Fig. 25 shows this scenario. Using the same notation as the calculation of DSB noise factor, the SSB noise factor can be expressed as

$$F_{SSB} = \frac{S_i}{N_i} \times \frac{2N_iG + N_{no}}{S_iG} = 2 + \frac{N_{no}}{N_iG} \tag{3.3}$$

From (3.2) and (3.3),

$$F_{SSB} = 2F_{DSB} \tag{3.4}$$

It shows that the SSB noise figure is 3 dB higher than the DSB noise figure. When designing a mixer, it should be made clear which noise figure is targeted at. The simulator like SpectreRF gives the SSB noise figure. So if the DSB noise figure is required, a 3-dB difference should be subtracted manually.

In practical system, the image noise in a SSB case is filtered using an image

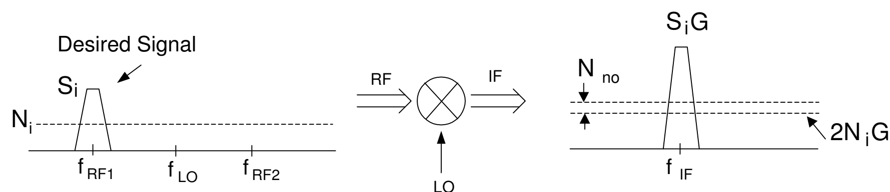



Fig. 25. Single-side band conversion

rejection filter (IRF) before the mixer or a complex filter after the mixer (see Fig. 26). So, the image noise experiences different gain with the noise within the signal band. Suppose the gain of the IRF or complex filter at image frequency is $G_{im}$, the noise factor is

$$F_{SSB,IRF} = \frac{S_i}{N_i} \times \frac{N_i G + N_i G_{im} G + N_{no}}{S_i G} = 1 + G_{im} + \frac{N_{no}}{N_i G}$$

With related to SSB or DSB noise factor,

$$\begin{aligned} F_{SSB,IRF} &= F_{SSB} - (1 - G_{im}) \\ & 2F_{DSB} - (1 - G_{im}) \end{aligned} \tag{3.5}$$

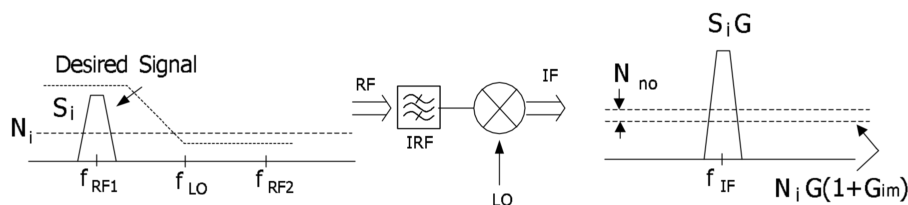



Fig. 26. Single-side band conversion with image rejection

### 3. Port-to-Port Isolation

In reality, signals always leak through different mechanisms from one port to another. The port-to-port isolation figure accounts for this unwanted transmission, it is defined as the ratio of the signal power available into one port of the mixer to the measured power level of that signal at one of the other mixer ports in a 50 Ω system.

The mixer has three ports. The desired transmission is from RF port at RF frequency to IF port at IF frequency. The leakage can occur among all the other ports. The most interested ones are discussed in detail as follows.

The signal leaked from the LO port to the RF port at a RF frequency (LO-to-RF leakage) will mix with the LO signal again, causing the so called self-mixing problem in direct conversion. Due to the non-zero reverse gain ($s_{12}$) of the LNA, the LO leakage may even reach the antenna through the LNA, then reflected back by the antenna/LNA interface (see Fig. 27).

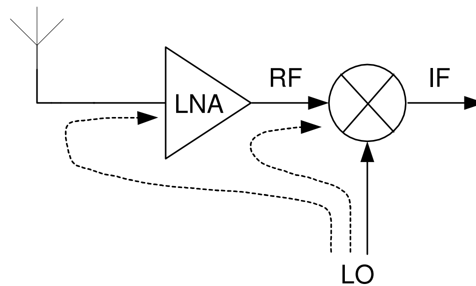



Fig. 27. LO-to-RF leakage

The LO-to-IF feed-through may cause desensitization of the block following the mixer which is usually a low pass filter. In order to guarantee a good mixing performance, the LO power is usually greater than the IF power. Although the LO frequency will be in the stop-band of the filter, the large LO signal will drive the

filter out of its desired operation region. Whereby it is desirable to have some passive filtering at the output of the mixer.

The RF-to-LO feed-through will allow the interferer and spurs present in the RF signal to interact with the LO. The RF-to-IF feed-through at the IF frequency may cause problems in direct conversion architecture due to the low-frequency even-order intermodulation product.

#### 4. Linearity Measurement

When a real mixer operates, not only do the desired tones mix (multiply) each other, their harmonics also experience the mixing process due to the non-linear characteristic of the mixing device. These unwanted products may fall into the spectrum band of the signal and degrade the desired signal. The non-linear behavior of the system is frequency dependent. We are interested in the dynamic linearity of the system, not just the DC non-linear characteristics.

In communication systems we usually use the following metrics to measure the mixer's linearity: 1-dB compression point ($P_{1\mathrm{dB}}$), second and third order intercept point (IP2 and IP3), spurious free dynamic range (SFDR) and compression free dynamic range (CFDR). The meanings of these terms for mixer are similar as for RF amplifiers, just be aware of the frequency conversion in the mixer.

### C. Circuit Topologies of Mixers

Down-converter and up-converter mixers are used in RF front-end to transform signal spectrum from one location to another. Linear, time-invariant systems can not generate spectral components not presented in the input. The mixer must be a non-linear or time-variant system. Virtually any nonlinear elements can be used as mixers. Some

nonlinearities just work better and more practical than others. The mixer can be passive or active depending on whether the mixer can provide conversion loss or gain. The mixer can also be unbalanced, single-balanced or double-balanced, depending on the signal operation symmetrical properties. Depending how the signal spectrum is transferred, mixers can be implemented to work in real signal domain or complex signal domain.

### 1. Diode Mixers

Diode mixers use the non-linearity of the diode to implement the frequency conversion. Fig. 28 shows the single-diode mixer.The single-diode mixer is the simplest and oldest passive mixer. The output LC tank is tuned to the desired IF, and the input is the sum of RF, LO and DC bias. This mixer can not provide any isolation between ports, neither can it achieve conversion gain. However, at very high frequencies (e.g. millimeter-wave band) this kind of mixer is extremely useful.

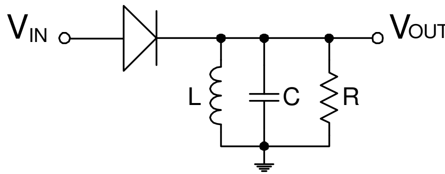



Fig. 28. Single-diode mixer

A single-diode mixer belongs to the unbalanced configuration. The single-balanced diode mixer uses two diodes as shown in Fig. 29. LO is large enough to make the diodes work as switches, regardless of the level of the RF signal. When the diodes are on, RF and IF are connected together, so the RF-IF isolation is poor. But the RF signal is common-mode for the transformer, so the RF-LO isolation is excellent.

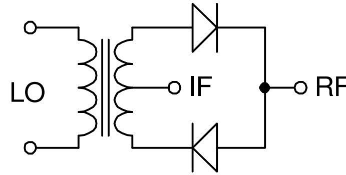



Fig. 29. Single-balanced diode mixer

A double-balanced diode mixer uses four diodes. Due to the symmetry of the circuit, isolations between each pair of ports are excellent, mainly limited by the device matching. The diode mixer is almost completely linear and the upper limit of the dynamic range is constrained by diode break-down. Typically, double-balanced mixers can achieve conversion loss of around 6 dB, and port isolations of at least 30 dB. Fig. 30 shows the typical configuration of this type of mixer.

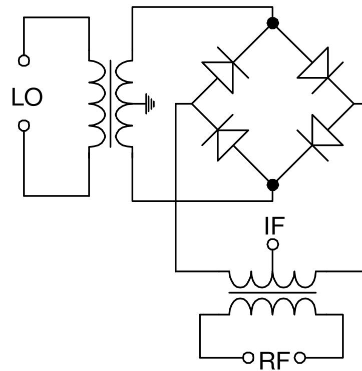



Fig. 30. Double-balanced diode mixer

2. Passive Mixer in CMOS Technology

This kind of mixer is more like the double-balanced diode mixer. The diodes are replaced by MOS transistors working as switches as shown in Fig. 31. MOS transistors

M1-M4 are working as switches and are driven by a large LO signal in the anti-phase. Thus, only one diagonal pair of transistors are turned on at any given time. More specifically, when the LO is high, M1 and M4 are on, $V_{IF}$ equals to $V_{RF}$, and when the LO is low, M2 and M3 are on, $V_{IF}$ equals to $-V_{RF}$. So it is equivalent to treat the mixing behavior as multiplication: the RF signal is multiplied by a square wave whose amplitude is either +1 or -1 and whose frequency is that of the LO signal.

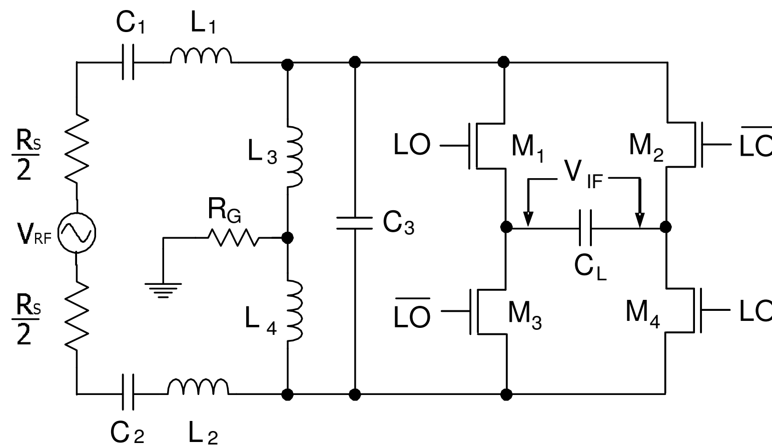



Fig. 31. CMOS passive double-balanced mixer

In Fig. 31, the input LC network provides matching and filtering. And due to the reactive matching network, the voltage conversion gain can be greater than 1, but this mixer can not provide power conversion gain. The noise figure and IIP3 are strong functions of the LO signal strength and depend on how the MOS transistors are switched. The implementation of a passive mixer is usually simple and there are no stability issues involved [13].

### 3. Gilbert-cell Mixer

Gilbert cell mixers are the most popular types of integrated mixers. They can be implemented either in BJT or CMOS technology. They are usually double-balanced mixers. Fig. 32(a) shows one possible implementation using MOS transistors. Good port isolation (40∼60 dB) can be achieved as a result of circuit symmetry. If the input signals (RF and LO) are small enough such that all the transistors in Fig. 32(a) are working in their linear region, then the Gilbert-cell will behave like an analog multiplier. But this is not an efficient way to perform the RF frequency conversion using a Gilbert-cell, it will generate a prohibitive high noise figure and has a strong LO dependence of conversion gain.

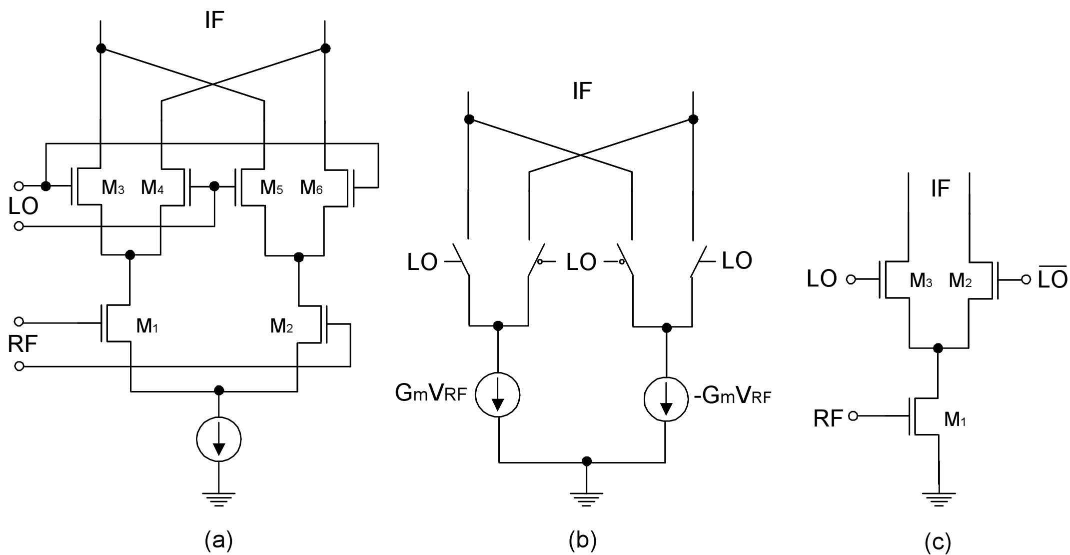



Fig. 32. Gilbert-cell mixer (a) transistor implementation (b) working principle and (c) single-ended version

In principle, the Gilbert-cell mixer works as shown in Fig. 32(b). M1 and M2 work as a voltage to current (V-I) converter or transconductance stage, and M3∼M6 are driven by a large enough LO signal such that they work as current commuting

switches. The linearity of the mixer is limited by the linearity of the V-I converter. Additional linearization techniques are usually applied to the V-I converter to improve the linearity of the mixer. For direct conversion and low IF, the noise figure is limited by the flicker noise of the current switches and for higher IF, the noise figure is limited by the thermal noise of the circuit. The transconductance conversion gain can be expressed as

$$G_c = \alpha \frac{2}{\pi} g_m \tag{3.6}$$

where $g_m$ is the transconductance of the V-I converter. Factor $\alpha$ is referred to as switching efficiency. Because the practical mixer's switches are usually not ideal, loss is introduced due to non-ideal switching and is modeled by $\alpha$, which is usually smaller than its ideal value one. In some cases, single-balanced mixer can fulfill the requirement, then half of the Gilbert-cell can be used as shown in Fig. 32(c). This mixer consumes half of the power, has single-ended input and differential output. It is easy to interface with single-ended LNA.

### 4. Sub-Sampling Mixer

A properly designed track-and-hold circuit can work as a sub-sampling mixer. Fig. 33 shows a general structure of this kind of mixer. One big advantage of this circuit is that the LO signal is clocked at a relative low frequency, which eases the design of LO circuitry. But the sampler still must have a good time resolution which means the clock absolute time jitter must be a tiny fraction of the carrier period. Due to the sub-sampling, the noise folding effect makes the mixer present a large noise figure. The linearity of the sub-sampling mixer is usually very high, but due to jitter noise and thermal noise folding, its dynamic range may be inferior to other carefully designed mixers. Finally, the sub-sampling performance will be limited by the performance

of the operational amplifier (Opamp) used in the sampler. The limited achievable gain-bandwidth (GBW) of the Opamp will limit the RF input signal to a certain range.

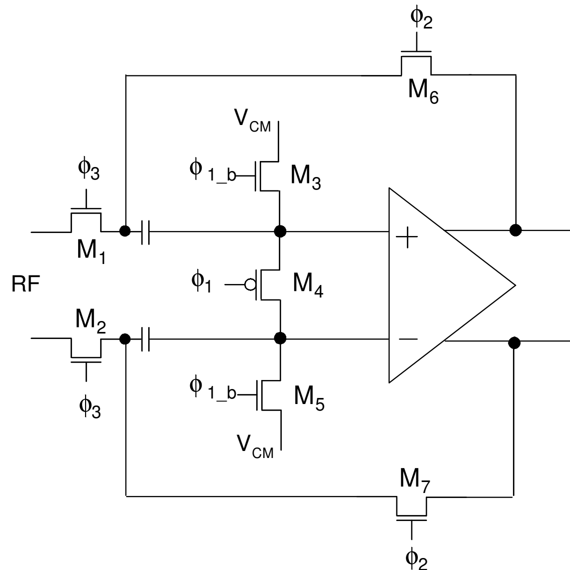



Fig. 33. Sub-sampling mixer

## 5. Harmonic Mixer

Unlike conventional mixers, harmonic mixers mix RF signals with the second or higher order harmonic of the LO signal. For standard mixers, the mixing product of interest is $LO \pm RF$. For harmonic mixers the $nLO \pm RF$ signal is desired. Usually $n = 2$ and the mixer is called sub-harmonic mixer.

The idea of harmonic mixing is to use the even harmonics of LO for its conversion product. The odd harmonics including its fundamental will be rejected either due to the odd symmetry of the system or by filtering or both.

One direct benefit of this idea is that the LO can run at half rate, which makes

VCO design easier. But because of the harmonic mixing, the conversion gain is usually small (several dB) and the noise figure is high. Harmonic mixers are attractive for the direct conversion applications due to the fact that they have low self-mixing DC offset.

Fig. 34 is one possible implementation of the harmonic mixer [14]. Two emitter-coupled BJT pairs work as two limiters. The odd symmetry of their transfer function suppress even order distortion including LO self-mixing. The small RF signal will modulate the zero crossing point of the relatively large LO signal. The output of the mixer is a rectangular wave in pulse width modulation fashion, a low pass filter will demodulate the signal. The leaked LO signal will generate a position modulated signal which can not be demodulated by the low pass filter. So ideally this mixer will not have DC offset.

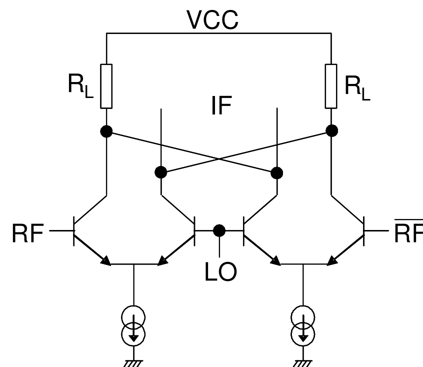



Fig. 34. Harmonic mixer

## D. Mixer Design for a Low-IF Bluetooth Receiver

Bluetooth is a wireless technology for small form-factor, low-cost, short-range radio links between mobile PCs, mobile phones, PDAs and other portable devices. It oper-

ates at the 2.4 GHz Industrial Scientific Medicine(ISM) band. Its modulation format is Gaussian frequency shift keying (GFSK) with an index of 0.28∼0.35. The data rate is 1 Mb/s and the channel spacing is 1 MHz. Low cost and possible system-on-chip (SoC) solutions are the most attractive features of Bluetooth technology. The digital system is an important part of the system and digital technology benefits most by using CMOS process. So the CMOS process is a must for SoC implementation. Comparing to the BiCMOS technology, CMOS process is preferred in the Bluetooth system design because it provides inexpensive implementation at a high integration level. Thanks to the recent development in the silicon fabrication technique, the current CMOS technology is able to accommodate the high speed RF circuit design in the GHz frequency band and hence can be applied in the Bluetooth receiver design. Therefore, TSMC 0.35 $\mu$m CMOS technology is chosen to realize the Bluetooth receiver system.

For a fully integrated CMOS implementation, there are two architecture candidates, direct conversion and low-IF. Direct conversion has the most possible high level integration and does not require image rejection. It needs less components and can achieve low power consumption. The difficulties faced by direct conversion are DC offset and flicker noise [15]. While DC offset can be removed by some dedicated circuitry, the flick noise presented in the mixer prevents one from adopting this architecture. The process, TSMC 0.35 $\mu$m CMOS, used in this implementation has flicker noise corner frequency around 1∼2 MHz. It is not possible to make the mixer's noise figure meet the specification for direct conversion. A low-IF receiver can also achieve a high level integration and has a possible low power requirement with careful system level and circuit level design [16]. For low-IF, flicker noise is less significant in the signal band. The DC offset can be easily removed with relatively simple circuitry. But one needs to consider the image rejection and folded-back interference problem.

So low-IF architecture is a better choice for this implementation. The IF frequency is decided to be 2 MHz due to the high flicker noise corner frequency. The image and fold back interferer is rejected by a complex filter following the mixer.

### 1. Low-IF Bluetooth Receiver Architecture

Fig. 35 shows the low-IF Bluetooth receiver block diagram. It has two identical mixers for the I and Q branches. The mixers in this receiver convert the 2.4 GHz frequency band RF signal to a 2 MHz band IF signal, then the down-converted I and Q signals are fed to a complex filter to make image/interferer rejection and channel selection. The specifications for the mixer required in this receiver are listed in Table VI.

Table VI. Mixer specifications for low-IF Bluetooth receiver

| Parameters | min. | typ. | max. | Unit |
|---|---|---|---|---|
| RF input frequency range | 2400 | | 2480 | MHz |
| LO input frequency range | 2400 | | 2482 | MHz |
| IF output frequency range | - | 2 | - | MHz |
| Voltage conversion gain | - | 12 | - | dB |
| LO drive level (internal) | - | 0 | - | dBm |
| Input 1dB compression point | - | 0.1 | - | dBm |
| IIP3 (Vrms) | - | 5.6 | - | dBm |
| Input referred noise | - | - | 7.5 | nV/$\sqrt{Hz}$ |

### 2. Implementation of the Down-Conversion Mixer

For the front-end of the Bluetooth receiver, the LNA is designed by designer Wenjun Sheng using the conventional source inductive degeneration technique [17]. The down-

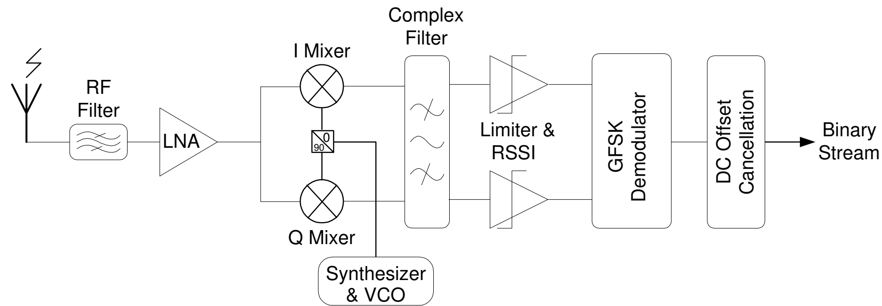



Fig. 35. The proposed low-IF Bluetooth receiver

conversion mixer is a modified Gilbert-Cell mixer [18], as shown in Fig. 36. In order to improve the output dynamic range, the tail current in the conventional Gilbert-cell is removed, this also allows low voltage operation and has higher linearity. It is obvious that without the tail current, the voltage headroom will be improved by a $V_{dsat}$ of a MOS transistor if a simple current source was used. The improved linearity without current source can be explained as below.

For a source coupled differential pair, in order to operate both of the two transistors in the saturation region, the input differential voltage should be within [19]

$$|V_{id}| < \sqrt{2}V_{od} \tag{3.7}$$

where $V_{od}$ is the gate over-drive voltage when $V_{id} = 0$, i.e. at quiescent DC biasing point. But if there is no tail current source, for the same DC biasing current, the differential input range for active operation will be expanded to

$$|V_{id}| < 2V_{od} \tag{3.8}$$

The differential pair has less linear input range due to the existence of the tail current source. The current source makes the two transistors in the differential pair

Fig. 36. Bluetooth receiver down-conversion mixer

interact with each other so the their active input region is limited more than two single transistors without such interaction. However, without tail current source, the circuit is not a truly differential structure any more which may have a poor common-mode rejection performance. Fortunately, if the LNA's output signals are well balanced and/or additional loop is used to suppress the common-mode signal, this architecture can be used.

The injected current $I_B$, called bleeding current, allows separate control of current

flowing through the drive stage ($M_1$ and $M_2$) from the current switches ($M_3 \sim M_6$). For the same drive bias current, bleeding reduces the current flowing through the current switches and load resistance. Since flicker noise can be expressed as [20]

$$i_{nf}^2 = K_f \frac{I_D^{A_f}}{C_{ox} W_{eff} L_{eff} f} \Delta f \tag{3.9}$$

Reducing the bias current will significantly reduce flicker noise contributed by the switches and allow large load resistors which increases the conversion gain. Furthermore, with current bleeding, the switches can work with smaller gate to source voltage, so for a given LO signal level, smaller charges are necessary to turn on or off the switches and conversion efficiency is improved.

On the other hand, bleeding can degrade the high frequency performance of the RF drive stage, because bleeding reduces the bias current of the switches therefore increasing the load impedance seen by the driving transistors. Ideally, the RF drive stage works as a voltage-current converter, it should see as small impedance at its output as possible, increased load impedance reduces voltage-current conversion efficiency. Due to Miller effect, increased voltage gain from the input of the RF stage to its load makes reverse isolation larger than without bleeding.

Current bleeding also adds additional noise sources due to the bleeding circuit, which is usually implemented using active devices. In this implementation, the bleeding current source is formed by a PMOS current source with a large device size. This arrangement makes the flicker noise contribution from the bleeding source smaller. Care must be taken to make the matching between the two bleeding current sources as good as possible. Otherwise, a large mismatch will drive the injection node towards the positive or negative power supply, if the switch current is too small to set the drain voltage of $M_1$ or $M_2$. Although additional common mode feedback circuits can be used to prevent this problem [21], it will require an additional tail current

source which is removed for the limited voltage headroom reason. Since the load of the mixer is resistive, the need of a common mode feedback circuit at the mixer's output is avoided.

The voltage conversion gain can be written as

$$A_{vc} = \frac{2}{\pi} g_{m,RF} R_L \tag{3.10}$$

where $g_{m,RF}$is the differential transconductance of $M_1$ and $M_2$. IF $M_1$ and $M_2$are perfectly matched, then $g_{m,RF} = g_{m,M_1} = g_{m,M_2}$, otherwise for the first order approximation, assume $\Delta g_m = |g_{m,M_1} - g_{m,M_2}| \ll g_{m,RF}$,

$$g_{m,RF} \approx \frac{g_{m,M_1} + g_{M,M_2}}{2} \tag{3.11}$$

IIP3 of the mixer is approximately given by [22]

$$V_{IIP3} \approx 2\sqrt{\frac{4}{3}\frac{g_{m,RF}}{K\theta}} \approx 4\sqrt{\frac{2}{3}\frac{V_{od}}{\theta}} \tag{3.12}$$

where $K = \frac{1}{2}\mu_{eff} C_{ox} \frac{W}{L}$, $\theta = \frac{1}{E_{sat} L}$ and $V_{od} = V_{GS} - V_{th}$. It is assumed that $\theta V_{od} \ll 1$. For a regular differential pair with tail current source, and using the same transistor size and bias condition, its IIP3 can be found to be [22]

$$V_{IIP3} \approx 4\sqrt{\frac{2}{3}} V_{od} \tag{3.13}$$

which is by a factor of $\frac{1}{\sqrt{\theta V_{od}}}$ smaller than the one used here without a tail current source. It is obvious from (3.10) and (3.12) that an increase in the drive stage transconductance will increase the conversion gain and linearity at the same time. But increasing $g_m$ will result in a larger power consumption. The noise analysis of the mixer is quite involved and can be found in [23].

The mixer's design flow is shown in Fig. 37. Because the signal reaches the mixer's input after being amplified by the LNA, we start with the linearity specification for

the RF driving stage. From (3.12), the required gate over-drive voltage $V_{od,RF}$ can be calculated, then from (3.10), the device size $W_{RF}$, bias current $I_{RF}$ and load resistor $R_L$ can be determined. The bleeding current is chosen by considering the output voltage headroom. In order to allow the current switches working efficiently, the LO swing should be larger than the switch differential pair's linear operation voltage range, which is $\sqrt{2}V_{od,SW}$. Therefore, the gate over-drive of the current switches $V_{od,SW}$, and their size $W_{SW}$ can be calculated. After the hand calculations, simulation verification is carried out. Usually the design parameters needs to be adjusted intensively through simulation and several iterations are required to achieve the targeted specifications.

### 3. Layout Considerations and Simulation Results

When doing the layout of the mixer, special care should be taken to ensure the I and Q mixers are symmetrical and the gate resistance is minimized. Matching techniques such as common centroid and inter-digitized pattern are applied. The length of the poly gate is kept short enough to reduce the gate resistance (large gate resistance will degrade the noise performance). For the layout of a poly-poly capacitor, if the bottom plate is floating, the parasitic capacitance from the bottom plate to the substrate should be considered. It is about 40% of the nominal capacitance. Decoupling capacitors may be needed to prevent the circuits from oscillation. Metal should be wide enough to carry large current. The current density allowed through metal is about 1 mA/$\mu$m. Guarding rings are place around the circuits to improve isolations from other blocks on the same die. The mixer uses resistors as load, guidelines for reducing resistor mismatch are: 1) Use dummy poly lines when possible. 2) Maximize the number of contacts per width of the device. 3) Avoid partial coverage of resistors by any layer of metal. 4) Local symmetry is important for matching, in a bank of

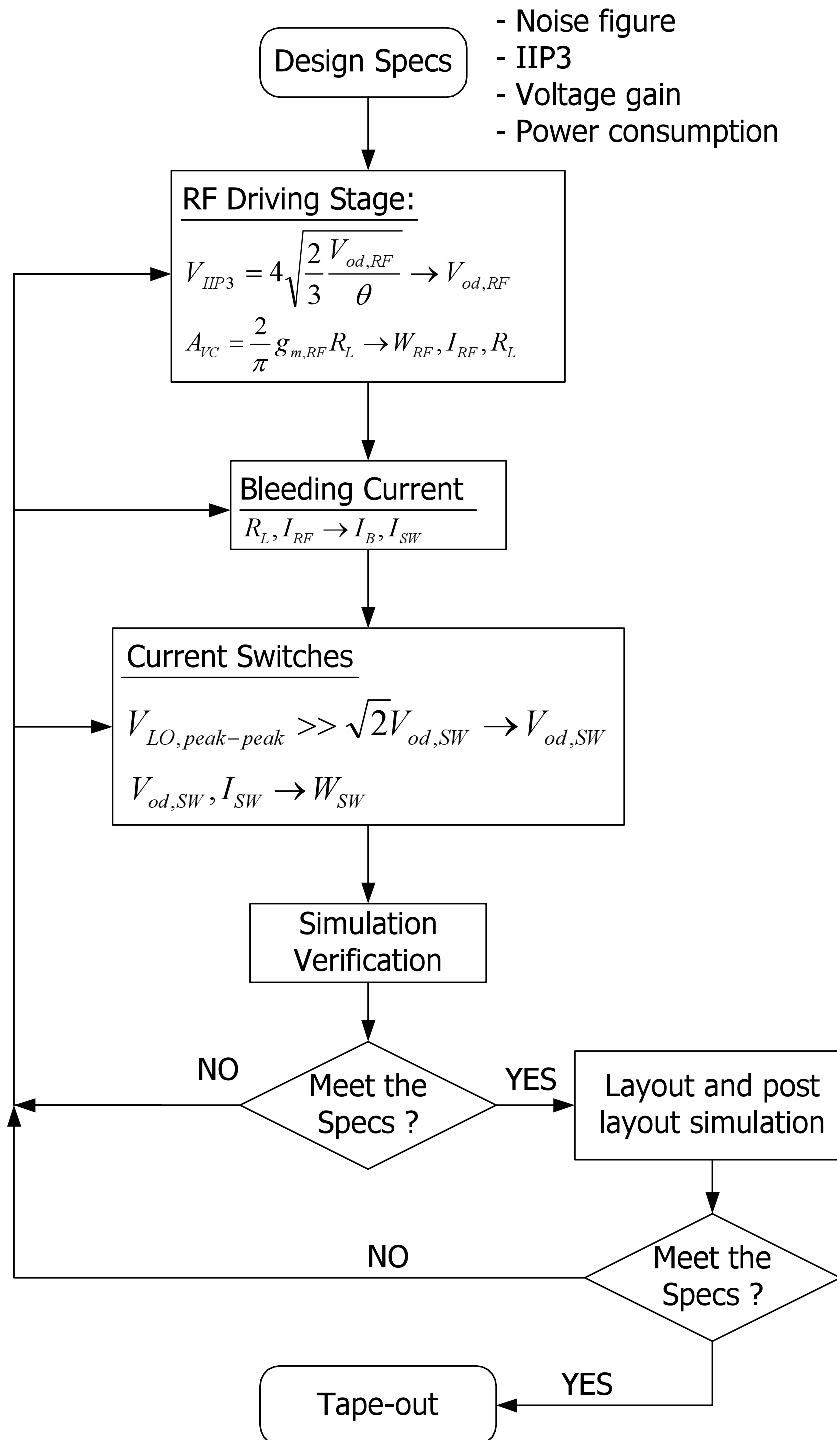



Fig. 37. Mixer design flow chart

identical resistors, resistors at the edge showed about 2% difference in value from their nearest neighbors. 5) Gradient effects are present and do effect matching, keep critically matched pairs as close as possible.

Fig. 38 is the layout of the mixer, it includes both I and Q branches. The circuit is simulated using Cadence SpectreRF environment. Because the mixer will be interfaced with LNA directly, its input impedance is first obtained around its operation frequency. And is given to the LNA designer. The LNA and mixer co-simulation is done to make sure they are compatible. Mixer's separate performance is than re-simulated and listed in Table VII.

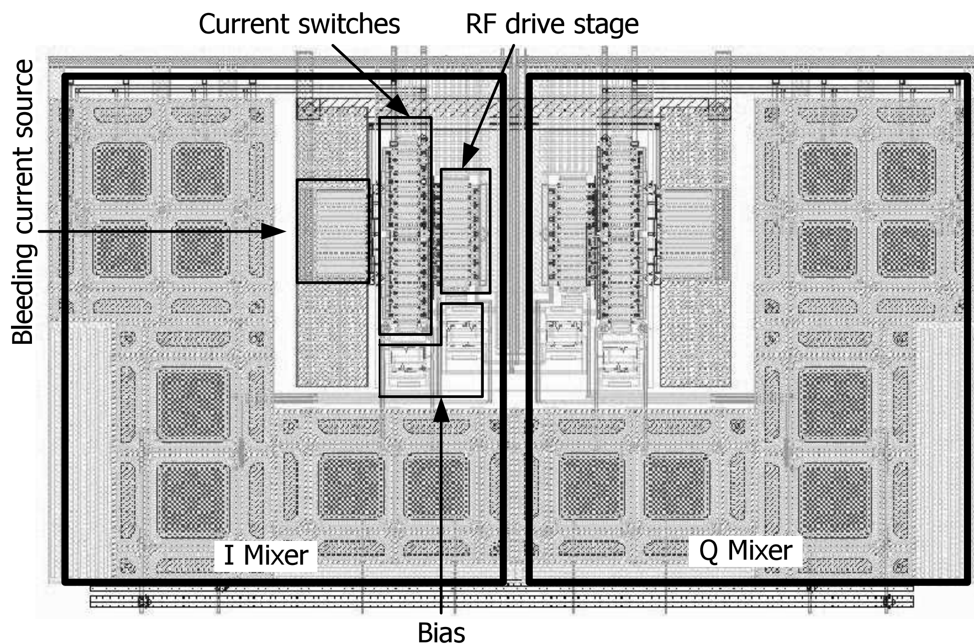



Fig. 38. Layout of the down-conversion mixers

Table VII. Bluetooth mixer simulation results

| Parameter | Value | Unit |
|---|---|---|
| Voltage conversion gain | 12.4 | dB |
| RF Input resistance | 214 | $\Omega$ |
| RF Input capacitance | 166 | fF |
| Input referred noise | 6.8 | nV/$\sqrt{Hz}$ |
| RF-LO isolation | 30 | dB |
| LO swing | 0 | dBm |
| Supply Voltage | 3 | V |
| Current consumption | 5 | mA |

### 4. Experimental Results of the Mixer Within the Receiver

The Bluetooth receiver IC is fabricated using the TSMC 0.35 $\mu m$ standard CMOS process through MOSIS service and packaged in a 48-pin TQFP plastic package. The die microphotograph is shown in Fig. 39, and it occupies 2.5 $mm \times 2.5$ $mm$ silicon area.

As an integrated part of the Bluetooth receiver, the mixer is tested together with the LNA. In order to guarantee the performance, no access to the internal high frequency nodes within the LNA and mixer is granted. Therefore, the LNA and mixer have been tested as a single block. Together they consume 10 mA current from a 3 V power supply.

The measured cascaded NF and voltage gain are 8.5 dB and 25 dB, respectively. Fig. 40 shows the input S11 measurement plot of the front-end. The S11 is less than -10 dB in the 2.4GHz band. The system IIP3 is about -10 dBm. Fig. 41 is the IIP3

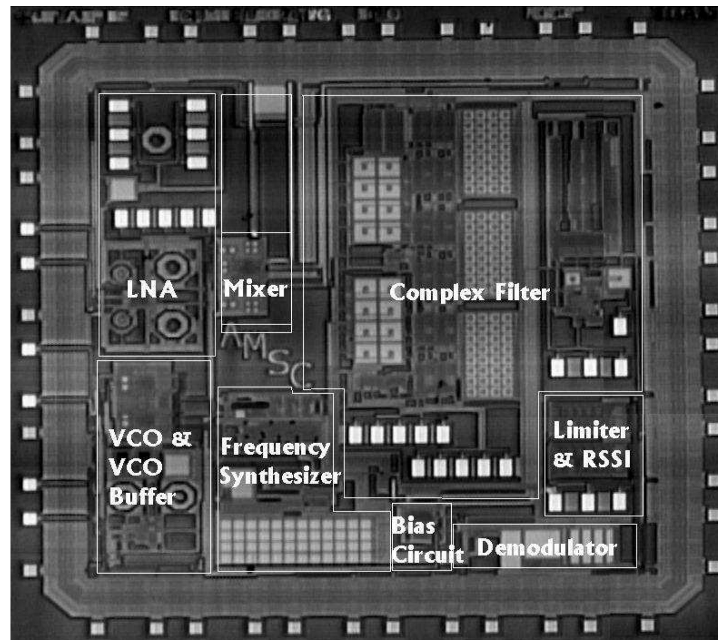



Fig. 39. Die photo of Bluetooth receiver

plot of the receiver system. The IIP3 was tested using two tones 3 MHz and 6 MHz away from the desired signal. The 3rd order intermodulation product was observed at the output of the complex filter and referred back to the receiver input to calculate IIP3. The achieved sensitivity is -82 dBm for the whole receiver.

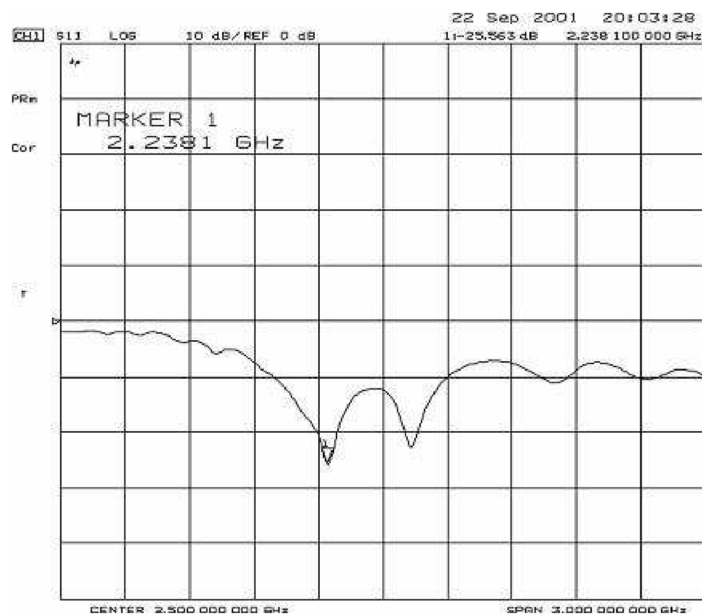



Fig. 40. Input return loss of the Bluetooth front-end

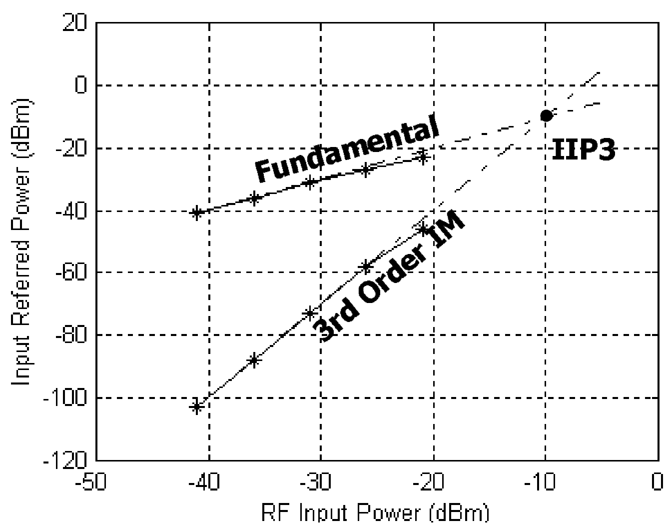



Fig. 41. Measured IIP3 of the Bluetooth receiver

# CHAPTER IV

## BLUETOOTH/WI-FI DUAL-STANDARD RECEIVER RF FRONT-END

With the rapid development of wireless access technique, more and more devices are Bluetooth enabled. At the same time, the wireless LAN (IEEE 802.11b, Wi-Fi) standard comes into our everyday life. While Bluetooth targets on short range cable-replacement and point-to-point communication applications [24], Wi-Fi is focused on wireless local network connection capabilities for portable devices such as laptops and PDAs [25]. These two standards do not compete with each other, but rather complement each other [26]. They do different things and tasks. There are circumstances where Bluetooth is preferable to Wi-Fi, for example, short range point to point or multi-point data transfer with low power consumption, data self-synchronization and updating. But for more complicated network functionalities and high speed data transfer, Wi-Fi will take over. One big application is wireless Internet access. All those functions are needed in high-end laptop and PDAs. So there are practical needs to require both standards within one application. Combining these two standards together into one single chip will reduce the fabrication cost and add more functionalities to the product.

Both Bluetooth and IEEE 802.11b standards work at the 2.4 GHz ISM band. Although the base band will be quite different between these two standards, the co-band operation makes the RF front-end having maximum compatibility. And it is reasonable to integrate the two standards in a single-chip with maximum circuit building block sharing. Table VIII lists the key features of both standards.

This chapter focused on the development of a dual-mode receiver RF front-end. In the following sections, the direct conversion Bluetooth/Wi-Fi dual-standard re-

Table VIII. Bluetooth and Wi-Fi standards key features

| | Bluetooth | Wi-Fi |
|---|---|---|
| Frequency band | 2.4-2.48 GHz | 2.4-2.48 GHz |
| Available speed | 720 Kbps | 5.5 Mbps |
| Operation range | 10 cm-10 m | 100 m |
| Security risk | Low | High |
| Targeted application | Wireless PAN | Wireless LAN |
| Form factor | Small | Large |
| Power requirement | Low | High |

ceiver architecture is discussed, the design considerations and circuits implementations of the RF front-end is presented as well as the measurement results.

### A. Direct Conversion Bluetooth/Wi-Fi Dual-Standard Receiver

As stated above, Wi-Fi belongs to the wireless local area network standard family. It operates at the same frequency band as Bluetooth, and there is a demand to combine these two standards together to make the product more competitive. In this design, direct-conversion architecture was explored for both standards to avoid the image rejection problem in IF architecture, maximizing block sharing between two standards to save power and silicon area. A major concern in the low-IF architecture is the selection of the intermediate frequency (IF). In order to benefit the most from the low-IF architecture, the IF frequency is usually optimized to suite the modulation format and signal bandwidth defined in a particular standard. Therefore, the IF will vary as the receiver switches between the standards to avoid significant degradation

of receiver performance. As a consequence, the channel select filter, whose center frequency is located at the IF, becomes difficult to design. In the direct conversion receiver architecture, the channel select filter will be just a low pass filter. It does not have a center frequency to adjust when the receiving mode changes. Furthermore, the signal bandwidth is also different from one standard to the other. The implementation of a bandwidth variable filter is more straightforward for low pass than that for band pass.

Although the direct conversion receiver enjoys the high level integration and potential of low power consumption, it suffers from some inherent implementation difficulties, such as DC offset and low frequency noise. The large DC offset and low frequency flick noise make it difficult to design a direct conversion receiver in current CMOS technologies for narrow band systems. Additional circuits and techniques, such as a digital calibration circuit, have to be used to alleviate the performance degradation due to the large DC offset. While there is no obvious way to solve the flick noise problem in CMOS technology for narrow band systems like Bluetooth. BiCMOS technology, on the other hand, has much lower intrinsic DC offset and flicker noise. Simple measures, like a 2nd or 3rd order high pass filter, can provide enough suppression to the DC offset and flick noise. Therefore, BiCMOS process is chosen for this dual-standard direct conversion receiver to relax design complexity for a reasonable increase of cost.

Fig. 42 is the block diagram of the Bluetooth/Wi-Fi dual-standard receiver. In this receiver, the front-end blocks, LNA and mixer, are completely shared between Bluetooth and Wi-Fi standards. The received signal in the Wi-Fi standard can be 10 dBm stronger than the Bluetooth signal. To accommodate the larger dynamic range of the 802.11b received signal, the LNA has two gain modes. It switches between a 15 dB gain or a 15 dB attenuation according to the received signal power.

An integer-N frequency synthesizer is used to generate two times the desired local oscillation (LO) frequency to avoid the LO pull-in problem in a transceiver due to the coupling from the power amplifier to the VCO. The output of VCO is then passed through a divided-by-two circuit to generate quadrature LO signals for the I and Q channels. The received signal is converted to DC by the down-conversion IQ mixer. A $G_m - C$ low-pass channel selection filter with programmable bandwidth is used to accommodate both standards. Then the signal is conditioned by the variable-gain amplifier and sent to analog-to-digital converter (ADC). The ADC is a parallel pipelined structure. For Bluetooth, its sampling rate is 10 MHz, and has an 11-bit resolution. For Wi-Fi mode, the sampling rate is 44 MHz, and its resolution reduces to 8-bit. DC offsets cancelation technique is used to remove both static and dynamic offsets [27].

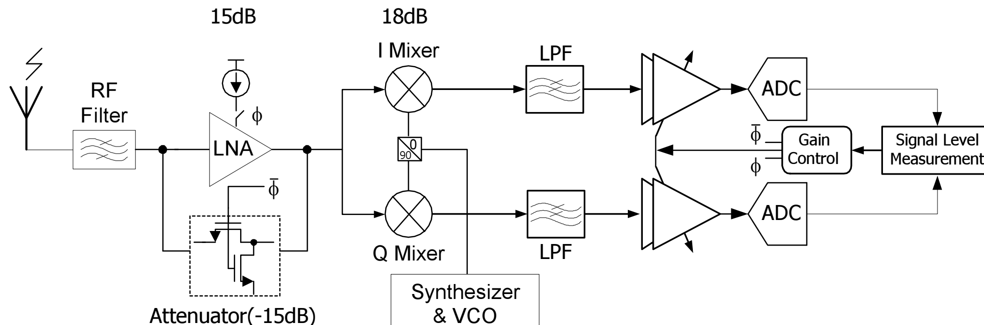



Fig. 42. Bluetooth/Wi-Fi receiver

## B. RF Front-End Design Considerations

For this design, a fully-differential structure is used to reject the common-mode noise coming from the digital parts and other sources as much as possible. The differential structure can also isolate the influence of bond-wire. For a single-ended circuit, bond-

wire for signal grounds directly becomes part of the circuit and must be modeled well in order to have an accurate prediction in simulation. In a differential structure, bond-wire is in series with the tail current and is in both DC and common-mode path, so it does not or has little effect on the circuit differential operation. Thus its performance is more reliable and predictable. The differential structure consumes as twice of the power as single-ended one for the same performance, but the improved isolation and reliability justifies the power increase.

For a differential LNA, the desired operation mode is the differential mode. But the common mode operation also presents. An important issue or consideration is the common mode stability. Fig. 43 shows the source degenerated differential LNA input stage, its differential half-circuit and its common mode equivalent half-circuit. In the figure, $L_g$ is the gate inductor, $L_s$ is the source degeneration inductor, they together resonate with gate capacitor $C_1$. $C_1$ is the total capacitance between the gate and source, it is composed of gate-source overlapping capacitance $C_{ov}$, gate-source channel capacitance and additional capacitance added externally. $2C_2$ is the parasitic capacitor of the current source. For the differential operation, this capacitor does not appear in the half circuit. For the common mode, half of this capacitor will present in series with the inductor $L_s$. The resistance of the current source is assumed to be sufficiently large.

For the differential half-circuit, the input impedance [10] can be expressed as

$$Z_{in,diff} = j\omega\left(L_g + L_s\right) + \frac{1}{j\omega C_1} + \frac{g_m}{C_1}L_s \tag{4.1}$$

At the resonate frequency $\omega = \frac{1}{\sqrt{(L_g+L_s)C_1}}$, it is a pure positive resistance $\frac{g_m}{C_1}L_s$, where $g_m$ is the small signal transconductance of transistor $M_1$. So there is no obvious stability problem for the differential operation.

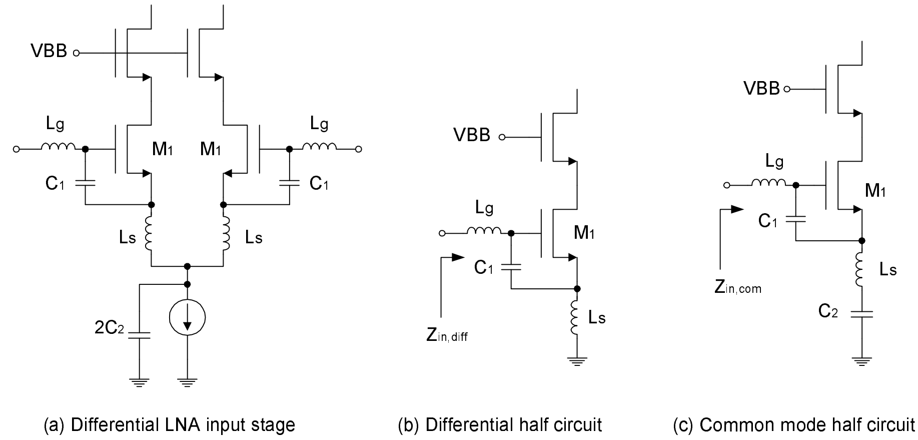



Fig. 43. Differential LNA common mode stability

For the common mode half-circuit, the input impedance is given by

$$Z_{in,com} = j\omega\left(L_g + L_s\right) + \frac{C_1 + C_2}{j\omega C_1 C_2} + \frac{g_m}{C_1} L_s - \frac{g_m}{\omega^2 C_1 C_2} \tag{4.2}$$

The real part of $Z_{in,com}$is

$$R_{in,com} = \frac{g_m}{C_1}\left(L_s - \frac{1}{\omega^2 C_2}\right) \tag{4.3}$$

It can be seen that if this real part goes negative, then there exists potential instability. One may want to increase the capacitance of $C_2$ in order to keep $R_{in,com}$ greater than zero. This is not the right way. A large common-mode capacitance will disturb the differential operation and even worse, increase the common-mode gain. We should keep this capacitance as small as possible. For an ideal tail current source of the differential pair, the common-mode gain is infinity, so the oscillation will not be able to sustain. Of course one should check the sign of (4.3) and further evaluate if this negative resistance will cancel the source impedance's real part thus causing a problem.

Another concern is the interface between the LNA and mixer. We have a direct conversion receiver. There is no image rejection filter between the LNA and mixer, so the LNA does not need to drive $50\,\Omega$ impedance and the mixer does not have to provide $50\,\Omega$ input matching. This is another degree of freedom for the direct conversion front-end design. The design strategy is that the mixer will be designed first. Then the mixer input impedance becomes known and will be driven by the LNA. Another way is to have a rough estimate of the impedance level of the mixer input and then the LNA and mixer can be designed at the same time by different designers. The LNA and mixer must be brought together to verify their interface behavior. For off-chip interconnection, the effect of bond wires on the circuit must be modeled and considered with the circuit design.

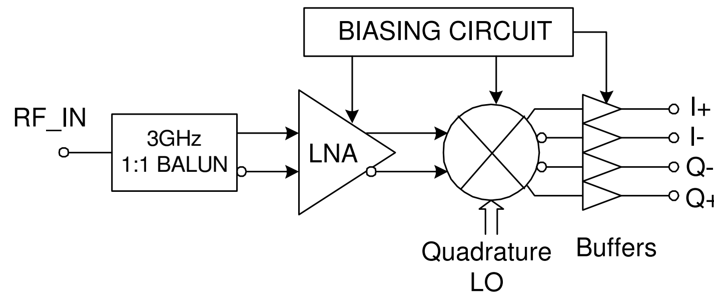



Fig. 44. RF front-end block diagram of BT/Wi-Fi Receiver

## C. Circuits Implementations

The dual-mode RF front-end includes the LNA, mixer, buffers and their biasing circuit. Fig. 44 gives the block diagram of the RF front-end. This is a direct conversion front-end, so no image rejection filter (IRF) is required between the LNA and mixer. This improves the aspect of integration of this architecture.

Table IX lists the specifications of the LNA and mixer for the Bluetooth and Wi-Fi operation. Notice that Wi-Fi has the most stringent requirement. So the RF front-end is shared by both the Bluetooth and Wi-Fi modes and targeted at the most stringent specifications in the table.

Table IX. Block specifications for the BT/Wi-Fi RF front-end

| Parameters | LNA | Mixer | Unit |
|---|---|---|---|
| Voltage Gain | 15/15 | 18/18 | dB |
| Noise figure | 2/2 | 25/15 | dB |
| IIP3 | -8/-8 | 0/0 | dBm |
| IIP2 | 20/20 | 40/30 | dBm |

### 1. LNA Implementation

The SiGe BiCMOS technology is used to implement the front-end. The LNA is given in Fig. 45. It is an inductive degenerated differential structure and can provide two gain steps of 15dB and -15dB. Gain control is implemented by a differential attenuator built around the LNA as indicated in the dashed box in Fig. 45. The attenuator is formed using NMOS transistor $M_5 \sim M_9$. For high gain mode, all the NMOS transistors will be turned off by connecting their gates to ground. Thus the normal operation of the LNA will not be affected. For low gain operation, these transistors are driven into their triode region and the LNA's bias current $I_{tail}$ will be cut off. Under this mode, the LNA itself will not consume current and is by-passed. The resistor divider formed by the NMOS transistors is used to provide further attenuation (-15 dB). A capacitor $C_m$ is inserted into the attenuator to improve impedance matching for the low gain mode.

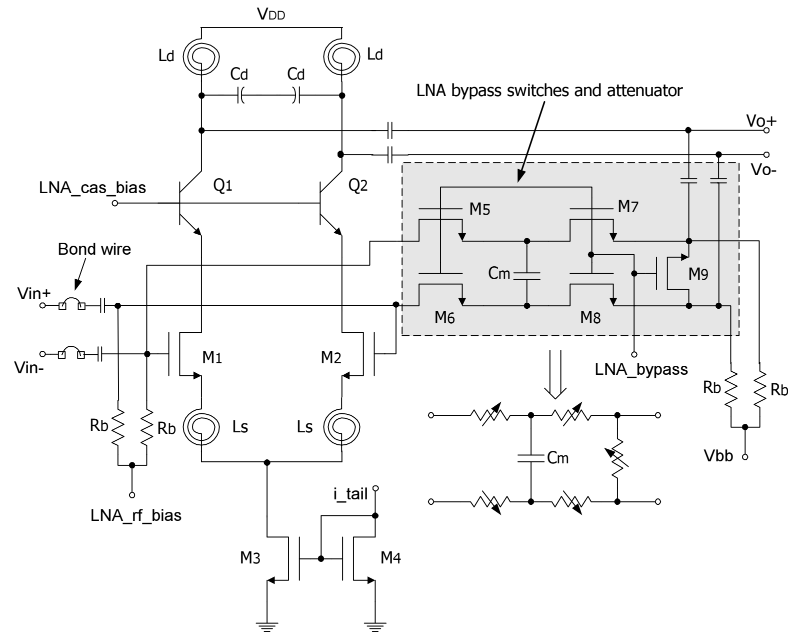



Fig. 45. LNA for BT/Wi-Fi receiver

All the matching conditions are established on-chip using the inductive source degeneration technique [10]. The only required off-chip components are a 1:1 RF BALUN to convert the single-ended signal to differential and two additional chip-inductors to provide matching.

The RF drive stage was chosen to be NMOS transistors due to the fact that MOS transistors are more linear than bipolar transistors for the same current consumption [28]. The cascoded bipolar transistors $Q_1$ and $Q_2$ provide stable and matched bias voltage to the drain of NMOS $M_1$ and $M_2$. Here bipolar is preferred because the base-emitter voltage change is much smaller than the gate-source voltage change for the same bias current variation. The transconductance available from NPN bipolar is much larger than the one from NMOS for the same current. Thus voltage at the drain of $M_1$ and $M_2$ does not change significantly which minimizes the voltage gain from

the gate to the drain of $M_1$ or $M_2$, therefore the Miller effect is reduced and reverse isolation is improved. This can be seen more clearly through the input impedance of the cascoded stage, which can be expressed as

$$Z_{cas,in} = r_e + \frac{\alpha_0 R_L}{g_{m,bjt} r_{ce}} \tag{4.4}$$

where $r_e = \frac{\alpha_0}{g_{m,bjt}} \approx \frac{1}{g_{m,bjt}}$. $\alpha_0$ is the emitter to collector current ratio and is very near unity for a modern NPN bipolar process. $R_L$ is the equivalent output load impedance. $r_{ce}$ is the bipolar transistor output impedance.

If the cascoded transistor is replaced by its MOS counterpart, its input impedance will be

$$Z'_{cas,in} = \frac{1}{g_{m,mos} + g_{mb,mos}} + \frac{R_L}{(g_{m,mos} + g_{mb,mos})\, r_{ds}} \tag{4.5}$$

where $g_{m,mos}$ and $g_{mb,mos}$ are the transconductance of a MOS transistor from gate and bulk terminal to drain terminal respectively. $r_{ds}$ is the drain output impedance. For the same bias current,

$$r_e < \frac{1}{g_{m,mos} + g_{mb,mos}}$$

and

$$g_{m,bjt} r_{ce} > (g_{m,mos} + g_{b,mos})\, r_{ds}$$

so

$$Z_{cas,in} < Z'_{cas,in}$$

. The gain from input of $M_1(M_2)$ to its drain is proportional to the cascoded transistors input impedance, therefore using bipolar as cascoded transistor this gain is smaller than that uses a MOS transistor.

Degenerative inductor $L_s$ (1 nH) and load inductor $L_d$ (3 nH) are on-chip planar spiral inductors. They are formed using the top layer metal (analog metal, AM) and there is a deep trench lattice pattern beneath them to reduce metal loss and substrate

loss respectively. Bond wires at the input are used to form the input biasing network. The use of bond wire has two benefits. First, its quality factor is much larger than on-chip inductors and thus, contributes less noise the overall noise factor. Second, it does not occupy die area and makes the chip more compact.

A systematic design procedure was followed from the beginning and the circuit was finely tuned through extensive simulations. Under matching conditions, there are several equations established for the input matching network. The operation frequency of the LNA relates the inductance and capacitance in the matching network as

$$\omega_o = \frac{1}{\sqrt{(L_g + L_s)\, C_t}} \tag{4.6}$$

In addition to the MOS transistor's intrinsic gate-source capacitance $C_{gs}$, another capacitor $C_{mim}$ is added between the gate and source for noise consideration (not shown in Fig. 45). $C_t$ is the total capacitance between the gate and source and

$$C_t = C_{gs} + C_{mim} \tag{4.7}$$

At $\omega_o$ the LNA is matched to the source impedance $R_s$ and

$$R_s = \frac{g_m}{C_t} L_s = 50\Omega \tag{4.8}$$

The input matching network's Q can be expressed as

$$Q = \frac{1}{2\omega_o C_t R_s} = \frac{1}{2\omega_o L_s g_m} \tag{4.9}$$

In the above equations, $g_m$ is the transconductance of $M_1$and $M_2$, $L_g$ is the required gate inductance implemented using bonding wire.

Now the proper Q value should be determined and the considerations about input Q is elaborated as follows. The RF filter before the LNA requires a certain load

termination, which is the input impedance of the LNA, for maximum power transfer and low sensitivity, and its performance is only guaranteed for a given tolerance for this termination impedance, for example, between $25\,\Omega$ and $100\,\Omega$. The variation in the reactance part of the matching network can also cause serious variation of the RF filter gain and pass band ripple. In order to maintain small variations in the impedance and reactance of the matching network, its Q value can not be too large, although a large Q is beneficial for low current consumption. Also large Q tends to degrade the linearity of the input stage, because the Q is also the voltage gain of the matching network, the input referred IIP3 will be reduced by a factor of Q. So the Q value is usually chosen to be 2-3 [29].

From (4.9), the required total gate-source capacitance for a specific Q can be calculated using

$$C_t = \frac{1}{2\omega_o R_s Q} \tag{4.10}$$

As stated above, the $C_t$ is composed of gate-source capacitance $C_{gs}$ of $M_1$ ($M_2$) and an MiM capacitance $C_{mim}$. The addition of $C_{mim}$ helps to optimize the noise performance and will be explained in further detail [11].

Fig. 46 shows the differential half circuit schematic (a) and small signal noise equivalent circuit (b) of the LNA. Here the cascoded bipolar transistors have minor influence on the noise behavior of the LNA, therefore its contribution to the total noise is discarded in the small signal circuit.

The capacitance $C_t$ affects the output noise current through the gate induced noise current $i_{n,g}$. Gate induced noise power can be expressed as

$$\overline{i_{n,g}^2} = 4kT\beta \frac{(\omega C_{gs})^2}{g_{do}} \Delta f \tag{4.11}$$

where $g_{do}$ is the transistor $M_1$'s output conductance for zero drain-source voltage, $\beta$

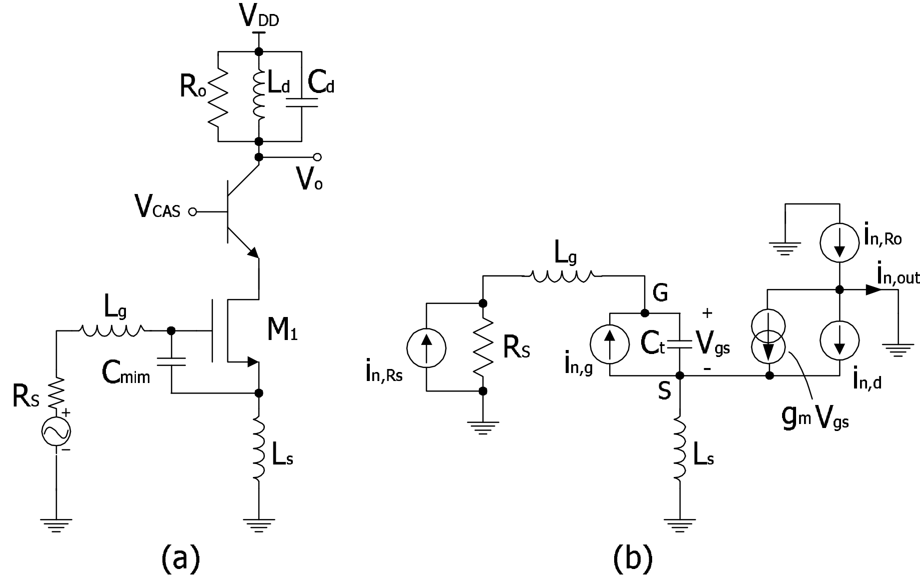



Fig. 46. Noise optimization

is the gate induced current noise factor and is about $\frac{4}{15}$ for long channel devices.

From the above discussions about the choice of the input quality factor, it is established that Q should be relatively small. However, in order to keep Q small, from (4.9), it is necessary to have large $C_t$. If all the capacitance is provided by the transistor's gate-source capacitance $C_{gs}$, the gate induced noise current will be quite large, since the gate induced noise grows with the square of $C_{gs}$, which can be easily seen from (4.11). Splitting $C_t$ into $C_{gs}$ and $C_{mim}$ decouples Q from $C_{gs}$ which allows for an adjustable reduction of Q for any given value of $C_{gs}$.

The effect of adding $C_{mim}$ can be further explained as follows. The output noise current due to gate induced noise current $i_{n,g}$ is

$$i_{n,o,g} = \frac{g_m}{j\omega_o C_t} \frac{jR_s\omega_o C_t - 1}{j2R_s\omega_o C_t} i_{n,g} \tag{4.12}$$

From (4.12) and (4.11) , one can obtain

$$\frac{\overline{i_{n,o,g}^2}}{\Delta f} \approx \frac{kT\beta}{g_{do}} \left(\frac{\omega}{\omega_o}\right)^2 \frac{g_m^2}{R_s^2 \left(\omega_o C_t\right)^2} \left(\frac{C_{gs}}{C_{gs} + C_{mim}}\right)^2 \propto P^2 \tag{4.13}$$

where P has been defined as

$$P \equiv \frac{C_{gs}}{C_t} = \frac{C_{gs}}{C_{gs} + C_{mim}} \tag{4.14}$$

The total capacitance $C_t$ is determined by Q through (4.9). Thus, by adding $C_{mim}$ and keeping $C_t$ fixed, the output noise current originated from the gate induced noise current is reduced by a factor of $\frac{1}{P}$.

The noise factor of Fig. 46(b) is [11]

$$F = 1 + aQ^2 W^{\frac{3}{2}} + \frac{a}{4} W^{\frac{3}{2}} + bQ^{-2} W^{-\frac{1}{2}} \tag{4.15}$$

where $a$ and $b$ are constant determined by the length of the transistor $M_1$, process parameters such as $\mu_{eff} C_{ox}$ and bias current.

For a fixed Q, there exists an optimal value for $W$, which can be obtained by taking the first order derivative of (4.15) in respect to $W$ and equating it to zero.

$$W_{opt} = \frac{A_b}{2Q^2} \sqrt{\frac{5}{6}} \frac{1}{\frac{4}{3} \omega_o R_s C_{ox} L} \tag{4.16}$$

where $A_b$ is the bulk charge factor [30].

Now that the size of the transistor is known from (4.16), $C_{gs}$ can be calculated from

$$C_{gs} = \frac{2}{3} C_{ox} W_{opt} L \tag{4.17}$$

$C_t$ has already been fixed by Q, so the required additional capacitance, $C_{mim}$, can be found through (4.7). The $g_m$ of $M_1$ can be fixed from the gain specification of the LNA, then the degeneration inductance $L_s$ will be obtained by (4.8). Finally the

biasing condition and current consumption can be determined.

After fine tuning through simulation, the width of $M_1$ and $M_2$ was fixed to be $96\,\mu m$ which was laid out by 24 fingers. The additional capacitance $C_{mim}$ is 277 fF and $C_{gs}$ is 140 fF for the size of the transistor mentioned above. The transconductance at the operating frequency of 2.4 GHz band is about 11 mA/V. The degeneration inductance $L_s$ is 1 nH and the required gate inductance $L_g$ is about 10 nH, which will be implemented by the bonding wire and off-chip surface mount inductors. It can be verified that the input impedance is about $26\,\Omega$ for the half-circuit shown in Fig. 46. So the differential input impedance is about $50\,\Omega$ as required.

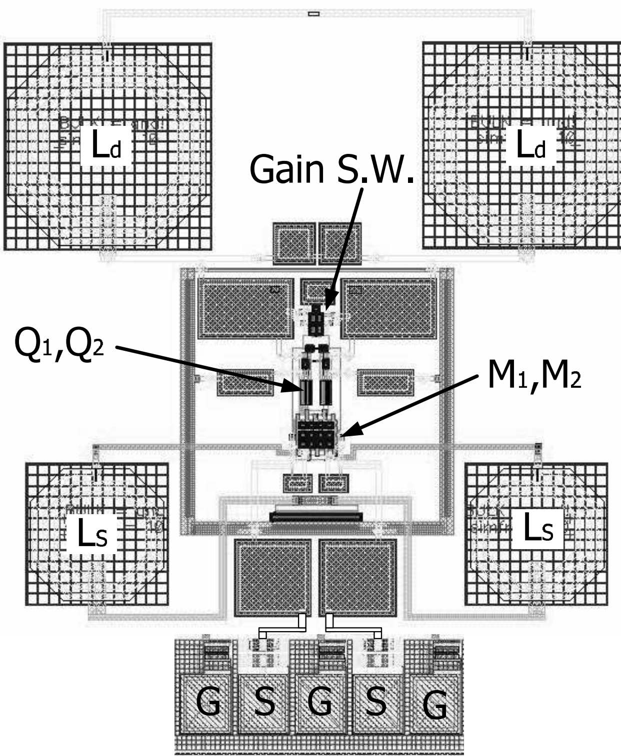



Fig. 47. LNA layout

Fig. 47 is the layout. Its area is $570\,\times\,580\,\mu m^2$. In the layout, interdigitate

and common-centroid techniques are used for transistor $M_1$ and $M_2$ to achieve good matching [31]. The differential input pads are formed into a G-S-G-S-G pattern to provide decoupling between each other.

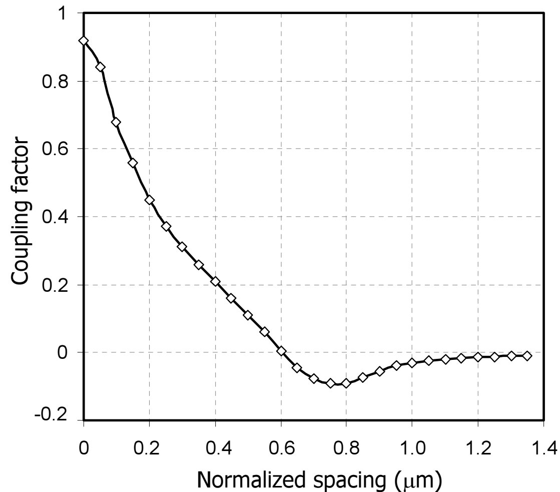



Fig. 48. Coupling factor between two shifted spirals

One can follow the LNA design flow given in Fig. 15, Chapter II. A single-ended LNA is designed first then transformed to the differential form by duplicating it and adding a tail current source.

It is worth mention that, for the differential structure, the two source degeneration inductors and two load inductors should be placed apart enough to avoid mutual coupling. In order to obtain a quantitative knowledge about the coupling effect, the mutual coupling factor $k$ between two inductors in adjacent metal layers is simulated using ASITIC [32]. The studied inductors are octagon spirals with a radius of $100\mu m$, metal spacing of $2\,\mu m$ and metal width of $8\,\mu m$. One of the inductors is laid out using metal layer 4, the other using metal layer 3. These two inductors are first put one on top of the other, then they are shifted apart. Mutual coupling factors ($k = \frac{M}{\sqrt{L_1 L_2}}$,

$M$ is the mutual inductance between $L_1$ and $L_2$) are calculated for different shifting and are plotted in Fig. 48. The x-axis is the normalized spacing and is defined by the spacing from center to center of the two spirals divided by the diameter of the spiral. When the two spirals are completely overlapped, $k$ is greater than 0.9 showing strong coupling. Coupling is reduced with the increment of normalized spacing. There exists a specific point where $k$ equals to zero. After this point, $k$ will increase with opposite polarity and will reach an extremum. When the normalized spacing is 1, the two spirals are shifted apart without any overlapping. The absolute value of the coupling factor $k$ for this configuration is about 0.03, which is already small enough. With the two spirals shifting further apart, the coupling factor becomes gradually smaller.

It can be seen from the above study that as long as the two spirals are put apart greater than 1.2 times of their diameter, the coupling can be reduced to a negligible value. In Fig. 47, the minimum normalized spacing among the four inductors is 2.6. Table X summarizes the simulation results of the LNA.

Table X. Chameleon LNA simulation results

| Parameter | Value | Unit |
|---|---|---|
| $S_{21}$ | 15/-15 | dB |
| NF | 1.6 | dB |
| IIP3 | -3 | dBm |
| $S_{11}$ | <-20 | dB |
| Pd | 16 | mW |

## 2. Mixer Implementation

The mixer shown in Fig. 49 is a fully differential Gilbert-cell based structure with I and Q branches sharing the same RF drive stage, therefore eliminating the RF drive stage mismatch compared to the conventional two separated I/Q mixers. The current commutating switches are NPN bipolar transistors which require less LO power than NMOS transistor switch pairs. This relaxes the required LO cross-coupling and isolation performance. Bipolar switching pair also has lower flicker noise than their NMOS counterpart. In addition, the bipolar transistors provide a higher $f_T$, which is required in order to achieve symmetrical on-off switching. This helps to minimize the second-order distortion caused by a non-ideal LO signal duty circle [33]. The RF driving stage uses NMOS transistors for high linearity. A further detailed discussion [34] of the LO leakage mechanism of the I/Q mixer is given as follows.

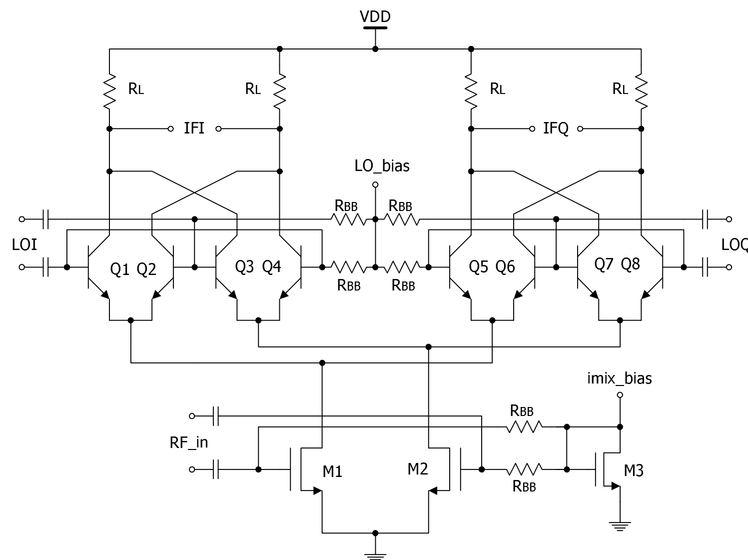



Fig. 49. Mixer for BT/Wi-Fi receiver

When there is no RF signal, due to the large LO signal, the voltage waveform

at the drain of $M_1$ or $M_2$ has a significant swing and it changes four times faster than that of the LO signal. Comparing with the case of using two separate mixers, the voltage swing at the drain of $M_1$ or $M_2$ is also much smaller. This is shown in Fig. 50. Therefore, a signal with four times the LO frequency will leak into the mixer's RF input port through capacitive coupling ($C_{gd}$, $C_{gs}$). Because this leakage is at high frequency and its amplitude is about 10 times smaller than that of the two separate mixer configuration, it will not cause problem. Due to device mismatch, this high frequency signal will also appear at IF port but will be filtered out by the load capacitance. Fig. 51 is the mixer layout view. Table XI summarizes the mixer's simulation results.

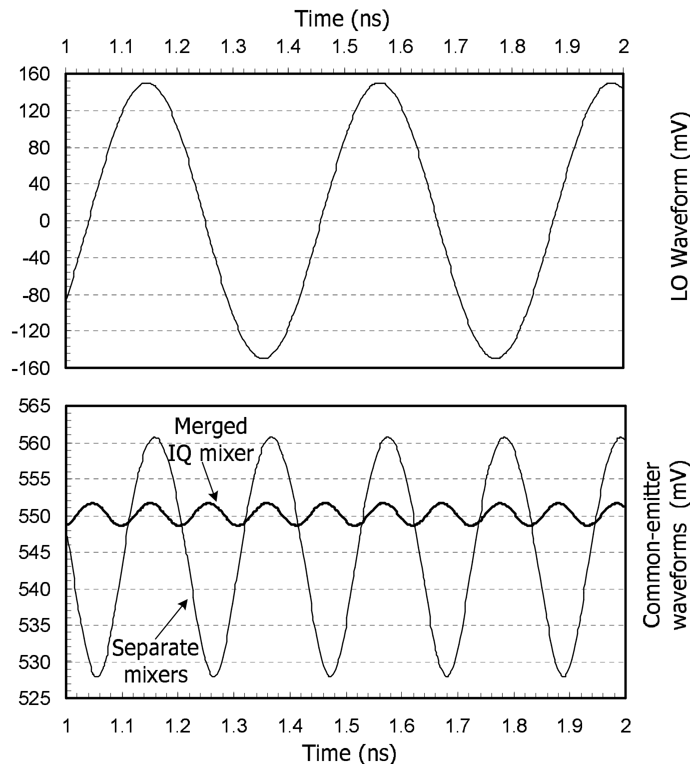



Fig. 50. Waveforms at switching pair common emitters

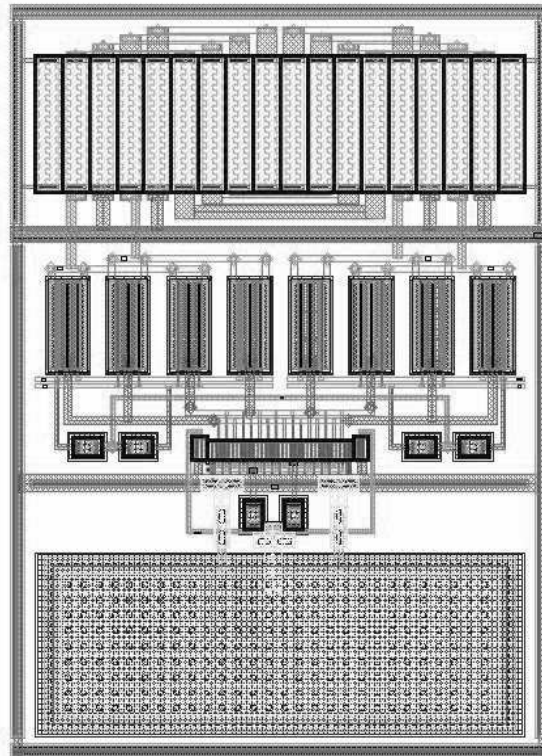


Fig. 51. Mixer layout

Table XI. Chameleon Mixer simulation results

| Parameter | Value | Unit |
|---|---|---|
| Conversion Gain | 19 | dB |
| DSB. NF | 10.6 | dB |
| IIP3 | +2 | dBm |
| LO drive | -10 | dBm |
| Pd | 8.8 | mW |

The design flow of the mixer can be generally followed as shown in Fig. 37, Chapter III. The differences are 1) the current switching pairs are formed by bipolar transistors here, their sizes are chosen just large enough to accommodate the current following through them, which helps to reduce the parasitics at the drain of the RF drive MOS transistors; 2) There is no bleeding current source in this implementation.

### 3. PTAT Biasing Circuit

The RF driving stage of the LNA and mixer are all NMOS transistors. The threshold voltage of a MOS device tends to have a temperature coefficient in the magnitude around $-2\,mV/^oC$, like the bipolar transistor base-emitter voltage $V_{BE}$. The carrier mobility is also temperature dependent and it tends to dominate the temperature behavior of a MOS transistor due to its exponential nature as shown in (4.18).

$$\mu\left(T\right) \approx \mu_0 \left(\frac{T}{T_0}\right)^{-\frac{3}{2}} \tag{4.18}$$

where $\mu_0$ is the mobility at reference temperature $T_0$. The transconductance $g_m$ of a MOS transistor is proportional to $\mu\left(T\right)$. With the increment of temperature, carrier mobility decreases, so does the $g_m$. This can be seen more easily by considering the MOS $g_m$ expression in the first order I-V approximation, i.e. the square law $I_D = K_n\left(V_{GS} - V_{th}\right)^2$ as

$$\begin{aligned} g_m &= 2K_n\left(V_{GS} - V_{th}\right) \\ &= 2\sqrt{K_n I_D} \end{aligned} \tag{4.19}$$

where $K_n = \frac{1}{2}\mu\left(T\right)C_{ox}\frac{W}{L_{eff}}$. So It can be derived from (4.19) and (4.18) that

$$g_m \propto \sqrt{\frac{I_D}{T^{3/2}}} \tag{4.20}$$

In order to combat the temperature-induced $g_m$ reduction, the bias current $I_D$should be increased with temperature. It is not easy to implement the $\frac{3}{2}$ expo-

nential increment of $I_D$, while a linear increment is implemented using PTAT (Proportional To Absolute Temperature) current source [35]. Fig. 52 is the simplified schematic of the biasing scheme for the proposed RF front-end.

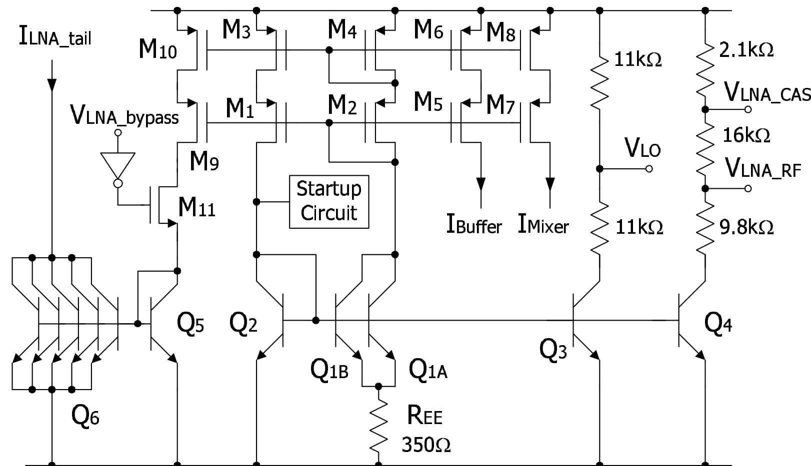



Fig. 52. Bias circuit for the RF front-end

$Q_1$ ($Q_{1A}$ and $Q_{1B}$ in parallel), $Q_2$, $M_1 \sim M_4$ and $R_{EE}$ form the PTAT current source. $Q_{1A}$, $Q_{1B}$ and $Q_2$ have the same emitter area $A_E$. The current mirror formed by $M_1 \sim M_4$ forces the current flowing through transistor $Q_1$ to be the same as the current flowing through $Q_2$. This current is noted as $I_{PTAT}$ and assumes the saturation current density of $Q_1$ and $Q_2$ is $J_o$, then the base-emitter voltage of transistor $Q_1$ and $Q_2$ can be written as

$$V_{BE1} = \frac{kT}{q} \ln \frac{I_{PTAT}}{2A_E J_o} \tag{4.21}$$

$$V_{BE2} = \frac{kT}{q} \ln \frac{I_{PTAT}}{A_E J_o} \tag{4.22}$$

The voltage developed across resistor $R_{EE}$ can be obtained by subtracting (4.21)

from (4.22)

$$\Delta V_{BE} = V_{BE2} - V_{BE1} = \frac{kT}{q} \ln 2 \tag{4.23}$$

This voltage is proportional to absolute temperature. If the temperature coefficient of resistor $R_{EE}$ can be neglected, then the PTAT current $I_{PTAT}$ is given by

$$I_{PTAT} = \frac{V_t}{R_{EE}} \ln 2 \tag{4.24}$$

where $V_t = \frac{kT}{q}$. For the $R_{EE}$ value shown in Fig. 52, the nominal current at $298\,K$ ambient temperature is about $52\,\mu A$. All the other currents and voltages are derived from this source. $V_{LNA_bypass}$ is the control terminal to turn off LNA and put it into attenuation mode.

From the above discussions, the biasing circuit can be designed as following. First the current mirroring ratios are chosen. These ratios are usually less than 10. Too small value will increase the power consumption of the biasing circuit, while too large value will degrade the current mirroring accuracy. For the LNA, this ratio is 4 and for mixer, this ratio is 8. Secondly, the PTAT current $I_{PTAT}$ is known from the current of the LNA and mixer, and the mirroring ratios. The required emitter resistor $R_{EE}$, can be calculated from (4.24). The third step is to chose proper value of resistors to generate all the required bias voltage. Finally, a start-up circuit is added to ensure the correct operation status of the PTAT current source and a control circuit is also inserted to turn on and off the LNA bias current. Table XII shows the nominal biasing voltages and currents for the RF front-end.

Table XII. RF front-end nominal biasing condition

| Bias point | $I_{buffer}$ | $I_{mixer}$ | $I_{LNA}$ | $V_{LO}$ | $V_{LNA_RF}$ | $V_{LNA_CAS}$ |
|---|---|---|---|---|---|---|
| Value | 12.8$\mu A$ | 410$\mu A$ | 2mA | 1.6V | 1.4V | 2.2V |

## D. Layout and Experimental Results

The RF front-end was fabricated using IBM 0.25 $\mu m$ SiGe BiCMOS technology through MOSIS. A die photo is shown in Fig. 53. The area of RF front-end is 740 $\times$ 770 $\mu m^2$, not including bond pads. Deep trench and substrate contact rings are placed around the front-end as indicated in Fig. 54. There are two layers of deep trench. The substrate contact ring is connect to a quiet ground. This arrangement improves the substrate noise isolation from the other part of the die.

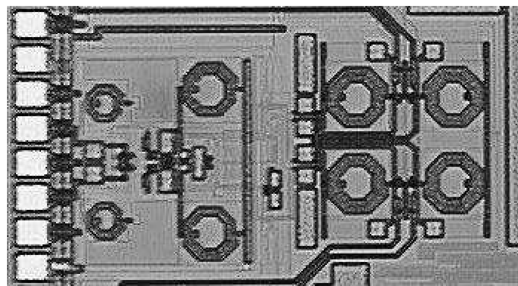


Fig. 53. Die photo of the front-end

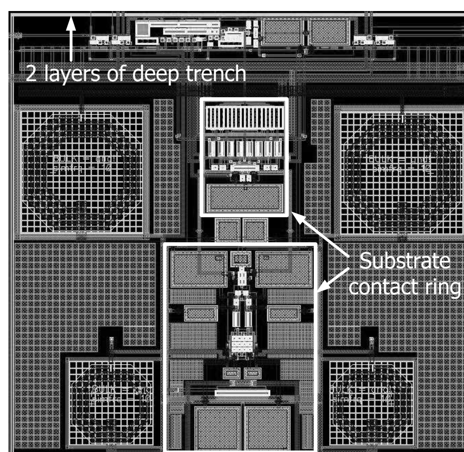



Fig. 54. Substrate noise isolation by deep trenches and guard rings

The testing board photo is shown in Fig. 55. This is the board for the whole dual-mode Bluetooth/Wi-Fi receiver with accessible mixer output. The LNA input SMA connector is put as close to the chip as possible. This is why there is a cut-in at the left edge of the board. The PCB is fabricated using FR4 material with a thickness of 0.031". This FR4 material is relatively less costly than dedicated high frequency lamination materials, and the thin thickness of the board makes the 50 Ωand 25 Ω transmission line width manageable.

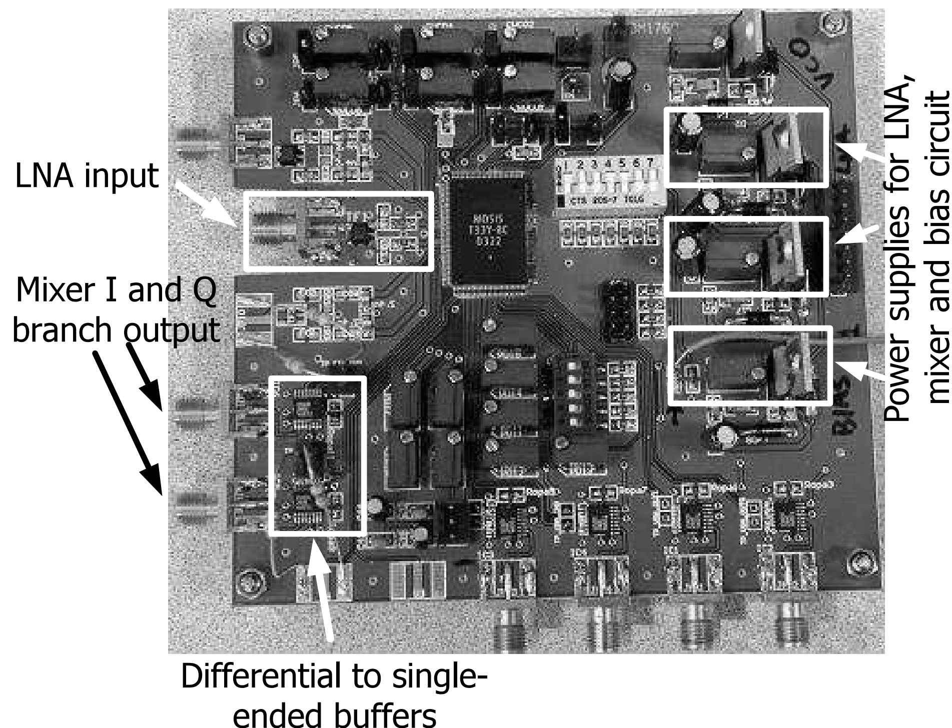
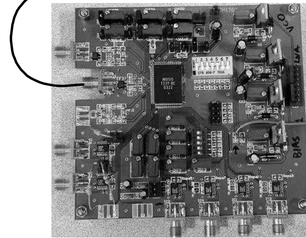



Fig. 55. RF front-end test board

The input impedance matching condition was checked using a network S-parameter analyzer HP8719ES. The testing setup is presented in Fig. 56, and the testing results are shown in Fig. 57 for high gain mode and Fig. 58 for low gain mode. In both cases the matching is better than -11 dB.

Network S-parameter analyzer (HP8719ES)



Test board

Fig. 56. Testing setup for input match

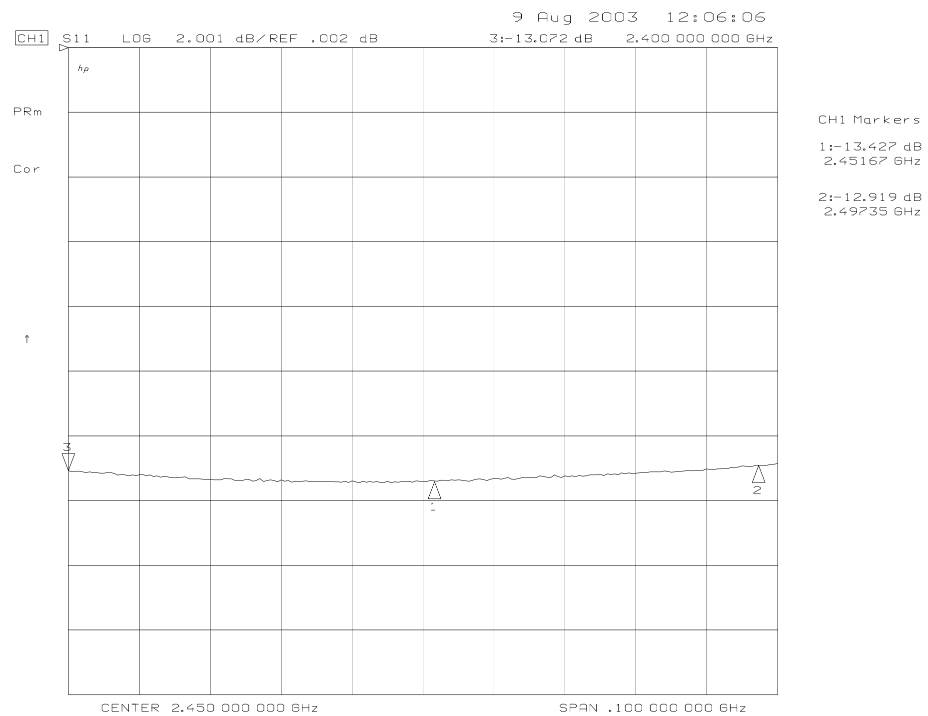



Fig. 57. Input matching for high gain mode

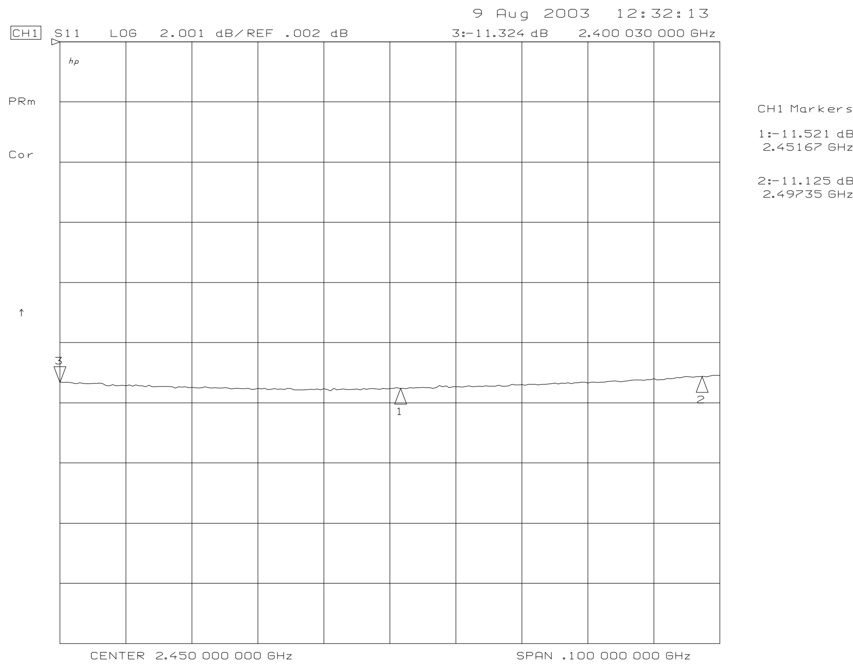



Fig. 58. Input matching for low gain mode

The intermodulation performance of the front-end was tested within the whole receiver using the two-tone test method. Fig. 59 shows the instrumental setup. Two signal generators SMIQ03 are used to generate the two testing tones. The two tones are then combined by a power combiner and fed into the LNA's input. The output spectrum is observed by the spectrum analyzer FSEB30 from the VGA output with the VGA gain setting of 12 dB.

The IIP3 curve in Wi-Fi mode is plotted in Fig. 60. The two tones are applied at 12 MHz and 25 MHz away from same side of the LO frequency respectively when the LNA is in high gain mode. The measured IIP3 is -13 dBm.

Fig. 61 shows the IIP2 plot. The two tones are applied at 12.2 MHz and 12.8 MHz away from same side of LO tone and the measured IIP2 is 10 dBm. These values meet the system specifications and indicate that the front-end performed as designed.

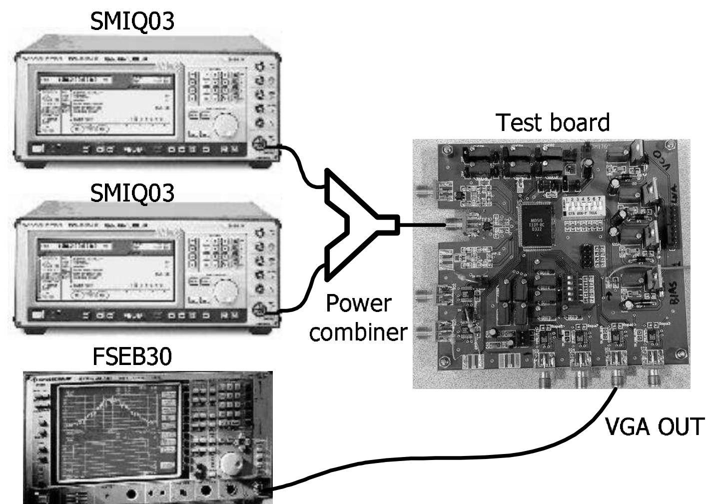



Fig. 59. Testing setup for IIP3 and IIP2 measurement

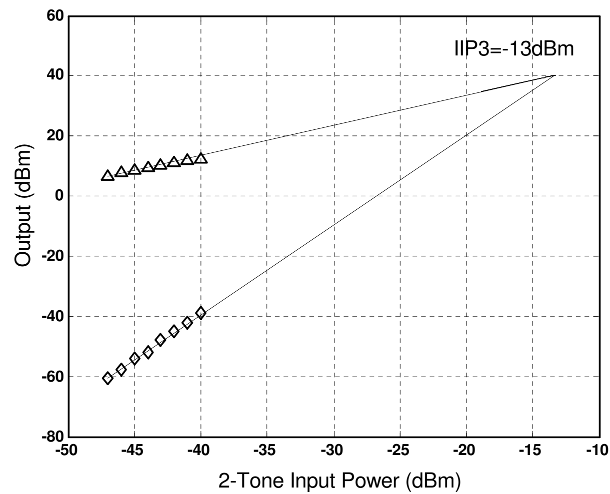



Fig. 60. IIP3 plot for 2-tone test at 12MHz and 25MHz offset

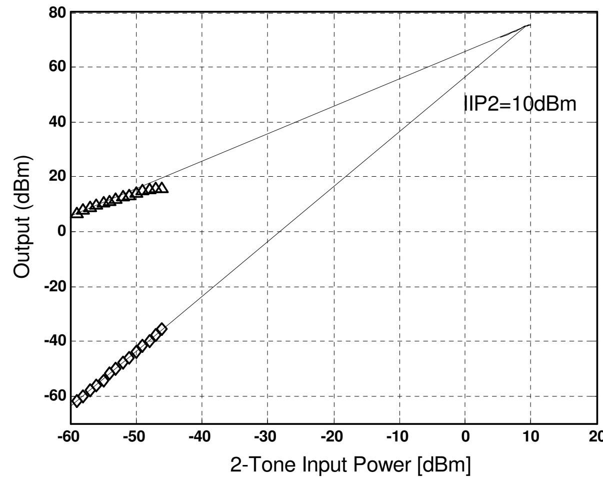



Fig. 61. IIP2 plot for 2-tone test at 12.2MHz and 12.8MHz offset

The I and Q branch matching performance testing setup is shown in Fig. 62. The input signal is swept to cover from 1 MHz to 10 MHz IF frequency range. The I and Q branch's amplitude and phase difference is observed by putting vector network analyzer HP89140 into vector mode and looking at the amplitude and phase response.

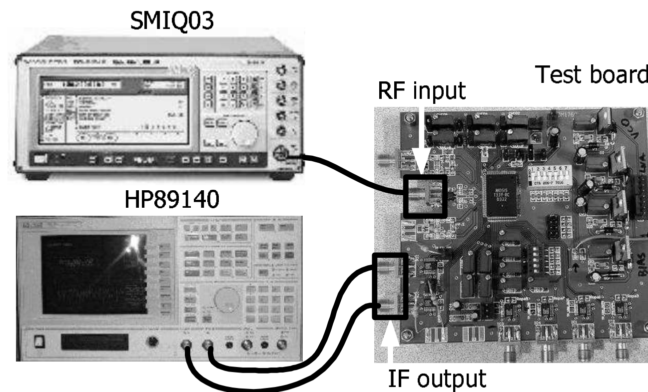



Fig. 62. I/Q mismatch measurement

The measured mismatch between I and Q outputs of the mixer across 10 MHz

IF frequency range is shown in Fig. 63. It shows that the amplitude mismatch is less than 1.2 dB, and phase mismatch is within 3.8 degrees. In the Bluetooth mode (up to 1 MHz) the phase mismatch is as large as 3.5 degrees and the amplitude mismatch 0.96 dB. For the Wi-Fi mode (up to 6MHz), the phase mismatch is smaller than 3.5 degrees and the amplitude mismatch smaller than 1 dB. The demodulator's SNR degradation due to the I/Q mismatch is less than 0.2 dB in both cases.

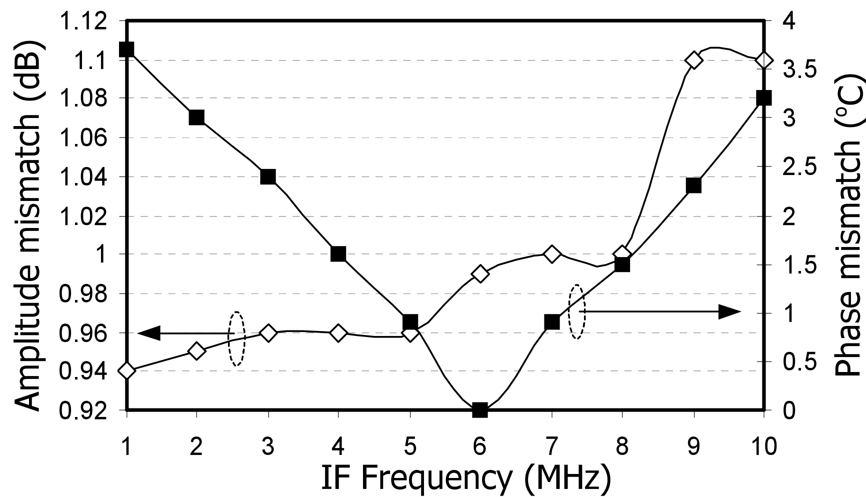



Fig. 63. I-Q mismatch performance

The noise figure and conversion gain measurement can be performed by using the spectrum analyzer FSEB30 from Rohde & Schwarz and noise source NC346B form Noise/COM. The testing setup is shown in Fig. 64. Before the actual measurement, the instruments are setup for calibration, then the front-end test board is inserted into the testing chain. Testing is automated by software supplied with the spectrum analyzer. The measured conversion gain is about 33 dB and noise figure is 5.5 dB across 10 MHz IF frequency range. Table XIII summarizes the measurement results.

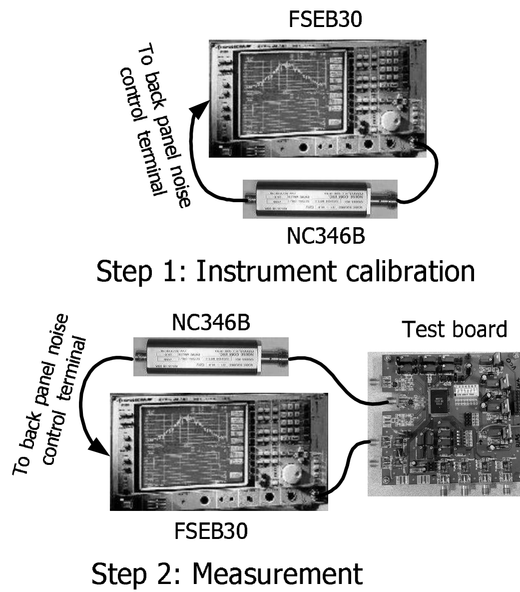



Fig. 64. Testing setup for noise figure and conversion gain

Table XIII. RF front-end measurement results

| Parameters | Value | Unit |
|---|---|---|
| Voltage gain | 33 | dB |
| IIP3 | -13 | dBm |
| Vdd | 2.5 | V |
| Current | 13.6 | mA |
| IIP2 | 10 | dBm |
| NF | 5.5 | dB |
| S11 | < -11 | dB |
| Phase mismatch | < 3.8 | degree |
| Amp. mismatch | <1.2 | dB |

# CHAPTER V

# LNA LINEARIZATION TECHNIQUES

Linearity is a key issue in RF systems. A circuit's non-linearity causes many problems such as inter-modulation and gain compression. The development of micro-processor technique makes the CMOS process much cheaper than the others. The system-on-chip target, also demands CMOS technology. For the same current consumption, NMOS transistors are more linear than bipolar transistors. Still, the linearity of MOS transistors can not meet the stringent requirements of state-of-the-art applications such as CDMA/AMPS. For example [36], in the IS-98 CDMA standard, the IIP3 of the LNA is set by the single-tone desensitization requirement:

$$\text{IIP3} \geq 59.5 - L_{TX-RX} + L_{TX} \text{ [dBm]} \tag{5.1}$$

where $L_{TX-RX}$ is the duplexer TX-RX isolation in TX band and $L_{TX}$ is the duplexer TX-antenna insertion loss. Typically, $L_{TX-RX}$ is about 53 dB and $L_{TX}$ is about 2.7 dB, which requires the LNA's IIP3 to be better than +9.2 dBm.

Linearity based on negative feedback is not suitable for high frequency applications due to stability issues and gain reduction. Thus, a lot of linearization techniques that focus on linearizing MOSFET transistors are introduced. The basic idea of linearization here is to use an additional transistor's non-linearity to compensate or cancel the nonlinearity of the main operation device using a feed-forward technique. The conventional way involves MOS transistors working in triode or weak inversion to provide linearization. With the development of technology, bipolar is available in CMOS technology. Although its performance is not as good as that in BiCMOS process, it is sufficient to provide linearization. The proposed linearization technique

is derived from multi-gated linearization. A time-invariant memoryless system can be represented using a Taylor series. Volterra analysis should be applied to non-linear systems with memory effect. Based on the non-linearity analysis, the techniques for linearizing the LNA will be introduced, and then the proposed linearization technique using hybrid transistors, i.e. using both MOS and BJT is discussed in details.

## A. Non-Linearity Analysis

### 1. Non-Linear System Representations

For a memoryless non-linear circuit or system, its transfer function can be represented by Taylor series:

$$\begin{aligned} y &= \sum_{k=0}^{+\infty} a_k x^k \\ &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots \end{aligned} \tag{5.2}$$

The system non-linearity is usually characterized through the two-tone test. Suppose the two testing signals are two sinusoids with the same amplitudes and different frequencies: $x_1 = A\cos(\omega_1 t)$ and $x_2 = A\cos(\omega_2 t)$. Substituting $x = x_1 + x_2$ into (5.2), it is easy to shown that the linear term $(x_1 + x_2)$ of the two-tone test is

$$A\cos\omega_1 t + A\cos\omega_2 t \tag{5.3}$$

the 2nd-order term $(x_1 + x_2)^2$ can be shown as

$$\begin{aligned} A^2 &+ \tfrac{1}{2}A^2\cos 2\omega_1 t + \tfrac{1}{2}A^2\cos 2\omega_2 t \\ &+ A^2\cos(\omega_1 - \omega_2)t + A^2\cos(\omega_1 + \omega_2)t \end{aligned} \tag{5.4}$$

and the 3rd-order term $(x_1 + x_2)^3$ can be expended into

$$\begin{pmatrix} & \frac{9}{4}\cos\omega_1 t + \frac{9}{4}\cos\omega_2 t \\ + & \frac{1}{4}\cos 3\omega_1 t + \frac{1}{4}\cos 3\omega_2 t \\ + & \frac{3}{4}\cos\left(2\omega_1 - \omega_2\right)t + \frac{3}{4}\cos\left(2\omega_2 - \omega_1\right)t \\ + & \frac{3}{4}\cos\left(2\omega_1 + \omega_2\right)t + \frac{3}{4}\cos\left(2\omega_2 + \omega_1\right)t \end{pmatrix} \times A^3 \tag{5.5}$$

The 3rd-order intermodulation (IM3) is one of the most important considerations in the RF small signal amplifier design. The 3rd order intermodulation component is contributed by the odd-order nonlinearities. The major contribution comes from the 3rd-order term $(x_1 + x_2)^3$ in the Taylor series. But the higher odd-order terms also have contribution, especially when 3rd order cancellation techniques are used, the 5th-order contribution may be pronounced. The $(x_1 + x_2)^5$ term contributes to the fundamental, 3rd-order and 5th-order intermodulation components as follow:

$$\begin{pmatrix} & \frac{25}{4}\cos\omega_1 t + \frac{25}{4}\cos\omega_2 t \\ + & \frac{25}{8}\cos\left(2\omega_1 - \omega_2\right)t + \frac{25}{8}\cos\left(2\omega_2 - \omega_1\right)t \\ + & \frac{5}{8}\cos\left(3\omega_1 - 2\omega_2\right)t + \frac{5}{8}\cos\left(3\omega_2 - 2\omega_1\right)t \end{pmatrix} \times A^5 \tag{5.6}$$

Higher than 5th-order non-linearities usually have less contributions and will generally be ignored.

Fig. 65 summarizes the frequency components of the two-tone test as mentioned above. For the 3rd-order and 5th-order terms, only the fundamental and intermodulation components are shown in the figure. Notice that odd-order nonlinearities generate spurs around the fundamental and compress or expand the fundamental amplitude. Even-order nonlinearities can generate DC and low frequency components which may change the DC biasing point of the circuits.

If the system has memory effect, its Taylor expansion should be replaced by

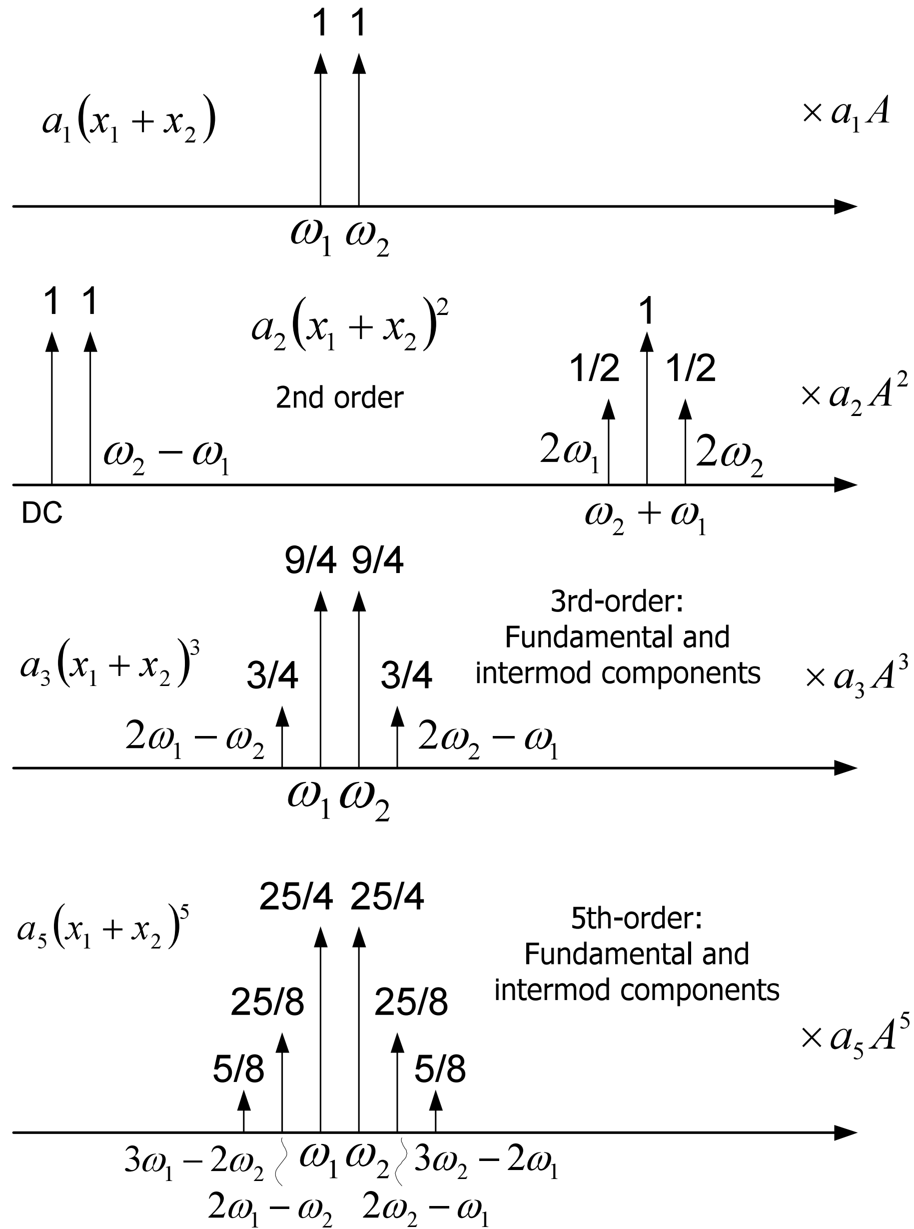



Fig. 65. Frequency components in two-tone test

Volterra series (see Appendix B):

$$\begin{aligned} y &= \sum_{k=0}^{+\infty} H_k\left(\omega_1, \omega_2, \cdots, \omega_k\right) \circ x^k \\ &= H_0 + H_1\left(\omega_1\right) \circ x + H_2\left(\omega_1, \omega_2\right) \circ x^2 + H_3\left(\omega_1, \omega_2, \omega_3\right) \circ x^3 + \ldots \end{aligned} \tag{5.7}$$

where $H_k\left(\omega_1, \omega_2, \cdots, \omega_k\right)$ is called Volterra kernel. The DC term $H_0$ can be obtained from Taylor expansion and will be dropped for AC analysis. The operator "$\circ$" means that the magnitude and phase of each frequency term in $x^k$ is to be changed by the magnitude and phase of $H_k\left(\omega_1, \omega_2, \cdots, \omega_k\right)$ [37]. The Taylor series only shows the low frequency effect (or the amplitude) of circuit non-linearity for a circuit with memory, while Volterra series contains both amplitude and phase information. For a general non-linear system, the locations of frequency components are the same as depicted in Fig. 65, but their amplitudes and phases will be functions of frequency $\omega_1$ and $\omega_2$. The rest of the text in this section will obverse several non-linearity effects using Volterra or Taylor notations.

### 2. Non-Linearity of Fully Differential Circuits

For a fully-differential circuit, assume its two inputs are $x_1 = x$ and $x_2 = -x$, then keeping up to 5th-order non-linearity terms, its two outputs will be

$$y_{o1}\left(x_1\right) = H_1 \circ x + H_2 \circ x^2 + H_3 \circ x^3 + H_4 \circ x^4 + H_5 \circ x^5 \tag{5.8}$$

and

$$y_{o2}\left(x_2\right) = -H_1 \circ x + H_2 \circ x^2 - H_3 \circ x^3 + H_4 \circ x^4 - H_5 \circ x^5 \tag{5.9}$$

The differential output will be

$$y_o = y_{o1} - y_{o2} = 2\left(H_1 \circ x + H_3 \circ x^3 + H_5 \circ x^5\right) \tag{5.10}$$

So a perfect fully-differential circuit does not have even-order terms in its differential output. But in reality, there will be asymmetry in the circuit. This can be the mismatch between two signal paths or the input signals itself are not fully-balanced. This effect can always be modeled by a DC offset $\Delta$ at the input signal as $x_1 = x + \Delta$ and $x_2 = -x$. Under this condition, the differential output can be shown to be

$$\begin{aligned} y_o &\approx 2\left(H_1 \circ x + H_3 \circ x^3 + H_5 \circ x^5\right) \\ &+ H_1\Delta + \left(2H_2 \circ x + 3H_3 \circ x^2 + 4H_4 \circ x^3 + 5H_5 \circ x^4\right) \end{aligned} \tag{5.11}$$

It can be seen that input signal offset or system mismatch will cause the differential output not only having DC offset $H_1\Delta$ but also the even-order terms. The output DC offset is also a function of frequency because $H_1$ is generally frequency dependent.

### 3. IM3 Due to 5th-Order and 2nd-Order Non-Linearity

IM3 is directly generated by 3rd-order non-linearity of the system but higher odd-order nonlinearities also have contribution to IM3. For example, in Fig. 65 the fifth-order nonlinearity term can also generate an IM3 term and cause the IM3 curve to deviate from 3:1 slope.

At small signal amplitude, when 5th-order effects can be ignored, the amplitudes of the IM3 terms are proportional to the 3rd power of the input amplitude. If the signal amplitude is significantly large, however, 5th-order distortion will start to affect the IM3 responses. When some of the 3rd-order non-linearity cancellation techniques are used to reduce the 3rd-order distortion, the IM3 curve usually deviates from the 3:1 slope for a moderately large input signal. The curve will compress or expand determined by the phase of the fifth-order term. If the phases of the 3rd and 5th order coefficients are coherent, 5th-order distortion will expand the IM3 curve, whereas if

the phases are opposite, the IM3 curve will be compressed.

Second-order non-linearity contributes to IM3 indirectly through feedback in the non-linear system. Specifically, the second-order components $2\omega_1$, $2\omega_2$ and $\omega_1 \pm \omega_2$ feedback to the input of the system (through $C_{gd}$ of MOS transistor, e.g.) and mixed with the fundamentals again to generate 3rd-order terms. The IM3 term contributed by the 3rd-order distortion and second-order distortion can be written using Volterra series coefficients sun as

$$\begin{aligned} IM3\left(2\omega_1-\omega_2\right) &= \tfrac{3}{4}H_3\left(\omega_1,\,\omega_1,\,-\omega_2\right)X\left(\omega_1\right)X\left(\omega_1\right)X^*\left(\omega_2\right) \\ &+ H_2\left(\omega_1,\,\omega_1-\omega_2\right)X\left(\omega_1\right)X^*\left(\omega_2-\omega_1\right) \\ &+ H_2\left(-\omega_2,\,2\omega_1\right)X^*\left(\omega_2\right)X\left(2\omega_1\right) \end{aligned} \tag{5.12}$$

$$\begin{aligned} IM3\left(2\omega_2-\omega_1\right) &= \tfrac{3}{4}H_3\left(\omega_2,\,\omega_2,\,-\omega_1\right)X\left(\omega_2\right)X\left(\omega_2\right)X^*\left(\omega_1\right) \\ &+ H_2\left(\omega_2,\,\omega_2-\omega_1\right)X\left(\omega_2\right)X\left(\omega_2-\omega_1\right) \\ &+ H_2\left(-\omega_1,\,2\omega_2\right)X^*\left(\omega_1\right)X\left(2\omega_2\right) \end{aligned} \tag{5.13}$$

where $H_n$ is the n-th order Volterra kernel, $n = 2, 3$. $X(\omega_k)$ represents the input two tones in frequency domain, $k = 1, 2$. $X\left(2\omega_1\right)$, $X\left(2\omega_2\right)$ and $X\left(\omega_2-\omega_1\right)$ are the second-order intermodulation terms feeded back to the input of the system. $X^*\left(\omega\right)$ is the complex conjugate of $X\left(\omega\right)$. Fig. 66 shows how the envelope terms and harmonics of the second-order distortion contribute to IM3 terms.

Quite often one may observe that the IM3 response at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are asymmetric in amplitudes. Suppose equal amplitudes in the two-tone test input signals at $\omega_1$and $\omega_2$, non-linear terms caused by 3rd-order distortion $H_3$ in the first term of (5.12) and (5.13) usually match each other. Asymmetry arises from the facts that i) the envelop term $X\left(\omega_2-\omega_1\right)$ appears in opposite phase in the lower and upper sidebands, which is clear from the complex conjugate operation in the second term of (5.12) and (5.13), and ii) the response of fundamental tones may be different. This

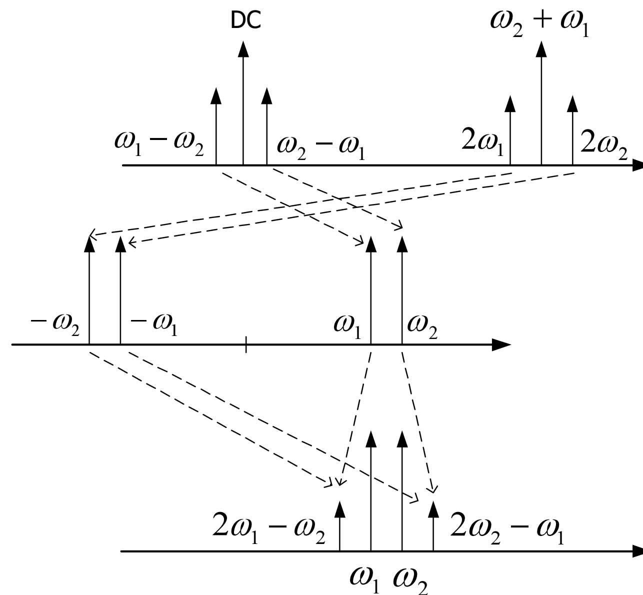



Fig. 66. Second-order distortion contributing to IM3 terms

usually is not the case because a flat pass band is generally desired for fundamentals. But the 2nd-order harmonics $X(2\omega_1)$ and $X(2\omega_2)$ in the third term of (5.12) and (5.13) are already far away from each other and further away from the pass band, so their response can be quite different.

## 4. Non-Linearity Due to Output Impedance

For a MOS transistor, non-linearity is largely introduced by the non-linear behavior of its transconductance, but its output impedance is also non-linear and contributes to distortion. At low frequency, the circuit can be assumed to be memoryless and Taylor analysis can be applied.

To start, write the MOS transistor's small signal output impedance as

$$r_{out} = \frac{1}{\lambda I_D} \tag{5.14}$$

where $\lambda = \frac{1}{V_A}$, $V_A$ is Early voltage. $I_D$ is drain current. If the drain current contains DC term $I_{DQ}$ and AC term $i_o$, then

$$r_{out} = \frac{1}{\lambda \left(I_{DQ} + i_o\right)} \tag{5.15}$$

Assume $r_{out}$ dominates the output AC impedance, thus the output voltage due to this impedance is

$$v_o = i_o r_{out} = \frac{i_o}{\lambda \left(I_{DQ} + i_o\right)} = \frac{1}{\lambda} \frac{i_o/I_{DQ}}{1 + i_o/I_{DQ}} \tag{5.16}$$

It is known for $|x| < 1$, $\frac{x}{1+x} = x - x^2 + x^3 - x^4 + x^5 + \ldots$. Usually$\left|\frac{i_o}{I_{DQ}}\right| < 1$, thus

$$v_o \approx r_o \left( i_o - \frac{1}{I_{DQ}} i_o^2 + \frac{1}{I_{DQ}^2} i_o^3 \right) = \beta_1 i_o + \beta_2 i_o^2 + \beta_3 i_o^3 \tag{5.17}$$

where $r_o = \frac{1}{\lambda I_{DQ}}$, $\beta_1 = r_o$, $\beta_2 = -\lambda r_o^2$, $\beta_3 = \lambda^2 r_o^3$. The output AC current $i_o$ can be related to the input voltage $v_i$ by

$$i_o = -G_m\left(v_i\right) \approx -\left[\alpha_1 v_i + \alpha_2 v_i^2 + \alpha_3 v_i^3\right] \tag{5.18}$$

where $\alpha_1 = g_m = K V_{od} \frac{2+\theta V_{od}}{1+\theta V_{od}}$, $\alpha_2 = \frac{K}{(1+\theta V_{od})^3}$, $\alpha_3 = \frac{-\theta K}{(1+\theta V_{od})^4}$.

From (5.17) and (5.18), one can arrive at

$$v_o = -\alpha_1 \beta_1 v_i - \left(\alpha_2 \beta_1 + \alpha_1^2 \beta_2\right) v_i^2 - \left(\alpha_3 \beta_1 + 2\alpha_1 \alpha_2 \beta_2 + \alpha_1^3 \beta_3\right) v_i^3$$

or

$$v_o = -A_v v_i + \left[\lambda A_v^2 - \alpha_2 r_o\right] v_i^2 + \left[2\lambda \alpha_2 A_v r_o - \lambda^2 A_v^3 - \alpha_3 r_o\right] v_i^3 \tag{5.19}$$

where $A_v = \alpha_1 \beta_1 = g_m r_o$. If the non-linearity induced by output impedance can be ignored, i.e. $\lambda \sim 0$ then

$$v_o = -A_V v_i - \alpha_2 r_o v_i^2 - \alpha_3 r_o v_i^3 \tag{5.20}$$

Compare (5.19) and (5.20) the total nonlinearity term may be improved or degraded depending on the sign and relative value of the terms contributed from output impedance non-linearity. If the output of the transistor is loaded by an additional impedance $Z_L$ which is linear and $|Z_L| \ll r_{out}$, then the total output impedance can be approximated by $Z_L$and the input $G_m$ non-linearity dominates the overall non-linearity behavior.

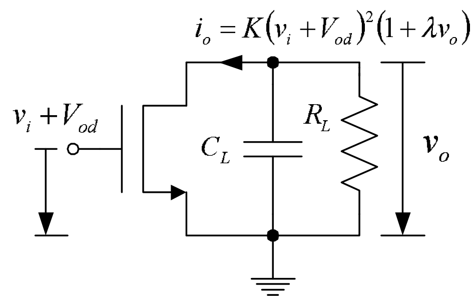



Fig. 67. Load non-linearity large signal model

At high frequency, the load impedance will be dominated by the load capacitance, thus memory effect can not be ignored. The output voltage should be expressed using Volterra series:

$$v_o = H_0 + H_1 \circ v_i + H_2 \circ v_i^2 + H_3 \circ v_i^3 + \cdots \tag{5.21}$$

Fig. 67 illustrates the large signal model of the circuit under consideration. For simplicity and clarity, it is assumed that the MOS transistor is a long channel device and has a large signal transfer function of

$$i_o = K\left(v_i + V_{od}\right)^2 \left(1 + \lambda v_o\right) \tag{5.22}$$

where $V_{od} = V_{gs0} - V_{th}$, $V_{th}$ is transistor's threshold voltage. The output voltage for

sinusoid inputs is

$$v_o = -i_o Z_L = -K\left(v_i + V_{od}\right)^2\left(1 + \lambda v_o\right) Z_L \tag{5.23}$$

where $Z_L$ represents the impedance of $R_L$ and $C_L$ in parallel. Keep (5.21) up to 3rd-order term and substitute it into (5.23):

$$\begin{aligned} H_0 + H_1 \circ v_i + & \\ H_2 \circ v_i^2 + H_3 \circ v_i^3 \; = \; & D_0 + D_1 v_i + D_2 v_i^2 \\ + \; & \lambda D_0 \left(H_o + H_1 \circ v_i + H_2 \circ v_i^2 + H_3 \circ v_i^3\right) \\ + \; & \lambda D_1 \left(H_o + H_1 \circ v_i + H_2 \circ v_i^2 + H_3 \circ v_i^3\right) v_i \\ + \; & \lambda D_2 \left(H_o + H_1 \circ v_i + H_2 \circ v_i^2 + H_3 \circ v_i^3\right) v_i^2 \end{aligned} \tag{5.24}$$

where

$$D_0 = -K V_{od}^2 Z_L \tag{5.25}$$

$$D_1 = -2K V_{od} Z_L \tag{5.26}$$

$$D_2 = -K Z_L \tag{5.27}$$

In order to find the Volterra kernel $H_k$, equate the same order term of $v_i$ in both sides of (5.24) and use the following relationships:

$$K V_{od}^2 R_L \ll \frac{1}{\lambda} \tag{5.28}$$

$$\omega C_L \gg \frac{1}{R_L} \tag{5.29}$$

$$g_m = 2K V_{od} \tag{5.30}$$

$$g_o = K \lambda V_{od}^2 \tag{5.31}$$

(5.28) holds because MOS transistor's Early voltage is usually much greater than its output DC voltage. (5.29) means the circuit's output impedance is dominated

by the load capacitance at high frequency and the circuit has strong memory effect. (5.30) and (5.31) are the linear small signal transconductance and output conductance respectively. It can be found that

$$H_0 = \frac{D_0}{1+\lambda D_0} \approx -KV_{od}^2 R_L \left(1 - g_o R_L\right) \tag{5.32}$$

$$H_1\left(\omega\right) = \frac{1+\lambda H_0}{1+\lambda D_0} D_1 \approx -\frac{g_m}{j\omega C_L}\left(1 - g_o R_L\right) \tag{5.33}$$

$$H_2\left(\omega_1,\, \omega_2\right) = \frac{D_2 + \lambda\left(D_1 H_1 + D_2 H_0\right)}{1+\lambda D_0} \approx -\frac{K\left(1 - g_o R_L\right)}{j\left(\omega_1+\omega_2\right) C_L} \tag{5.34}$$

$$H_3\left(\omega_1,\, \omega_2,\, \omega_3\right) = \frac{\lambda\left(D_1 H_2 + D_2 H_1\right)}{1+\lambda D_0} \approx -\frac{K\lambda g_m Z_3\left(\omega_1,\, \omega_2,\, \omega_3\right)}{\left(\omega_1+\omega_2+\omega_3\right) C_L^2} \tag{5.35}$$

where $Z_3\left(\omega_1,\, \omega_2,\, \omega_3\right) = \frac{1}{3}\left(\frac{1}{\omega_1+\omega_2} + \frac{1}{\omega_1+\omega_3} + \frac{1}{\omega_2+\omega_3} + \frac{1}{\omega_1} + \frac{1}{\omega_2} + \frac{1}{\omega_3}\right)$.

It can be shown by substituting $g_m = 2KV_{od}$, $\alpha_2 = K$ and $\alpha_3 = 0$ into (5.19) that the low frequency Taylor series of the circuit in Fig. 67 is

$$v_o = -g_m r_o v_i + 3K r_o v_i^2 - 2K\lambda g_m r_o^2 v_i^3 \tag{5.36}$$

Here the DC term is dropped and it is assumed $R_L \gg r_o$. Notice that due to the output capacitive loading, both the amplitude and phase of the non-linear term are quite different from the case without memory effect.

### 5. Third-Order Distortion of Inductive Source-Degenerated LNA

Inductive source-degenerated LNA in CMOS technology is a very popular LNA structure. It handles the trade-off among input impedance matching, noise figure and gain gracefully. This text will study the 3rd-order non-linearity of this type of LNA. Figs. 68 are the schematic view (a) and equivalent circuit for non-linearity analysis (b).

It is assumed that the dominant non-linearity comes from the transconductance

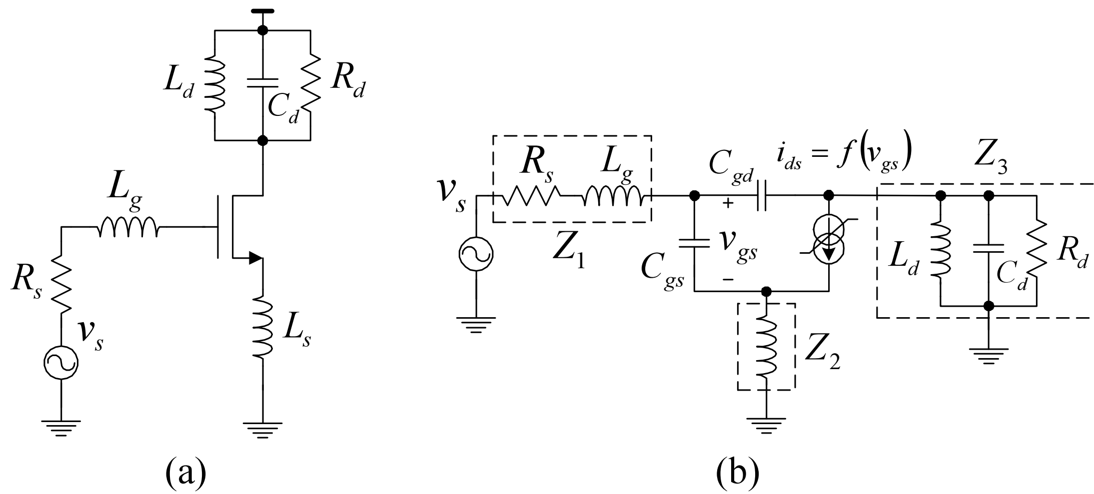



Fig. 68. Inductive degenerated CMOS LNA

$g_m$ of the LNA core device, here the MOS transistor. The drain AC current can be expanded using Taylor series up to 3rd-order term as

$$i_{ds} = f\left(v_{gs}\right) = g_m v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 \tag{5.37}$$

The distortion analysis for BJT was studied in [38], [39] using Volterra series. The equations can be adapted for MOS transistors by noting that MOS transistors have infinite DC current gain ($\beta \to \infty$) and no base-emitter diffusion capacitance ($\tau = 0$) [40]. In a two-tone test, the two tones are near each other, $\omega \approx \omega_2 \approx \omega_1$, such that their separation, $\Delta\omega = \omega_2 - \omega_1$, is much smaller than $\omega$. $R_s$ is the signal source resistance. The $IMD_3$ at drain node and the input referred IP3 power can be expressed as

$$IMD_3 = \frac{3}{4} \cdot \left|H\left(\omega\right)\right| \cdot \left|A_1\left(\omega\right)\right|^3 \cdot \left|\varepsilon\left(\Delta\omega,\, 2\omega\right)\right| A_s^2 \tag{5.38}$$

and

$$IIP3\left(2\omega_2 - \omega_1\right) = \frac{1}{6R_s \cdot \left|H\left(\omega\right)\right| \cdot \left|A_1\left(\omega\right)\right|^3 \cdot \left|\varepsilon\left(\Delta\omega,\, 2\omega\right)\right|} \tag{5.39}$$

where $A_s$ is the amplitude of the testing tones and

$$H(\omega) = \frac{1 + j\omega C_{gs}\left[Z_1(\omega) + Z_2(\omega)\right] + j\omega C_{gd} Z_1(\omega)}{g_m - j\omega C_{gd}\left[1 + Z_2(\omega)\left(g_m + j\omega C_{gs}\right)\right]} \tag{5.40}$$

$$A_1(\omega) = \frac{1}{g_m + g(\omega)} \frac{1 + j\omega C_{gd} Z_3(\omega)}{Z_x(\omega)} \tag{5.41}$$

$$\varepsilon(\Delta\omega,\, 2\omega) = g_3 - g_{oB}(\Delta\omega,\, 2\omega) \tag{5.42}$$

$$g_{oB}(\Delta\omega,\, 2\omega) = \frac{2}{3} g_2^2 \left[\frac{2}{g_m + g(\Delta\omega)} + \frac{1}{g_m + g(2\omega)}\right] \tag{5.43}$$

$$g(\omega) = \frac{1 + j\omega C_{gd}\left[Z_1(\omega) + Z_3(\omega)\right] + j\omega C_{gs}\left[Z_1(\omega) + Z_x(\omega)\right]}{Z_x(\omega)} \tag{5.44}$$

$$Z_x(\omega) = Z_2(\omega) + j\omega C_{gd}\left[Z_1(\omega) Z_2(\omega) + Z_1(\omega) Z_3(\omega) + Z_2(\omega) Z_3(\omega)\right] \tag{5.45}$$

In the above equations, $H(\omega)$ relates the equivalent input IM3 voltage to the IM3 response of the drain current non-linear terms. $A_1(\omega)$ is the linear transfer function from the input voltage $v_s$ to the gate-source voltage $v_{gs}$. $\varepsilon(\Delta\omega,\, 2\omega)$ shows how the non-linear terms in (5.37) contribute to the 3rd-order distortion. The first term in (5.42) comes from the 3rd-order non-linearity and the second term comes from the 2nd-order non-linearity as explained in (5.12) and (5.13). The 2nd-order feedback paths here include the gate-drain capacitor and the degeneration inductor.

$|H(\omega)|$ and $|A_1(\omega)|$ depend on the in-band source and load impedances which are usually selected to provide desired gain, noise figure and impedance match. Therefore lower distortion is achieved by reducing $|\varepsilon(\Delta\omega,\, 2\omega)|$. If $g_3$ dominates non-linearity which is generally the case, reducing it can significantly improve IIP3. In a bipolar LNA, out-of band termination or matching is usually used to make the term $|\varepsilon(\Delta\omega,\, 2\omega)|$ small thus IMD3 will be reduced. This is possible because $g_3$ and $g_{oB}$ have the same sign for bipolar transistor. $|\varepsilon(\Delta\omega,\, 2\omega)|$ is the difference between these two quantities. By making $g_3 \approx g_{oB}$, $|\varepsilon(\Delta\omega,\, 2\omega)|$ can be maintained at a pretty small value. But for a MOS transistor, it will be shown later that $g_3$ and $g_{oB}$ have

different signs. More specifically, in the saturation region, since $g_3$ is negative while $g_{oB}$ is positive, the only way to reduce $|\varepsilon(\Delta\omega,\, 2\omega)|$ is to make both $g_3$ and $g_{oB}$ small values.

So when $g_3$ has already been reduced by the third-order cancellation technique, it is also important to reduce $g_{oB}$ by keeping $g(\Delta\omega)$ and $g(2\omega)$ large. It is assumed in the above discussions that $\Delta\omega$ is at a very low frequency. Therefore for the inductive source-degenerated LNA, $Z_1(\Delta\omega) \approx R_s$, $Z_2(\Delta\omega) \approx 0$ and $Z_3\Delta\omega \approx 0$. So $g(\Delta\omega)$ has a very large value with respect to $g_m$, and $g_{oB}(\Delta\omega,\, 2\omega)$ can be approximated by

$$g_{oB}(\Delta\omega,\, 2\omega) \approx g_{oB}(2\omega) = \frac{2}{3}\frac{g_2^2}{g_m + g(2\omega)} \tag{5.46}$$

If the load LC tank has a high enough quality factor, then $Z_3(2\omega)$ is a small quantity and $Z_1(2\omega) = R_s + j2\omega L_g$, $Z_2(2\omega) = j2\omega L_s$. Under input impedance match condition, $\omega_T L_s = \frac{g_m}{C_{gs}} L_s = R_s$ holds. Putting all these considerations together, it can be shown that $|g(2\omega)|$ is proportional to $\frac{1}{\omega L_s}$. So in order to have small $g_{oB}$, one must select a small $L_s$, which means less degeneration. This is contrary to the general belief that degeneration improves linearity. Actually, inductive degeneration will degrade linearity. An additional out-of-band termination network can be added at the input to make $Z_1(2\omega) \approx 0$ [39]. If a termination network is used, $|g(2\omega)|$can be about $2 \sim 3$ times larger. Therefore, IIP3 can have an improvement by at least 3 dB. This $|g(2\omega)|$ increment assumes that the out-of-band termination network is added directly at the gate of the MOS transistor, thus it has to be implemented on the chip. In practice, this termination network can be implemented as a high Q LC parallel tank resonant at $\omega$, provides a very small impedance path to ground at frequency $2\omega$. Its performance will be limited by the quality of inductors available in the process and it will introduce additional noise, and probably affect the in-band impedance match. So it is usually not desirable for the LNA design. Off-chip termination using

low loss quarter-wave transmission lines is another way, but the improvement will be relatively small and is not worth the complexity, and added cost. To conclude, in order to make a linear LNA, the 3rd-order coefficient $g_3$ of the intrinsic transistor's Taylor series should be small and if possible, adding an out-of-band termination network and/or using a smaller degeneration inductor will also help. The rest of this chapter will study the techniques to provide a more linear LNA core circuit.

## B. Theoretical Analysis of Multi-Gated Linearization Technique

The low noise amplifier (LNA) has stringent requirements on operation frequency, noise, and linearity. Therefore, a LNA usually uses a minimum number of transistors. For example, a single-ended LNA usually contains only one or two transistors in its main signal path. In order to improve the LNA's linearity, more components have to be added into the signal path. Resistive degeneration and shunt-series feedback are traditionally used, but they introduce additional noise and require more power. A method based on multi-gated transistor (MGT) third-order non-linearity cancellation is discussed in [41]-[44]. This method directly reduces the 3rd-order coefficient of a LNA's core devices. There is no theoretical analysis done on the previously reported MGT. This section will explore the nature of the multi-gated linearization technique and the later sections will present a different way to implement the 3rd-order intermodulation cancellation by using a hybrid structure, i.e. combining the MOS and bipolar transistors available in a CMOS technology.

The input-output transfer characteristics of a MOS transistor can be expressed using the Taylor series (5.2) and is repeated here by dropping DC and terms higher than 3rd-order:

$$i_{ds}\left(v_{gs}\right) = g_m v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 \tag{5.47}$$

In order to obtain a large IIP3, $g_m$ should be kept almost unchanged or even larger and try to reduce $g_3$ and $g_2$.

The transfer characteristics of a short channel MOS transistor that is valid in all operation regions can be expressed as [45]

$$i_{DS} = K\frac{\chi^2}{1+\theta\chi} \tag{5.48}$$

where

$$\chi = 2\eta\phi_t \ln\left(1 + e^{\frac{V_{gs}-V_{th}}{2\eta\phi_t}}\right) \tag{5.49}$$

$V_{th}$ is the threshold voltage, $\phi_t$ is the thermal voltage $\frac{kT}{q}$, $\mu_e$ is effective mobility and $K = 0.5\mu_e C_{ox}\frac{W}{L}$. $\theta$ approximately models source series resistance, mobility degradation due to vertical E-field and velocity saturation effect. It is a function of channel length and is independent of body effect. $\eta$ is the rate of exponential increase of drain current with gate-source voltage in the sub-threshold region and the size of the moderate inversion region, which has values between 1 and 2 [45]. In moderate or strong inversion, (5.48) reduces to

$$i_{DS} = K\frac{\left(V_{gs0} - V_{th} + v_{gs}\right)^2}{1+\theta\left(V_{gs0} - V_{th} + v_{gs}\right)} \tag{5.50}$$

Here the gate-source voltage is expressed as the sum of the a DC bias voltage $V_{gs0}$ and small signal AC voltage $v_{gs}$, and $V_{od} = V_{gs0} - V_{th}$ is the gate source over-drive voltage.

Expanding (5.50) using the Taylor series in terms of $v_{gs}$ and neglecting the DC component and components higher than 3rd-order, the coefficients in (5.47) are determined by

$$g_m = \frac{KV_{od}\left(2+\theta V_{od}\right)}{\left(1+\theta V_{od}\right)^2},\ g_2 = \frac{K}{\left(1+\theta V_{od}\right)^3},\ g_3 = -\frac{\theta K}{\left(1+\theta V_{od}\right)^4} \tag{5.51}$$

The low frequency 3rd-order intercept point for the transistor in strong inversion is

$$A_{IIP3,strong}^{2} = \frac{4}{3}\frac{V_{od}}{\theta}\left(2 + \theta V_{od}\right)\left(1 + \theta V_{od}\right)^{2} \tag{5.52}$$

and a lower bound can be found by assuming $\theta V_{od}$ is sufficiently small to be

$$A_{IIP3,strong}^{2} > \frac{8}{3}\frac{V_{od}}{\theta} \tag{5.53}$$

In weak inversion, (5.48) can be reduced to

$$i_{DS} = K\left(2\eta\phi_t\right)^2 e^{\frac{V_{gs0} - V_{th} + v_{gs}}{\eta\phi_t}} \tag{5.54}$$

Further expressed in small signal term

$$i_{ds} = I_{s0} e^{\frac{v_{gs}}{\eta\phi_t}} \tag{5.55}$$

where $I_{s0} = K\left(2\eta\phi_t\right)^2 e^{\frac{V_{od}}{\eta\phi_t}}$. Treated the same way as in the strong inversion case, the Taylor series coefficients of (5.54) are

$$g_m = \frac{I_{s0}}{\eta\phi_t},\ g_2 = \frac{I_{s0}}{2\left(\eta\phi_t\right)^2},\ g_3 = \frac{I_{s0}}{6\left(\eta\phi_t\right)^3} \tag{5.56}$$

and the 3rd-order intercept point is

$$A_{IIP3,weak} = 2\sqrt{2}\eta\phi_t \tag{5.57}$$

Fig. 69 shows the second and third-order terms of a NMOS transistor versus its gate-source biasing levels. It can be easily seen that between the moderate/strong and weak inversion, the 2nd-order term has the same sign, and both are positive. The 3rd-order term has a different sign. In the moderate or strong inversion, it has negative sign, and in the weak inversion it has a positive sign. In the moderate inversion region, there is a point where the 3rd order term is zero. Of course, it is very hard to bias a transistor exactly at this optimal point for minimum 3rd-order

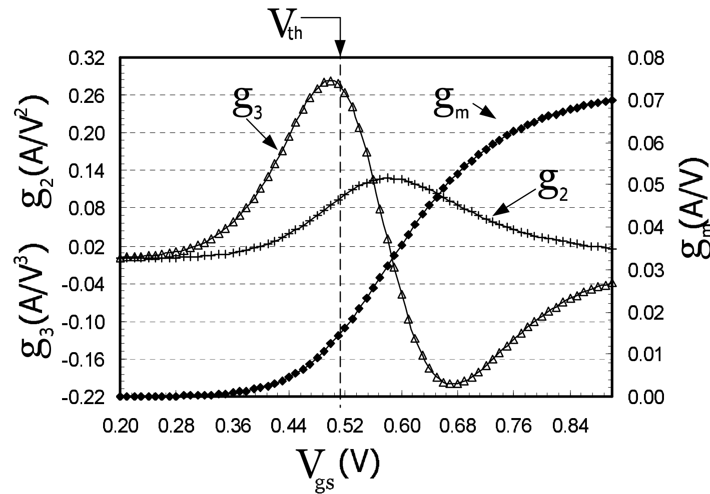



Fig. 69. MOS transistor 2nd-order and 3rd-order distortion terms

non-linearity. The alternative is to use two or more transistors and bias them at different inversion regions and trim the size of the transistors such that the positive term will cancel the negative term. This way, the minimum 3rd-order distortion is achieved. [40]-[44] shows the idea of using multiple NMOS transistors connected in parallel form and biased at different gate drive levels to achieve an extremely linear device for RF circuit applications such as LNAs and power amplifiers.

The multi-gated configuration using two NMOS transistors is shown in Fig. 70. Transistor $M_1$ is the main transistor working in the strong inversion region. Transistor $M_2$ is the auxiliary device biased in the weak inversion region.

Fig. 71 shows the linearity (IIP3 and IIP2) versus bias voltage $V_{Baux}$ of the auxiliary transistor. The main transistor is biased at 0.74 V. Notice that when the bias voltage of the auxiliary transistor is around $0.5 \sim 0.55\,V$, the combined device response has the best IIP3. The IIP3 improvement is about 10 dB compared to only having transistor $M_1$. Note that the linearity measurement here is for multi-gated

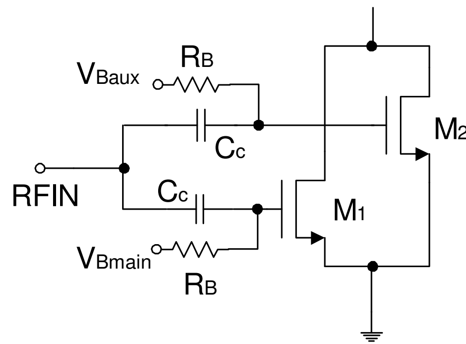



Fig. 70. Multi-gated linearization using two NMOS transistors

transistors only. The observed output is the combined drain current. The actual circuit using the multi-gated core will have less IIP3. The current consumption only increases slightly because $M_2$ is operating in the weak inversion.

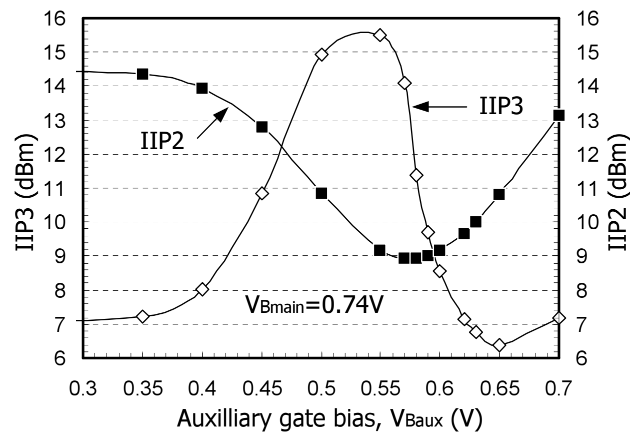



Fig. 71. NMOS multi-gated transistor linearity v.s. auxiliary bias voltage

Notice that the IIP2 almost has its worst value as demonstrated in Fig. 71. This is due to the fact that for different gate biases, although the 3rd-order terms can have different signs, so they can be canceled out by combining the current, the 2nd-order

terms will always have the same signs. By adding the output currents of the main and auxiliary transistors, the 2nd-order term will be added constructively, deteriorating the IIP2.

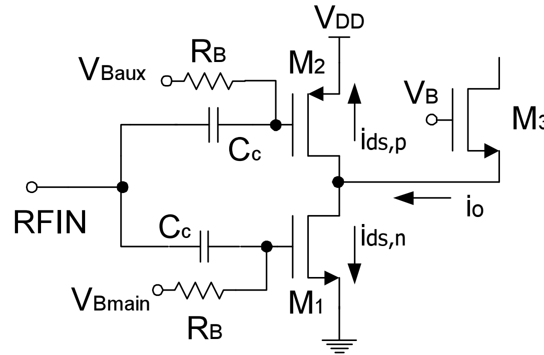



Fig. 72. Multi-gated linearization using NMOS and PMOS

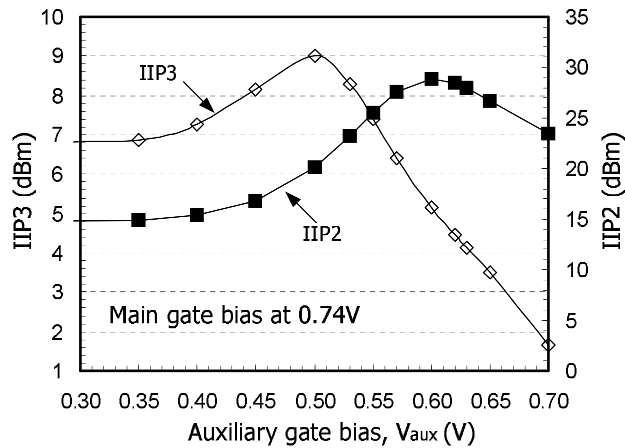



Fig. 73. Complementary multi-gated transistor linearity plot

By using PMOS instead of NMOS for the auxiliary transistor (complementary multi-gated transistor, CMGT), IIP3 and IIP2 can be improved at the same time. Fig. 72 illustrates this configuration. When the AC signal of the NMOS transistor

$(M_1)$ is increasing, the corresponding AC input signal of the PMOS transistor $(M_2)$ will decrease. So if the input for $M_1$ is $v_{gs}$, the input for $M_2$ will be $-v_{gs}$. Thus if we ignore the DC and higher order distortion terms, the output current of NMOS $M_1$ and PMOS $M_2$ can be written as

$$i_{ds,n} = g_{m,n}v_{gs} + g_{2,n}v_{gs}^2 + g_{3,n}v_{gs}^3 \tag{5.58}$$

$$i_{ds,p} = -g_{m,p}v_{gs} + g_{2,p}v_{gs}^2 - g_{3,p}v_{gs}^3 \tag{5.59}$$

The overall output current of the CMGT is the difference between the currents of those two transistors:

$$\begin{aligned} i_o &= i_{ds,n} - i_{ds,p} \\ &= (g_{m,n} + g_{m,p})\, v_{gs} + (g_{2,n} - g_{2,p})\, v_{gs}^2 + (g_{3,n} + g_{3,p})\, v_{gs}^3 \end{aligned} \tag{5.60}$$

It is observed that as the total transconductance increases, the IM2 term decreases because $g_{2,n}$ and $g_{2,p}$ have the same sign, and the IM3 term decreases because $g_{3,n}$ and $g_{3,p}$ have different signs as shown in (5.51) and (5.56). Fig. 73 shows the IIP3 and IIP2 curves of the CMGT configuration. Because PMOS transistors have an inferior performance compared to NMOS transistors, the IIP3 improvement is not as significant as the NMOS MGT. It is also clear that the IIP2 and IIP3 do not share the same optimal bias voltage. This is due to the different non-linear characteristics of NMOS and PMOS transistors. If the future process can match the non-linear behavior and speed of NMOS and PMOS transistors, the linearity improvement will be more significant.

Instead of making $g_3$ small by using MGT and $g_{oB}$ small by out-of-band termination, [47] proposed a method to change the phase and amplitude of the MGT's in-band $g_3$ by tapping the degeneration inductor into the source of the transistor working in the weak inversion. The consequence is that $g_3$ is modified to match the

phase and amplitude of $g_{oB}$, therefore, an improved IIP3 is achieved.

## C. Proposed Linearization Scheme Using BJTs in CMOS Process

### 1. Hybrid LNA: BJT as Auxiliary Transistor

Using a MOS transistor biased at weak inversion may have potential speed limitations [48]. A new implementation method to cancel the IM3 term is proposed. The goal is to keep the main transistor $M_1$ in the strong or moderate inversion where its IM3 has a negative sign and another transistor is added to provide a positive signed IM3. Instead of using the MOS transistor $M_2$ biased in its weak inversion, a bipolar transistor $Q_2$ is used. Fig. 74 depicts the configuration of the proposed LNA linearization method (Hybrid LNA). This implementation requires both MOS and BJT available in the process. BiCMOS is a natural choice, but in a specific RF CMOS process, BJT is sometimes available as a byproduct. Although its performance can not compete with that in a BiCMOS process, for linearization purposes, it may be good enough. For example, in the TSMC 0.18 $\mu m$ RF CMOS process, a $2\,\mu m \times 2\,\mu m$ bipolar transistor biased at $11\,\mu A$ base current, its cut-off frequency $f_T$ is 28 GHz and its $\beta_{ac}$ is 22. A NMOS RF transistor with dimensions $185\,\mu m \times 0.18\,\mu m$ biased at $0.64\,V$ gate-source voltage has a cut-off frequency of about 30 GHz.

The current-voltage transfer function of a bipolar transistor can be expressed by

$$i_{ce} = I_{so} e^{\frac{V_{BEQ+v_{be}}}{\phi_t}} = I_Q e^{\frac{v_{be}}{\phi_t}} \tag{5.61}$$

where $V_{BEQ}$is the base-emitter bias voltage, $I_{so}$ is the saturation current and $I_Q = I_{so} e^{\frac{V_{BEQ}}{\phi_t}}$. Note that (5.61) resembles (5.55), so its 3rd-order term coefficient can be obtained from (5.56) as

$$g_{3,\,bjt} = \frac{g_m^3}{6 I_Q^2} \tag{5.62}$$

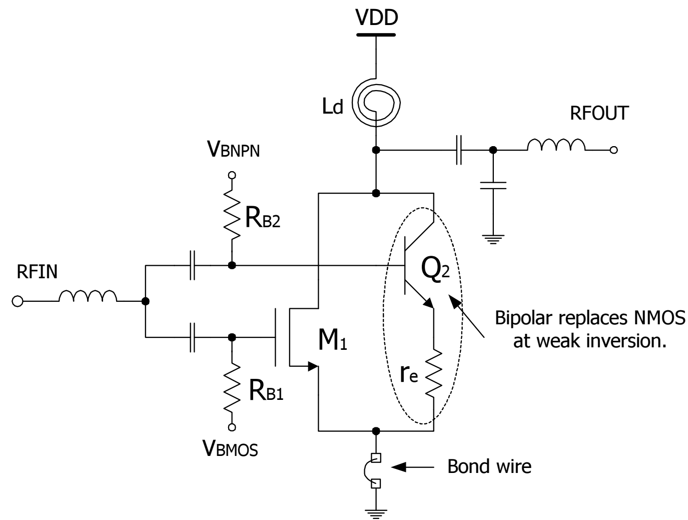



Fig. 74. Bipolar as auxiliary transistor for 3rd-order linearization

From (5.51), NMOS transistor's 3rd-order term coefficient in moderate/strong inversion is

$$g_{3,\,mos} = -\frac{\theta K}{(1+\theta V_{eff})^4} \tag{5.63}$$

It is observed from (5.62) and (5.63) that $g_{3,\,bjt}$ and $g_{3,\,mos}$ have different signs. If their magnitudes are matched, the overall 3rd-order term can be canceled. Usually the absolute value of the bipolar's IM3 coefficient $g_{3,\,bjt}$ is much larger than that of the NMOS transistor's IM3 coefficient $g_{3,\,mos}$. Therefore, $g_{3,\,bjt}$ needs to be scaled down properly in order to provide maximum cancellation. This is achieved by resistive emitter degeneration as shown in Fig. 74. Before one can continue, it is important to show that the memory effect in the degenerated BJT is weak, so it does not change the phase of the non-linear terms. For this purpose, Volterra analysis of the degenerated BJT is carried out.

## 2. Volterra Analysis of Resistive-Degenerated BJT

Considering memory effect, the collector current $i_o$ of an emitter-degenerated bipolar transistor can be expanded using the Volterra series as:

$$i_o = B_0 + B_1 \circ v_i + B_2 \circ v_i^2 + B_3 \circ v_i^3 + \cdots \tag{5.64}$$

Where all the signal quantities are assumed to be in the sinusoidal form, e.g., $i_o$ means $A_{io} \cos(\omega t + \phi)$. In order to calculate Volterra kernel $B_k$, a large signal high frequency model is shown in Fig. 75. In this model, only the collector current non-linearity is considered. The base current $i_b$ and resistor $r_b$ (as shown in the dashed box) are ignored and all the capacitors are assumed linear. Base-collector capacitor $C_\mu$ will be combined with the MOS transistor's gate-drain capacitor $C_{gd}$, so it is excluded from the model. The difference between the emitter and collector current is also ignored, and $R_e$ is the degeneration resistor. A more strict analysis can be found in [38].

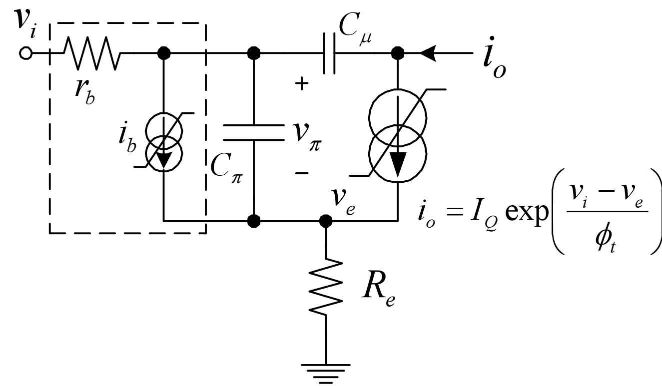



Fig. 75. Emitter-degenerated BJT large signal model

If the quiescent collector current is $I_Q$ then the dynamic collector current is

$$i_o = I_Q e^{\frac{v_i - v_e}{\phi_t}} \approx I_Q \left[1 + \frac{v_i - v_e}{\phi_t} + \frac{1}{2}\left(\frac{v_i - v_e}{\phi_t}\right)^2 + \frac{1}{6}\left(\frac{v_i - v_e}{\phi_t}\right)^3 + \cdots\right] \tag{5.65}$$

where $v_i$ and $v_e$ are AC signals. Using KCL at the emitter node:

$$(v_i - v_e)\, j\omega C_\pi + i_o = v_e g_e \tag{5.66}$$

where $g_e = R_e^{-1}$. Solving for $v_e$ and substituting it into (5.65) and replacing $i_o$ by its Volterra expansion (5.64):

$$\begin{aligned} &B_0 + B_1 \circ v_i + B_2 \circ v_i^2 + B_3 \circ v_i^3 + \cdots = \\ &\quad I_Q \Big\{ 1 + \tfrac{1}{I_t} \left[ -B_0 + (g_e - B_1) \circ v_i - B_2 \circ v_i^2 - B_3 \circ v_i^3 + \cdots \right] \\ &\quad + \tfrac{1}{2I_t^2} \left[ -B_0 + (g_e - B_1) \circ v_i - B_2 \circ v_i^2 - B_3 \circ v_i^3 + \cdots \right]^2 \\ &\quad + \tfrac{1}{6I_t^3} \left[ -B_0 + (g_e - B_1) \circ v_i - B_2 \circ v_i^2 - B_3 \circ v_i^3 + \cdots \right]^3 + \cdots \Big\} \end{aligned} \tag{5.67}$$

where $I_t = \phi_t (g_e + j\omega C_\pi) = \phi_t g_e (1 + j\omega C_\pi R_e)$. $B_k$ can be solved by equating the same order of $v_i$ at both sides of (5.67).

The zero-th order kernel or DC term $B_0$ is

$$B_0 = I_Q e^{-\frac{B_0}{\phi_t g_e}} \tag{5.68}$$

It shows that $B_0$ is smaller than the quiescent DC current $I_Q$ due to non-linearity. $\phi_t g_e$ is usually at the magnitude of around $1\,mA$ while $B_0$ is at the same magnitude of $I_Q$ which is about several tens of $\mu A$. Therefore $B_0$ can be approximated by

$$B_0 \approx I_Q \left( 1 - \frac{I_Q}{I_{t0}} \right) \approx I_Q \tag{5.69}$$

where $I_{t0} = \phi_t g_e$.

The 1st-order or linear kernel $B_1$can be shown to be

$$B_1(\omega) \approx \frac{g_m}{1 + g_m R_e} \left( 1 - j\frac{\omega}{\omega_\pi} \right) \tag{5.70}$$

where $\omega_\pi = 2\pi f_\pi = \frac{g_e}{C_\pi}$, $g_m = \frac{I_Q}{\phi_t}$. Typically $g_e \approx 0.05$, $C_\pi \approx 16fF$, so $f_\pi$ is at the order of several hundreds of GHz and for moderate operating frequencies $\frac{\omega}{\omega_\pi} < 0.01$

holds. Therefore, the phase angle of $B_1$ is very small (no greater than 1 degree) and for practical purposes, $B_1$ can be treated as frequency independent.

The 2nd-order kernel $B_2$ can be approximated by

$$B_2(\omega_1, \omega_2) \approx \frac{1}{2I_Q} \frac{g_m^2}{(1+g_m R_e)^3} \left[1 - j\frac{2(\omega_1+\omega_2)}{\omega_\pi}\right] \tag{5.71}$$

and its frequency dependence is also very weak.

The 3rd-order kernel $B_3$ is

$$B_3(\omega_1, \omega_2, \omega_3) \approx \frac{1}{6I_Q^2} \frac{g_m^3}{(1+g_m R_e)^5} (1 - 2g_m R_e) \left[1 - j\frac{3(\omega_1+\omega_2+\omega_3)}{\omega_\pi}\right] \tag{5.72}$$

For the worst case $\omega_1 = \omega_2 = \omega_3$, the phase angle of $B_3$ is no greater than 6 degrees. Thus $B_3$ can also be regarded as frequency independent. It should also be noticed that if the degeneration resistance is chosen right, $B_3$ can become zero or change polarity.

The above derivations have justified the memory effect in the resistive-degenerated BJT is very weak and can be ignored. Resistive degeneration can be used to scale the magnitude of the 3rd-order coefficient to match that of MOSFET. In this case (5.62) corresponds to

$$g_{3,\,bjt} = \frac{1}{6I_Q^2} \frac{g_m^3}{(1+g_m R_e)^5} (1 - 2g_m R_e) \tag{5.73}$$

Fig. 76 compares the simulated 3rd-order coefficients obtained at DC to the one at 3 GHz. The theoretical curves for DC and 3 GHz actually overlap with each other due to the extremely weak frequency independence. Because the theoretical analysis does not consider the BJT's base and emitter's extrinsic resistance, so the theoretical curves are shifted from the simulated ones, and the non-linearity predicted by the theoretical curves is a little bit larger, but the trend is well predicted. Fig. 77 shows the 3rd-order cancellation effect of the proposed hybrid configuration. The MOS and

BJT are biased separately and then their bias voltages are swept and their output currents are used to calculate the 3rd-order terms.

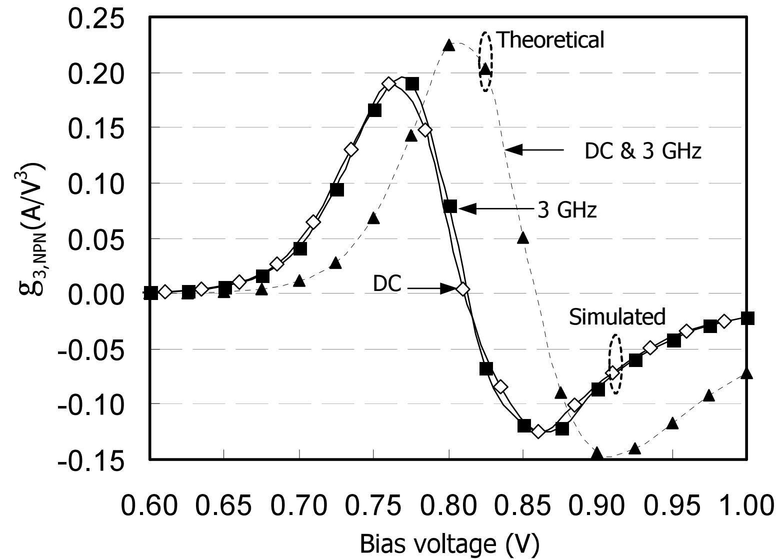



Fig. 76. Resistive-degenerated BJT 3rd-order coefficient at DC and 3GHz

## 3. Input Matching and Noise Contributions

The input matching network can be designed using the inductive source-degeneration technique [10]. The degeneration inductor will be implemented using bond wire. The gate inductance will be implemented in part from the bond wire and the other part from off-chip surface mount inductor. The on-chip inductor is usually not used at input because its size is usually large, so it will consume too much chip area. Another reason to put the gate inductor off-chip is that the on-chip inductor does not have a good quality factor, so the loss of the on-chip inductor will degrade the noise performance of the low noise amplifier.

Fig. 78 is the small signal equivalent circuit for the input impedance calcula-

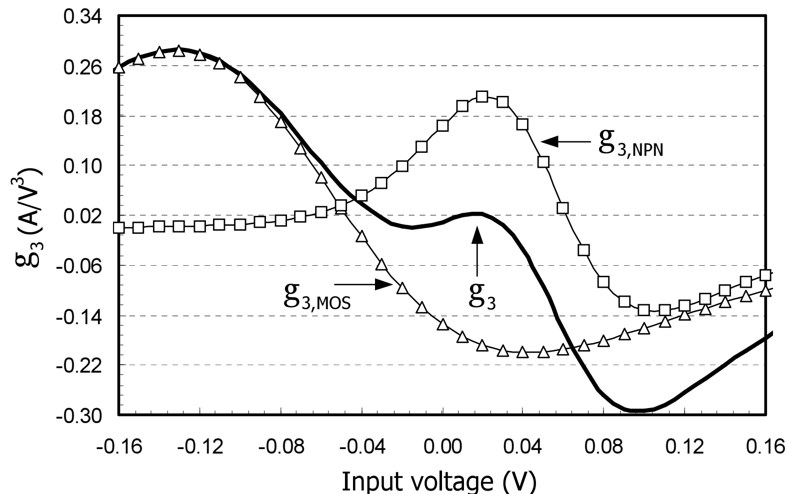



Fig. 77. 3rd-order terms of the BJT, NMOS and their combination

tion. $C_t$ accounts for the total capacitance between the gate and source of the MOS transistor $M_1$ in Fig. 74.

$$C_t = C_{gs1} + C_\pi \tag{5.74}$$

where $C_{gs1}$ is the gate-source capacitance of $M_1$, $C_\pi$ is the bipolar base-emitter capacitance. $g_\pi$ is the conductance introduced by the bipolar transistor and can be expressed as

$$g_\pi = \frac{1}{r_\pi} = \frac{I_Q}{\beta \phi_t} \tag{5.75}$$

The bipolar's emitter degeneration resistor $R_e$ is about $20 \sim 40\ \Omega$. This is a relatively small value and will be ignored in the following analysis for simplicity.

Assuming $\omega^2 C_t^2 \gg g_\pi$, the input impedance of the configuration can be derived as

$$Z_{in} \approx j\omega\left(L_s + L_g + \frac{g_m g_\pi L_s}{\omega^2 C_t^2}\right) + \frac{1}{j\omega C_t} + \frac{g_m}{C_t}L_s + \frac{g_\pi}{\omega^2 C_t^2} \tag{5.76}$$

It is observed that the bipolar transistor introduces a shunted RC network between the gate and source of the MOS transistor as shown in Fig. 78. The base-emitter

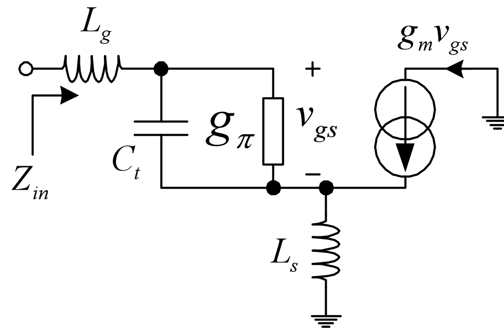



Fig. 78. Small signal circuit for input impedance calculation

capacitance $C_\pi$ will shift the input matching frequency to a lower frequency. The resistance $r_\pi$ in parallel with this additional capacitor is at the magnitude of several kilo-ohms. This resistance will vertically shift the S11 curve upward. Fig. 79 shows the simulated S11 plot with and without the bipolar transistor activated. Therefore, considering the effects of bipolar input impedance on the total input impedance, the input can still be matched to a specific value using inductive degeneration.

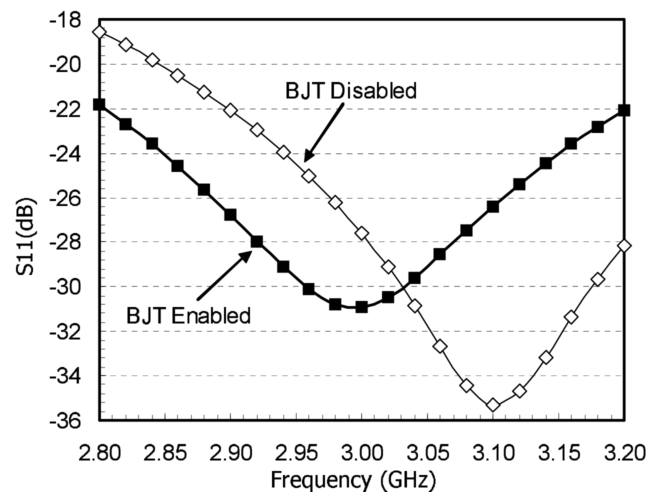



Fig. 79. Effect of bipolar transistor on input matching

Due to the low bias current, the bipolar contributes a small amount of noise to the whole circuit. Simulation shows that the bipolar transistor adds less than 2.4% to the overall noise of the circuit while the MOS transistor contributes about 14%. Table XIV lists the noise contribution ratios of different devices at 3 GHz. Here the input matching network is designed to achieve the best impedance match. The noise figure calculated from the values given in the table is about 2.2 dB.

Table XIV. Device noise contribution ratios of single-ended hybrid LNA († signal generator's internal resistance)

| Components | Noise ratio |
|---|---|
| $R_s^{\dagger}$ | 60% |
| MOS transistor | 14% |
| Bipolar transistor | 2.4% |
| Other devices | 23.6% |

### 4. Sensitivity to Bias Condition, Process Corners and Temperature

In the proposed linearization technique, an NPN transistor in a CMOS process is used. It is necessary to verify that the NPN transistor is fast enough to work at the RF frequency. Table XV lists the simulated $f_T$ of the NPN transistor and NMOS transistor used in Fig. 74 against process corners. It can be seen that the NPN transistor almost has the same speed as the NMOS transistor in FF and SS corners.

Figs. 80 gives the IIP3 of the NMOS-NPN combination with different bias conditions of the MOS and bipolar device. Two operation frequencies are given: 2.4 GHz and 3 GHz. The two test tones are placed 2 MHz apart. The two plots in Figs. 80 for different frequencies are almost the same. This means that IIP3 is not sensitive

Table XV. $f_T$ of a $2 \times 2$ NPN transistor

| Corner | NMOS | NPN |
|---|---|---|
| TT | 30GHz | 28GHz |
| FF | 53GHz | 35GHz |
| SS | 21GHz | 22GHz |

to operational frequency as proved previously.

It is observed from the figures that if the MOS transistor is biased at 0.685 V and the bipolar transistor is biased at 0.8 V, the configuration has the optimal IIP3 condition. If the MOS transistor gate is biased too low or too high, there will exist two IIP3 maxima, one for a lower bipolar base bias and the other for a higher bias. For $V_{bmos} = 0.685\,V$, if due to process variations, its value changes to 0.635 V or 0.700 V, the IIP3 does not vary significantly. If $V_{bnpn}$ is around 0.8 V, the IIP3 curves is pretty flat for $V_{bmos} = 0.635\,V$ and $V_{bmos} = 0.685\,V$. Notice that for $V_{bmos} = 0.7\,V$, the IIP3 at $V_{bnpn} = 0.8\,V$ has a local minima, so the variation of $V_{bnpn}$ within about $\pm 30\,mV$ will not degrade the IIP3.

Fig. 81 is the IIP3 versus bipolar base bias voltage against process corners. Here the MOS transistor is biased at 0.685 V. Process corners will make the optimal bias condition change, especially for the FF corner. The optimal bias voltage is shifted to a lower value for the FF corner. An On-chip corner dependent biasing scheme can be used to modify the biasing point.

Temperature behavior of the proposed configuration is shown in Fig. 82. It shows how the IIP3 profile changes with different bias voltage of the NPN transistor under different temperatures. The optimal biasing point of the bipolar transistor shifts with temperature as shown in Fig. 83. This biasing profile can be realized by deriving the

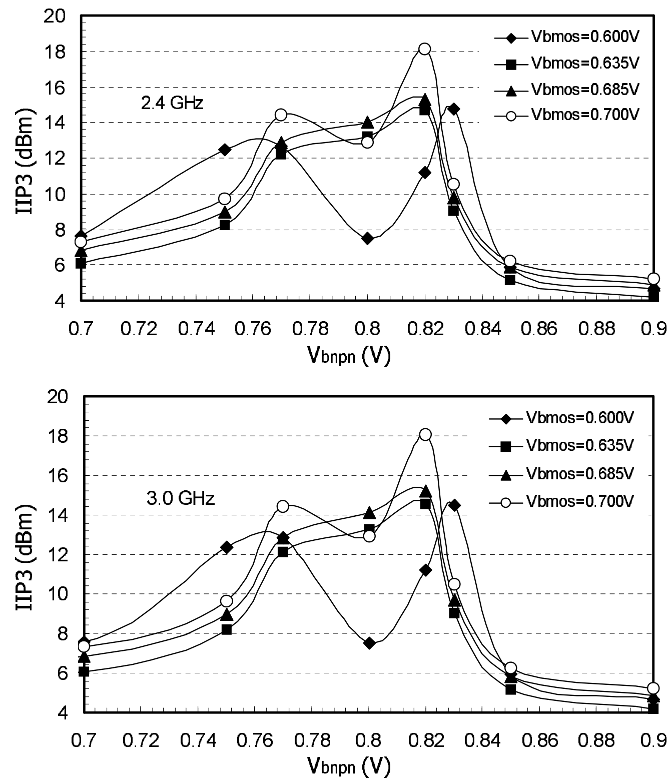



Fig. 80. IIP3 of NMOS-NPN combination vs. bias conditions

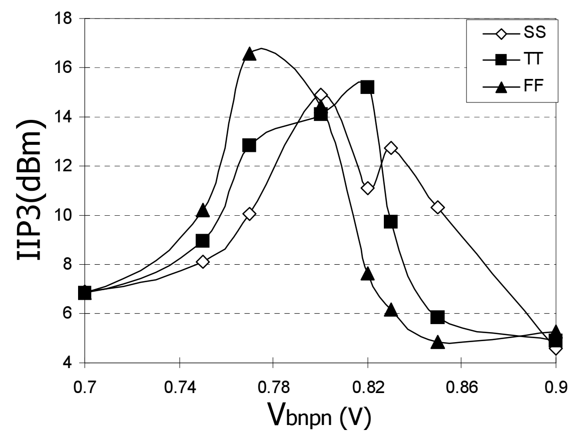



Fig. 81. IIP3 against process corners

base bias voltage from a PTAT current source running through a resistor. The NMOS transistor should be biased using a constant -$g_m$ biasing circuit to keep the $g_m$of the circuit constant over temperature variations.

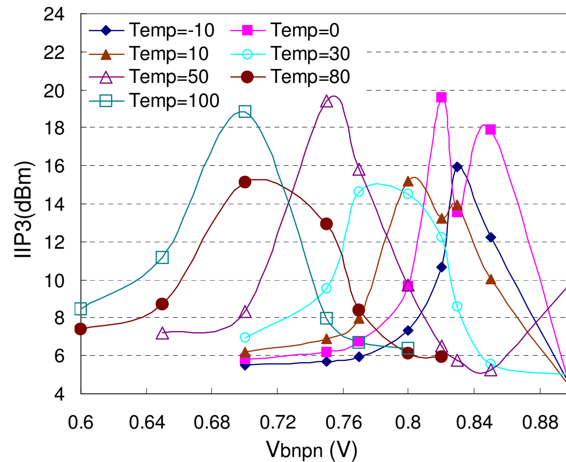



Fig. 82. IIP3 temperature behavior

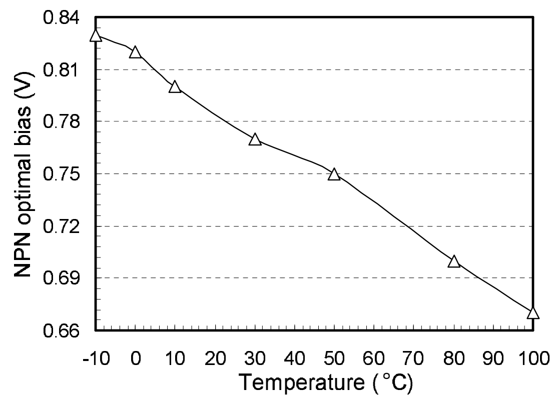



Fig. 83. NPN transistor optimal biasing profile

This single-ended hybrid LNA's design flow is depicted in Fig. 84. To begin with, a single-ended inductive-degenerated LNA is design according to the provided speci-

fications. In this stage, the linearity is not considered. Then an emitter-degenerated BJT is added to the circuit. The BJT and degeneration resistor's size and bias current can be determined from the MOS transistor and BJT's 3rd order coefficient simulation plot. After optimal conditions are found, the input impedance matching inductors $L_g$ and $L_s$ need to be adjusted to accommodate the added devices. Also the load should be fine-tuned to make gain return to the specification. The overall simulation verification is then carried out and the design usually needs to iterate to finally meet the required specifications.

## 5. The Differential Configuration

The IIP2 shown Fig. 71 almost has its worst value for the best IIP3. This is due to the fact that for different gate bias, although the 3rd-order terms can have different signs, so they can be canceled out by combining the current, the 2nd-order terms, however, will always have the same signs. By adding the output current of the main and auxiliary transistor, the 2nd-order terms will be added constructively, which will deteriorate IIP2. The single-ended hybrid LNA configuration also has this problem. In order to simultaneously keep the 2nd-order performance and provide 3rd-order compensation, a differential structure is proposed as in Fig. 85.

The 3rd-order Volterra kernel $H_3$ of a differential pair can be expressed as [37]

$$H_3\left(\omega_1,\, \omega_2,\, \omega_3\right) \approx g_3 \left(1 - j\frac{\omega_1 + \omega_2 + \omega_3}{3\omega_p}\right) \tag{5.77}$$

where $g_3$ is the 3rd-order coefficient of the Taylor series at DC. $\omega_p = \frac{g_m}{C_p}$. $g_m$ is the differential pair low frequency transconductance. $C_p$ is the total capacitance of the common-source or common-emitter node and it is assumed to dominate the memory effect of a differential pair. Usually $C_p$ is proportional to the size of the tail current source and thus, to the tail current. $g_m$ is also proportional to the tail current.

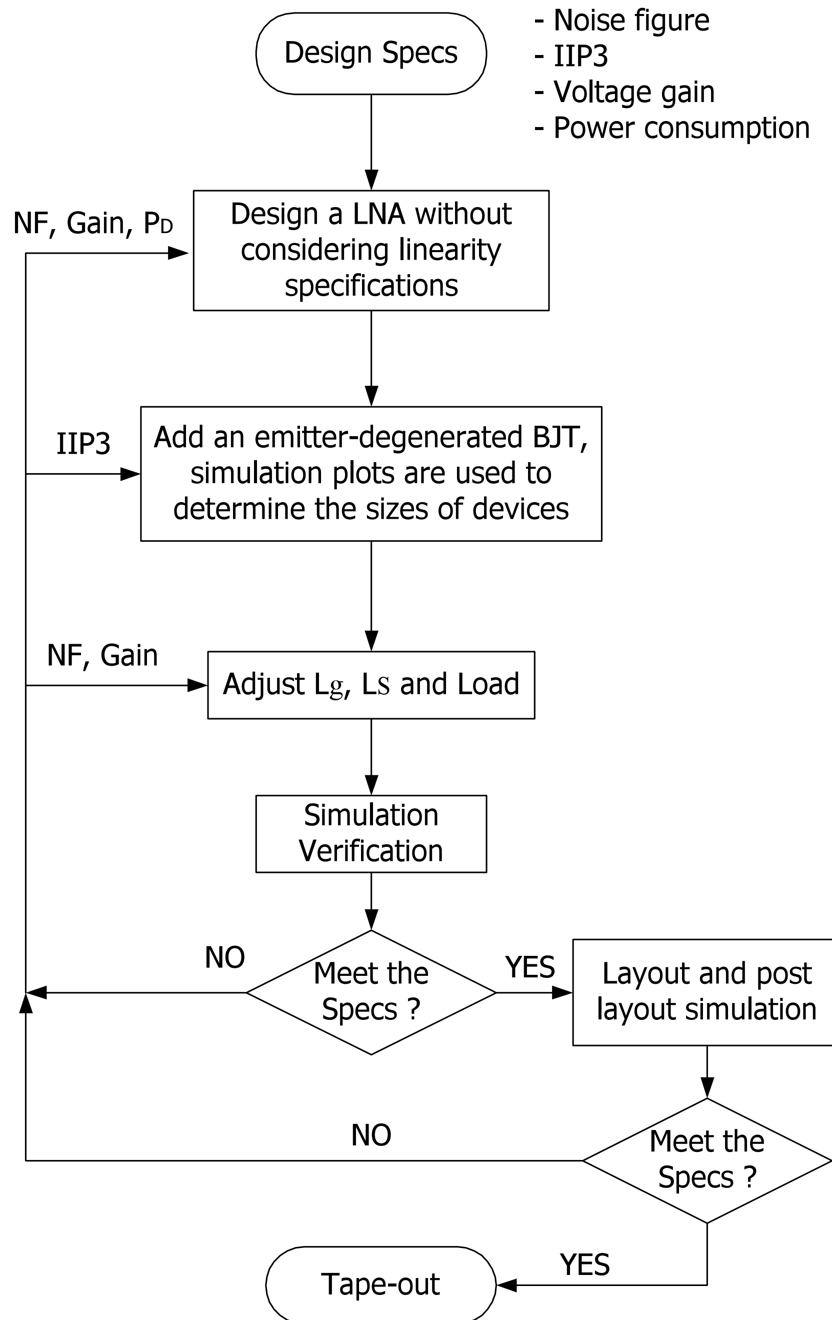



Fig. 84. Mixer design flow chart

Therefore, $\omega_p$ for MOS is at the same order as that for the BJT. The additional phase shift in $H_3$ for MOS is almost the same as that of BJT, i.e. for a bipolar or MOS differential pair, their 3rd-order terms will have the same sign. However, as mentioned before, for the same current consumption, bipolar's 3rd-order term is much larger than that of the MOS transistor. So with reduced bias current and resistive emitter degeneration, bipolar's 3rd-order term can be made to match that of the MOS transistor. By subtracting the output of the MOS and bipolar differential pair, the 3rd-order term can be canceled without significantly reducing the overall gain.

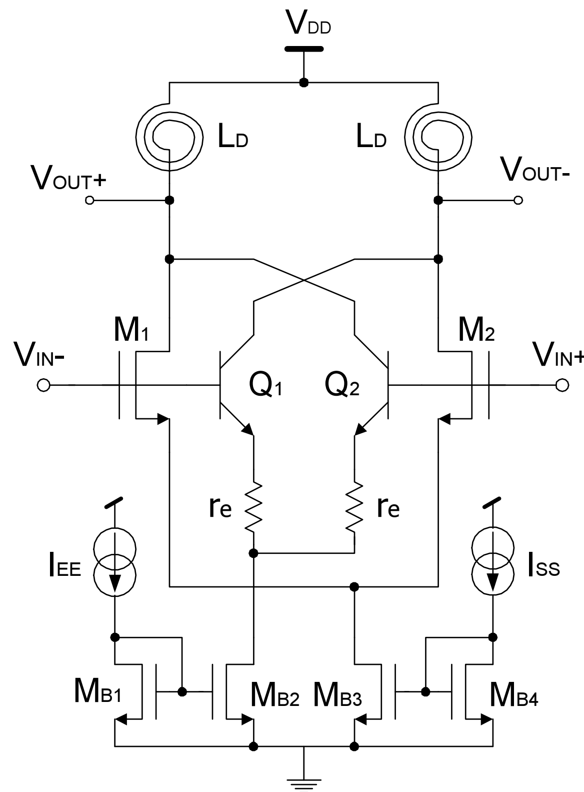



Fig. 85. Linearized differential LNA using bipolar differential pair

In the ideal case, where the matching between transistors is perfect, there should be no 2nd-order non-linearity at the biasing point of the differential pair. But in

practise, mismatch exists, and can be modeled as a biasing offset. The proposed differential method can not make the 2nd-order non-linearity substantially smaller than a single differential pair, but it can expand the input range for a small 2nd-order term, thus becoming more tolerant to the bias offset or device mismatch. On the other hand, if single-ended output is chosen from the differential structure, simulation shows that IIP2 can be improved by about 10 dB compared to without bipolar differential pair linearization. Fig. 86 shows the 2nd and 3rd-order non-linear terms of the MOS and bipolar differential pairs together with overall non-linearity terms. The 3rd-order cancellation and 2nd-order range expansion are easily identified from the plots.

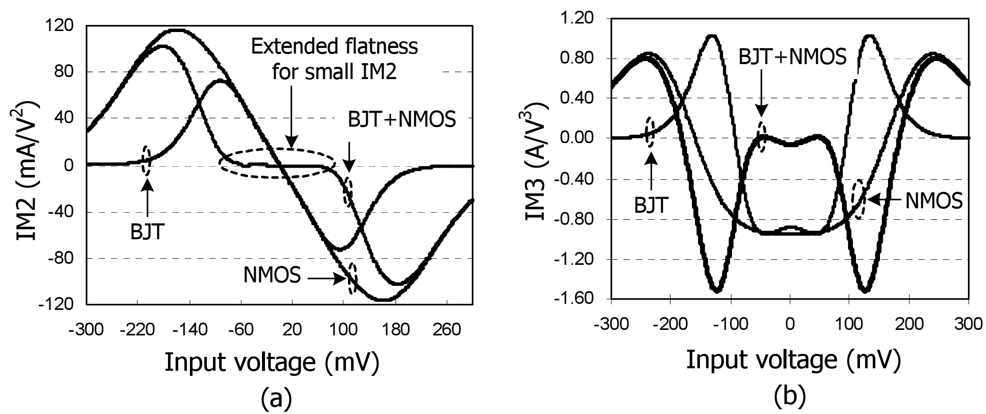



Fig. 86. (a) 2nd-order input range expansion and (b) 3rd-order cancellation of the proposed differential LNA

The differential structure uses a destructive combination of MOS and bipolar differential pairs. The output currents coming from the bipolar differential pair is subtracted from that of the MOS differential pair. While in the single-ended structure, the output currents generated from the MOS and bipolar transistors are constructively combined. This subtle difference makes the input matching of differential structure using inductive degeneration unfeasible. The degeneration inductor will

probably introduce an additional differential phase shift between the MOS and bipolar differential pairs which will deviate the non-linearity cancellation effect. So the input impedance matching is implemented using an LC network.

Again due to the destructive combine of MOS and bipolar pairs, the bipolar's noise contribution is more pronounced compared to the signal-ended structure. It contributes about 15% to the overall noise of the circuit. Table XVI shows how the noise contribution ratios break-down into different components. The noise figure calculated from the values given here is 3.7 dB.

Table XVI. Noise contribution ratios of differential configuration († signal generator's internal resistance)

| Components | Noise ratio |
|---|---|
| $R_s^{\dagger}$ | 42% |
| MOS transistors | 10% |
| Bipolar transistors | 15% |
| Other devices | 33% |

The differential hybrid LNA's design procedure is similar to the single-ended one previously shown in Fig. 84.

## D. Measurement Results and Comparisons

The single-ended and differential LNA's were designed using TSMC 0.18 $\mu m$ RF CMOS process. Testing chips were fabricated through MOSIS MEP program and packed in a QFN package. The die photomicrographs of these two circuits are shown in Figs. 87. The active size is 390 $\mu m$ × 390 $\mu m$ for the single-ended LNA and is 620 $\mu m$ × 490 $\mu m$ for the differential one.

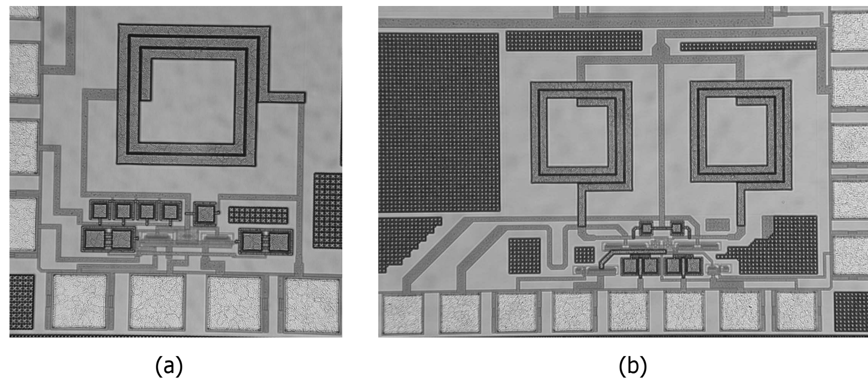



Fig. 87. Die photomicrographs of hybrid linearized LNA's (a) single-ended (b) differential

Fig. 88 shows the IIP3 plot of a single-ended linearized LNA using the linearization technique depicted in Fig. 74. Out-of-band termination is realized using the tuned LC tank as the load [49]. The two tones are put at 2700 MHz and 2701 MHz respectively. The S-parameter measurement plots are shown in Fig. 89. The S11 is better than -10 dB from 2.6 GHz to 2.9 GHz. Due to the single-ended structure and lack of cascoded transistor, the reverse isolation (S12) is not very good. The power gain is about 6.4 dB, noise figure is measured to be 2.1 dB around 2.7 GHz. It draws 6.4 mA from a single 1.2 V power supply.

The differential version (Fig. 85) of the hybrid LNA's operation frequency is centered around 2.5 GHz. Fig. 90 is the measured IIP3 plots of the circuit with and without the cancellation bipolar pair activated. The two testing tones are put at 2500 MHz and 2501 MHz respectively. When the bipolar pair is enabled, the fundamental term is reduced by about 2 dB and the IM3 term is reduced by 12.5 dB, therefore about 5 dB IIP3 improvement can be obtained with the bipolar cancellation pair. Fig. 91 is the measured S-parameters of the differential LNA. The S11 is better than -9 dB from 2.4 GHz to 2.6 GHz. The circuit's reverse isolation measured by S12

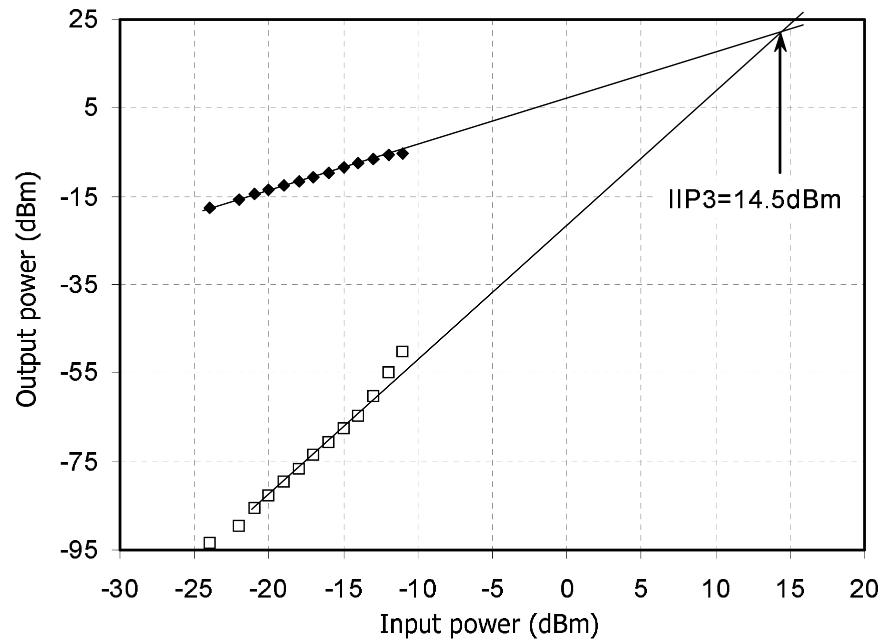



Fig. 88. IIP3 of the single-ended bipolar linearized LNA

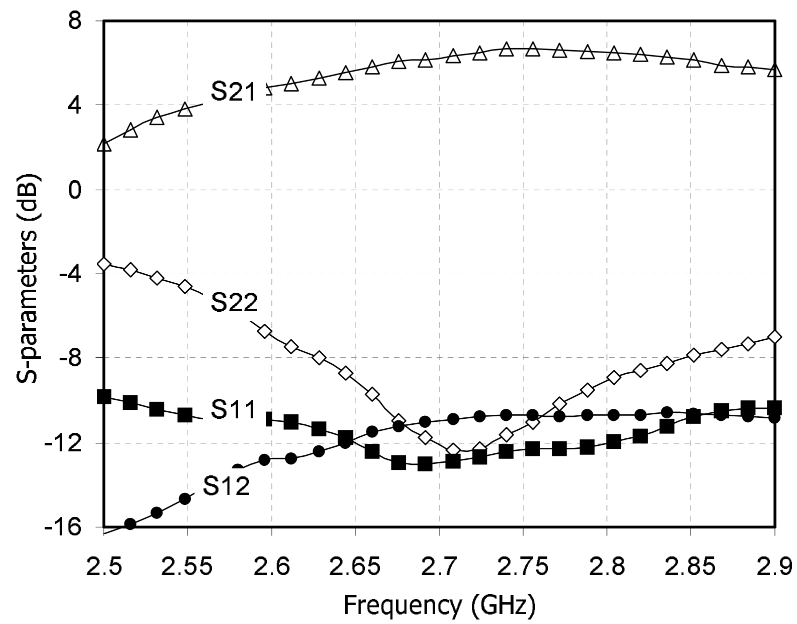



Fig. 89. Linearized single-ended LNA S-parameters

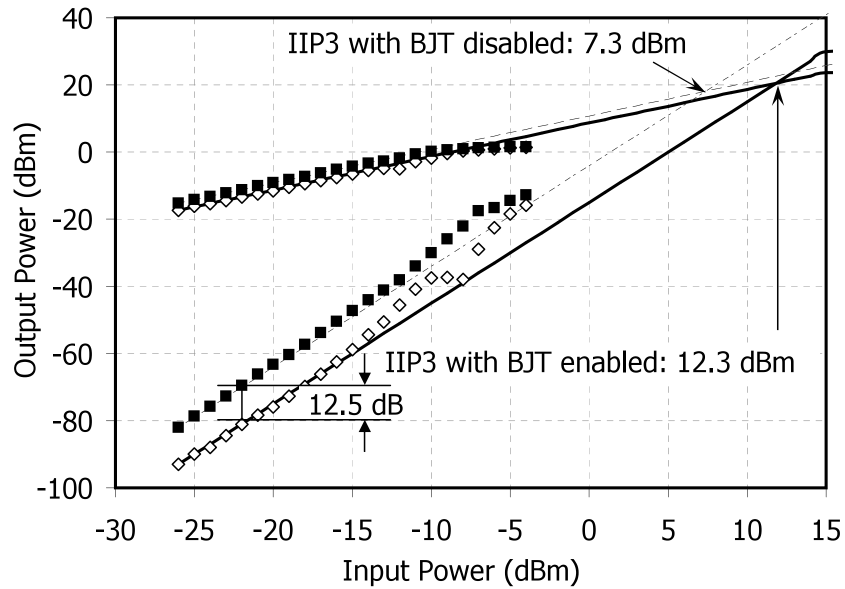



Fig. 90. IIP3 of differential LNA with and without bipolar cancellation pair activated

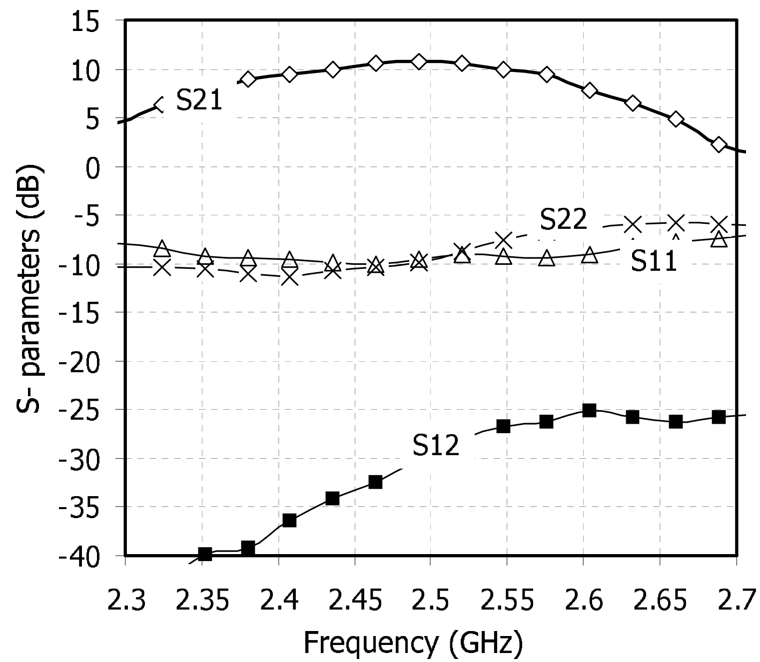



Fig. 91. Linearized differential LNA S-parameters

is much better than that of the single-ended LNA. The power gain is about 10 dB. Noise figure is measured to be 3.4 dB around 2.5 GHz. This differential LNA draws 11 mA from a single 1.8 V power supply. Because the differential structure inherently suppresses the second-order non-linear term, out-of-band termination is not required as in the single-ended case.

Table XVII compares the proposed technique with reported high linearity LNA's. The figure of merit (FOM) [50] is defined as

$$FOM = \frac{P_{IIP3} \times G}{(F-1)\, P_D} \tag{5.78}$$

where $P_{IIP3}$, $G$, $F$ and $P_D$ are the input-referred 3rd-order intercept point, power gain, noise factor and power dissipation respectively.

Table XVII. Comparison of the proposed linearization implementation with the state-of-the-art linear LNA's in the literature

| | Freq. (GHz) | Gain (dB) | NF (dB) | IIP3 (dBm) | Power (mW) | FOM |
|---|---|---|---|---|---|---|
| Single-ended [41] | 0.9 | 10 | 2.85 | 15.6 | 21.1 | 18.5 |
| Single-ended [proposed] | 2.7 | 6.4 | 2.1 | 14.5 | 8.9 | 22.8 |
| Differential [49] | 0.9 | 5 | 2.8 | 18 | 45 | 4.9 |
| Differential [proposed] | 2.5 | 10 | 3.4 | 12.3 | 19.8 | 7.2 |

The proposed linearized LNAs can work at a higher frequency and achieve an improved figure-of-merit. This is due to the fact that the emitter-degenerated BJT has a better controlled 3rd-order non-linearity by the degeneration resistance, and it requires less current to compensate the MOST's non-linearity. Therefore, the circuit can be designed without the BJTs first, and then adding the BJTs for linearization.

Note that the proposed design is implemented in CMOS technology, the bipolar device has a limited performance. One may wonder what if a true BiCMOS technology is used. To this end, a single-ended hybrid LNA is simulated using IBM 0.25$\mu m$ BiCMOS technology. Thanks to the superior bipolar transistor available in the BiCMOS process, the noise figure of this LNA is less than 1.2 dB at 3.0 GHz with the power gain of 9.5 dB and IIP3 of 12 dBm. The power dissipation is 6.6 mW. Thus the BiCMOS design has a figure-of-merit 67, which is much better than its CMOS counterpart.

# CHAPTER VI

# A MUTUAL-COUPLED DEGENERATED LNA AND ITS EXTENSION TO CONCURRENT DUAL-BAND OPERATION

Nowadays, wireless technologies are advancing faster then ever. The diverse range of wireless applications have demands on communication systems for versatility and flexibility. While different wireless applications are usually operating in different frequency bands, recently dual-band or even multi-band transceivers gain a lot of interest. Being the first active block in the receiver chain, the low noise amplifier (LNA) has to have multi-band capabilities. Most of the current wireless systems are narrow band. A wide band LNA can be used to cover a wide frequency range, but it consumes a large amount of power and at the same time amplifies interferences which are not on the actual reception band. Therefore, one would like the LNA's transfer function curve to looks like the superposition of two or more narrow band LNAs. In order for the front-end LNA to cover more than one frequency band, the matching network must be able to provide degrees of freedom equal to the number of bands covered. Methods already reported in the literatures are: (i) device switching [33], (ii) concurrent matching [51], and (iii) device switching plus concurrent matching [52]. The switched method can provide optimal design for every band while the concurrent approach can receive more than one band at the same time.

Current wireless mobile terminals provide dual- or multi-band operations mainly for cellular capacity reasons. While in the future multi-band/multi-mode terminals will be enabled to access different systems providing various services. Fig. 92 shows the mobile station receiver band in the spectrum for a clear review of frequency allocations. The frequency numbers in the figure are shown in MHz. Numbers shown

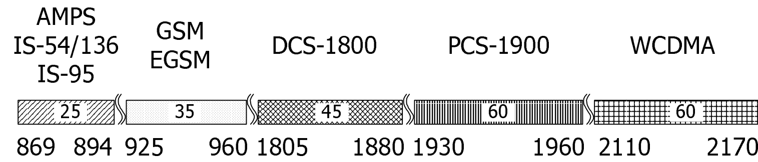



Fig. 92. Mobile station receiver band

in the rectangle representing the band are its bandwidth. A dual-band receiver front-end typically consists of a dual-band antenna, followed by a monolithic dual-band filter and dual-band LNA ate provides gain and impedance match at two bands.

An input impedance match method using mutual coupled inductors will be studied first and measurement results for an LNA working in the 900 MHz GSM band will be presented. Then the extension of a concurrent dual-band LNA will be proposed. The advantage of concurrent operation is it does not need any dual-band switch or duplexer and has maximum front-end circuit sharing which reduces the silicon area. Concurrent reception is more desired when two bands provide different services such as voice and data.

## A. Principle of Impedance Match Using Mutual Inductance

The inductive source-degenerated LNA is a very popular architecture in the integrated CMOS RF and microwave circuit design [10]. It nicely trade-offs among input matching, noise figure and gain specifications. A cascoded structure is usually used in the LNA to reduce Miller effect on input impedance and improve reverse isolation. An inter-stage inductor is used in [53] and [54] to provide inter-stage matching. In this section, the mutual coupling between the degeneration and inter-stage inductors is added to obtain another degree of freedom for the design.

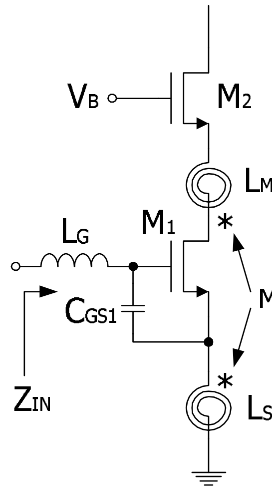



Fig. 93. Matching utilizing mutual inductance

### 1. Input Impedance

Fig. 93 shows the idea of using mutual inductance in the LNA input matching network. Its small signal circuit for input impedance calculation is depicted in Fig. 94. It is easily seen that the input impedance $Z_{in}$ can be written as the sum of inductor $L_G$'s impedance $sL_G$ and the impedance $Z_X$ looking into the circuit after inductor $L_G$:

$$Z_{in} = sL_G + Z_X \tag{6.1}$$

Impedance $Z_X$ can be calculated by solving the following equations

$$\begin{aligned} V_{s1} &= sL_s\left(I_i + I_{d1}\right) \pm I_{d1} sM \\ V_{G1} &= V_{s1} + \tfrac{I_i}{sC_{GS1}} \\ I_{d1} &= g_{m1} V_{GS1} = g_{m1} \tfrac{I_i}{sC_{GS1}} \\ Z_X &= \tfrac{V_{G1}}{I_i} \end{aligned} \tag{6.2}$$

where $M$ is the mutual inductance between $L_S$ and $L_M$. Its polarity is determined

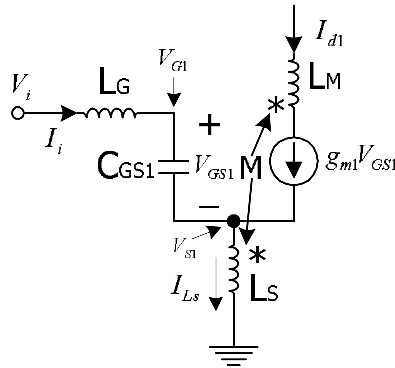



Fig. 94. Input impedance small signal circuit

by how the magnetic coupling is constructed. Coupling coefficient $k$ is defined as

$$k = \frac{M}{\sqrt{L_S L_M}} \tag{6.3}$$

A typical range for the $k$-factor achievable in silicon designs is $0.6 \leq k \leq 0.95$. $Z_{in}$ is given by

$$Z_{in} = s\left(L_G + L_S\right) + \frac{1}{sC_{GS1}} + \frac{g_{m1}}{C_{GS1}}\left(L_S \pm M\right) \tag{6.4}$$

For the mutual coupling polarity shown in Fig. 94 by the asterisk, minus sign will be assigned to the third term in (6.4). This impedance expression resembles the one without mutual inductance coupling except that the real part is modified by the mutual inductance. Under resonant frequency

$$\omega_o = \frac{1}{\sqrt{\left(L_G + L_S\right) C_{GS1}}} \tag{6.5}$$

$Z_{in}$ only presents resistance

$$R_{in} = \omega_T \left(L_S \pm M\right) \tag{6.6}$$

where $\omega_T = \frac{g_{m1}}{C_{GS1}}$.

For long channel approximation

$$\omega_T \propto \frac{1}{L^2}\mu\left(V_{GS1} - V_{th1}\right) \tag{6.7}$$

The device channel length $L$ is usually chosen to be process minimum for high frequency performance. $C_{GS1}$ is calculated from the optimum input quality factor $Q_I = \frac{1}{\omega_o C_{GS1} R_s}$. Therefore the device width $W$ is obtained by resolving$C_{GS1} = \frac{2}{3}WLC_{ox}$. Gate over drive voltage $V_{GS1} - V_{th1}$ can be fixed by power consumption constraint $I_D \propto \frac{W}{L}\left(V_{GS1} - V_{th1}\right)^2$or linearity requirement $IIP3 \propto V_{GS1} - V_{th1}$.

For a conventional inductive degenerated LNA, difficulty may arise from using high bias level to obtain high input linearity. Input match condition requires $R_s = \omega_T L_S$, where $R_S$ is usually $50\,\Omega$ or $75\,\Omega$. For high gate bias level, $V_{GS1} - V_{th1}$ is large, therefore $\omega_T$ is also large. This situation will probably require a very small degeneration inductance $L_S$ which is hard to design or already comparable to the bond wire inductance. Of course, bond wire can be used but it is less controllable by the designer and is affected by the placement of the die in the package which is unknown at the very beginning of the design stage. Small $L_S$ will also require large $L_G$ to provide the same resonant frequency $\omega_o$. While a large inductance generally has a larger series resistance which will degrade the noise figure of the LNA directly, because $L_G$ is series directly with the gate. The presence of the mutual inductance offers another degree of freedom for input impedance matching. By choosing the negative sign in (6.4), $R_s = \omega_T\left(L_S - M\right)$, larger $L_S$ can be used. If the design is targeting minimum current consumption and if $V_{GS1} - V_{th1}$ is small, a large $L_S$ may be required. In this case, the positive sign in (6.4) can be used to enable smaller degeneration inductance $L_S$.

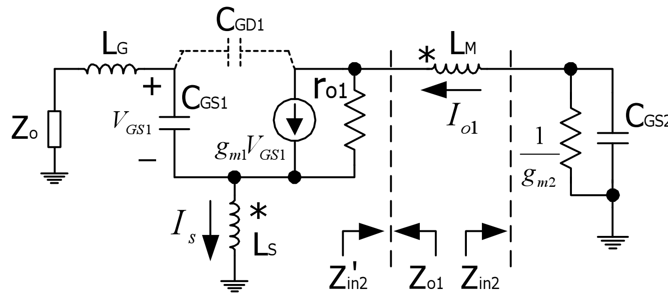



Fig. 95. Inter-stage impedance

2. Interstage Impedance

In Fig. 93, cascoded transistor $M_2$ reduces Miller effect and provides output-input isolation. Adding $L_M$ can provide inter-stage impedance transformation and help to further reduce Miller capacitance of $M_1$ therefore increasing reverse isolation. Fig. 95 shows the small signal equivalent circuit for inter-stage impedance calculation. $Z_o$ is the source impedance. $\frac{1}{g_{m2}}$ in parallel with $C_{GS2}$ represents the input impedance of the cascoded stage.

Intuitively, the common-gate configured transistor $M_2$ has an input resistance of about $\frac{1}{g_{m2}}$ (actually it is larger than $\frac{1}{g_{m2}}$ because the drain of $M_2$ is not shorted to AC ground. $\frac{1}{g_{m2}}$ is good approximation for low load impedance though). The $M_2$'s gate-source capacitance $C_{GS2}$ together with inductor $L_M$ forms a shunt-series impedance transformation network. This network transforms $\frac{1}{g_{m2}}$ to a smaller value. Therefore the voltage gain from the gate of $M_1$ to its drain terminal will be decreased due to a reduced load impedance, thus reducing the Miller feedback.

It can be shown that $M_1$'s drain current $I_{o1}$ and source current $I_s$have the following relationship

$$I_s = F_1 I_{o1} \tag{6.8}$$

$$F_1 = \frac{sL_G \pm sM + \frac{1}{sC_{GS1}} + Z_o}{sL_G + sL_s + \frac{1}{sC_{GS1}} + Z_o} \tag{6.9}$$

At input matched condition, $F_1$ can be simplified as

$$F_1 = 1 - j\frac{\omega_o}{\omega_T} \tag{6.10}$$

The impedance looking into the drain of $M_1$ is

$$Z_{o1} = 2r_{o1} + \frac{\omega_o^2}{\omega_T} L_s + j\frac{\omega_o}{\omega_T} Z_o \tag{6.11}$$

Usually $\omega_o$ is far below $\omega_T$ and $Z_o$ is $50\,\Omega$ or $75\,\Omega$, so the reactive part of (6.11) will be relatively much smaller than the resistive part. $Z_{o1}$ will be nearly resistive.

Looking away from the drain of $M_1$, the impedance can be shown to be

$$Z'_{in2} = \frac{1}{g_{m2}} \frac{1}{1 + \left(\frac{\omega_o}{\omega_{T2}}\right)^2} \pm \left(\frac{\omega_o}{\omega_T}\right)^2 \omega_T M + j\omega_o \left[(L_M \pm M) - \frac{1}{g_{m2}} \frac{\omega_{T2}}{\omega_{T2}^2 + \omega_o^2}\right] \tag{6.12}$$

where $\omega_{T2} = \frac{g_{m2}}{C_{GS2}}$. By choosing a proper $\omega_{T2}$, $Z'_{in2}$ can have only the resistive part and its value is smaller than $\frac{1}{g_{m2}}$. The effect is that there will be more current pumped into the cascoded stage thus improving efficiency.

It can be shown that the transconductance of the first stage does not get affected by the mutual coupling

$$G_{m1} = Q_I g_{m1} \tag{6.13}$$

By choosing negative polarity of the inductive coupling, the voltage gain from $M_1$'s gate to its drain is

$$A_{V1} = G_{m1} \left[ \left(2r_{o1} + \frac{\omega_o^2}{\omega_T} L_s\right) \parallel \left( \frac{1}{g_{m2}} \frac{1}{1 + \left(\frac{\omega_o}{\omega_{T2}}\right)^2} - \left(\frac{\omega_o}{\omega_T}\right)^2 \omega_T M \right) \right] \tag{6.14}$$

This gain can be made smaller than that without mutual inductive coupling therefore further reducing the Miller effect.

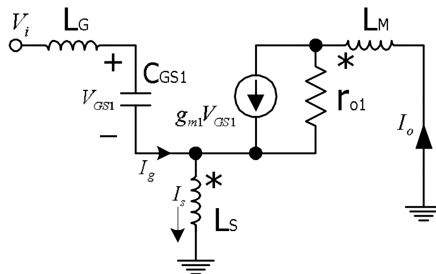



Fig. 96. Overall transconductance

3. Effective Transconductance

The effective transconductance from the input terminal to the current flowing through inductor $L_M$ can be found by observing the small signal equivalent circuit in Fig. 96.

At around frequency $\omega_o$, one can write the following equations

$$\begin{aligned} V_{GS1} &= Q_I V_i \\ I_s &= I_o + I_g \\ I_g &= j\omega_o C_{GS1} V_{CGS1} \\ V_i - (V_{GS1} + j\omega_o L_G I_g) &= j\omega_o L_s I_s \pm j\omega_o M I_o \end{aligned} \tag{6.15}$$

The effective transconductance $G_m$ can be shown to be

$$G_m = \frac{I_o}{V_i} = \frac{1 + Q_I\left(\frac{\omega_o^2}{\omega_{oG}^2} + \frac{\omega_o^2}{\omega_{oS}^2} - 1\right)}{j\left(\frac{\omega_o}{\omega_T}\right) Z_o} \tag{6.16}$$

where

$$\omega_{oG} = \frac{1}{\sqrt{L_G C_{GS1}}} \tag{6.17}$$

and

$$\omega_{oS} = \frac{1}{\sqrt{L_S C_{GS1}}} \tag{6.18}$$

It is easy to shown from (6.5), (6.18) and (6.17) that $\frac{\omega_o^2}{\omega_{oG}^2} + \frac{\omega_o^2}{\omega_{oS}^2} = 1$, so $G_m$ is further

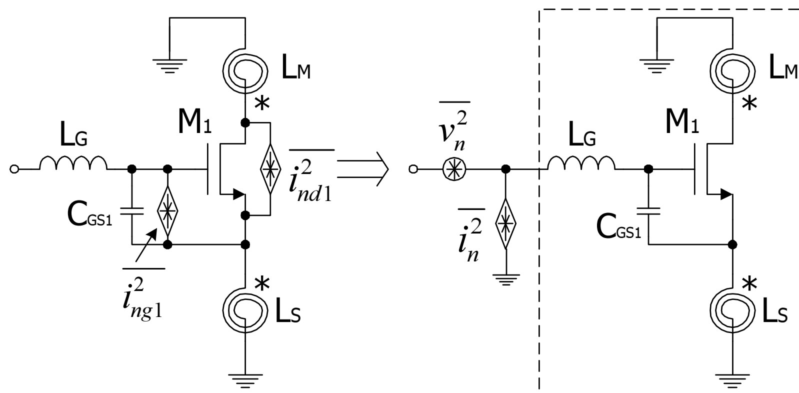



Fig. 97. Equivalent input noise sources of the inductive coupled LNA

simplified into

$$G_m = -j\left(\frac{\omega_T}{\omega_o}\right)\frac{1}{Z_o} \tag{6.19}$$

It is seen that the phase of effective transconductance is -90 degrees. In order to have large effective transconductance, the cut-off frequency of the RF transistor $M_1$ should be large or equivalently, $M_1$'s gate over-drive voltage need to be raised, which means larger power consumption.

## 4. Noise Analysis

Since in an LNA with the cascoded structure, the cascoded MOSFET has a little effect on the noise performance of the whole circuit [55], its noise contribution will be ignored in the following noise analysis. $M_1$'s drain noise current $\overline{i_{nd1}^2}$ and induced gate noise current $\overline{i_{ng1}^2}$ will be represented by a noise voltage $\overline{v_n^2}$ and noise current $\overline{i_n^2}$ as shown in Fig. 97.

It can be shown that at around operation frequency $\omega_o$,

$$i_n = i_{ng1} + i_{nd1} j \frac{\omega_o}{\omega_T} \tag{6.20}$$

and

$$v_n = -\frac{i_{ng1}}{j\omega_o C_{gs1}} \tag{6.21}$$

Since the mutual inductance does not appear explicitly in the above equations, it does not make the expression forms of input equivalent noise sources different from a regular inductive degenerated LNA.

Noise parameters $G_u$, $R_n$, $G_c$ and $B_c$ are found to be

$$G_u = \frac{\gamma}{\alpha} g_{m1} \left(\frac{\omega_o}{\omega_T}\right)^2 \left(1 - |c|^2\right) \tag{6.22}$$

$$R_n = \frac{\alpha\delta}{5g_{m1}} \tag{6.23}$$

$$G_c \approx 0 \tag{6.24}$$

and

$$B_c = -\omega_o C_{gs1} \left(1 + \frac{|c|}{\alpha}\sqrt{\frac{5\gamma}{\delta}}\right) \tag{6.25}$$

The minimum noise figure obtained for this circuit configuration is

$$F_{mim} = 1 + \frac{2}{\sqrt{5}}\frac{\omega_o}{\omega_T}\sqrt{\gamma\delta\left(1 - |c|^2\right)} \tag{6.26}$$

This minimum NF requires

$$G_s = G_{opt} = g_{m1}\frac{\omega_o}{\omega_T}\sqrt{\frac{5\gamma}{\alpha^2\delta}\left(1 - |c|^2\right)} \tag{6.27}$$

and

$$B_s = B_{opt} = \omega_o C_{gs1} \left(1 + \frac{|c|}{\alpha}\sqrt{\frac{5\gamma}{\delta}}\right) \tag{6.28}$$

Under perfect impedance match condition, $G_s = \frac{1}{Z_o} = \frac{1}{R_s}$ and $B_s = 0$, the noise figure is

$$F = 1 + \frac{\alpha\delta}{5}\frac{1}{g_{m1}R_s} + \left(\frac{\omega_o}{\omega_T}\right)^2 g_{m1}R_s \left(\frac{\alpha\delta}{5} + \frac{\gamma}{\alpha} + |c|\sqrt{\frac{4}{5}\gamma\delta}\right) \tag{6.29}$$

Several observations can be obtained from (6.29). The $\left(\frac{\omega_o}{\omega_T}\right)^2$ term shows the noise figure will increase with operation frequency if other parameters keep constant. If there is no correlation between the gate and drain noise current, i.e. $c = 0$, the noise figure could be smaller. There is an optimal value of $g_{m1}R_s$ product which makes the noise figure under perfect impedance match minimum. For $\gamma = 2$, $\delta = 4$, $\alpha = 0.85$ and $|c| = 0.4$:

$$g_{m1}R_s|_{opt} \approx 0.4\left(\frac{\omega_T}{\omega_o}\right) \tag{6.30}$$

Fig. 98 is the schematic of the proposed LNA, and Fig. 99 shows a detailed design flowchart. It is very similar with the design flow in Fig. 15, Chapter II. The major difference is the input impedance matching involves mutual inductance and the design trade-offs also include the mutual coupling factor.

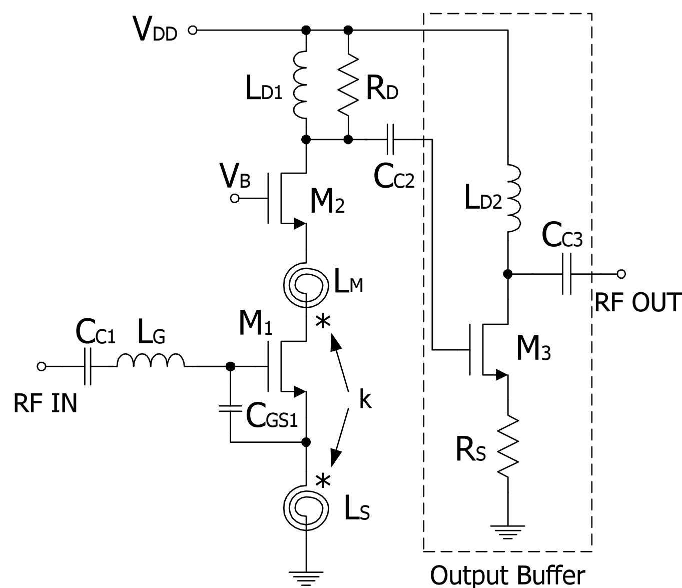



Fig. 98. The proposed mutual-coupled degenerated LNA

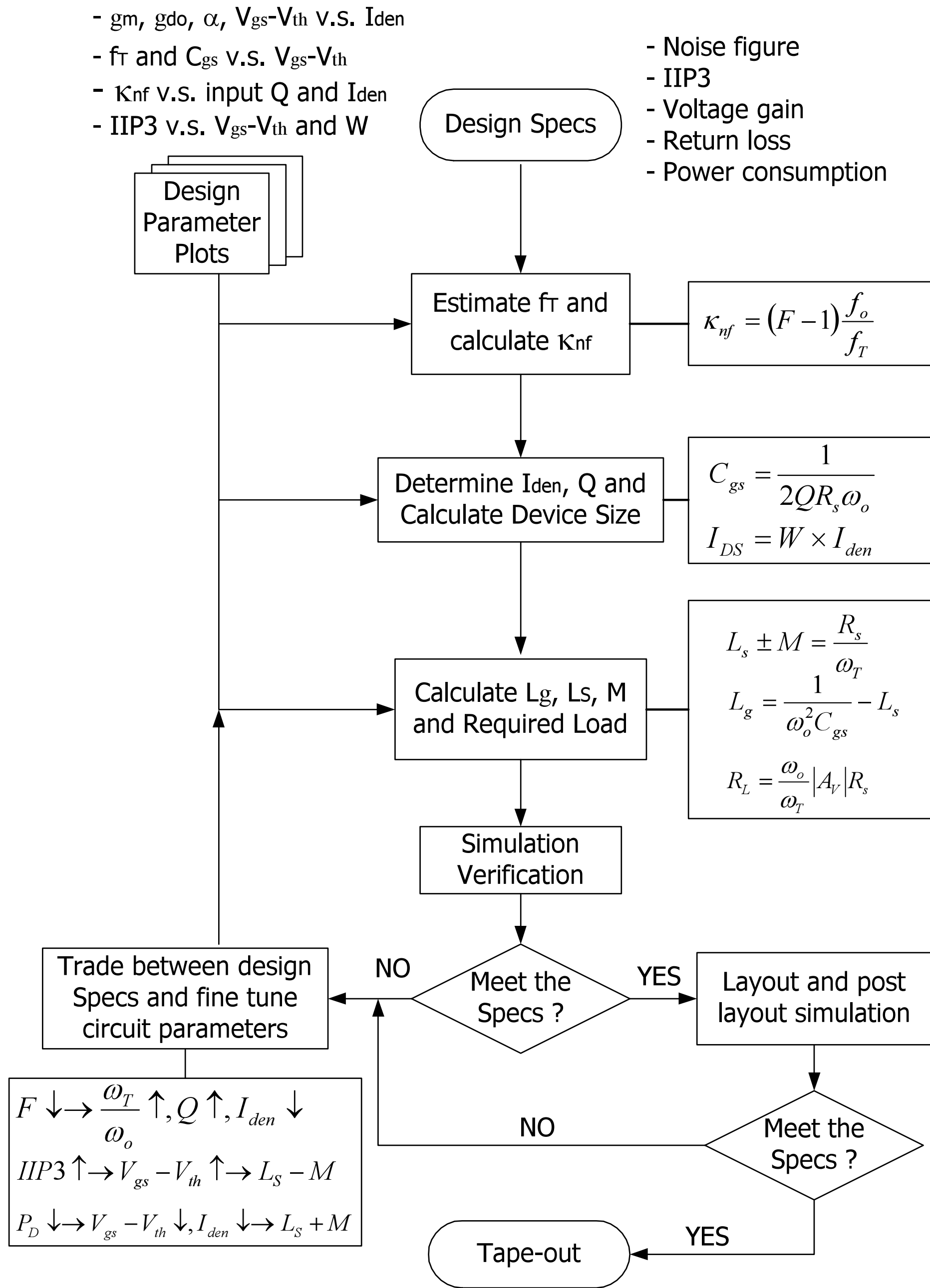



Fig. 99. Design flow of the mutual-coupled LNA

## B. Chip Measurement Results of the Mutual-Coupled Degenerated LNA

The proposed mutual-coupled source-degenerated LNA is implemented using TSMC 0.35$\mu m$ CMOS technology and fabricated through MOSIS service. Its die microphotograph is shown in Fig. 100.

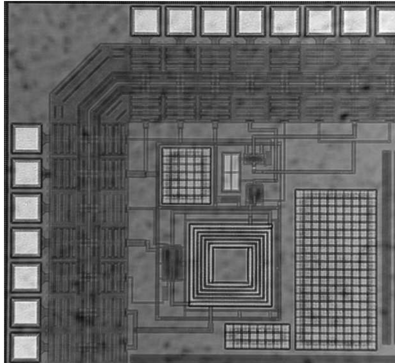


Fig. 100. Die microphotograph of the mutual-coupled degenerated LNA

The mutual coupled inductors $L_M$ and $L_s$ was implemented on-chip by two interleaved square spirals. The inductor $L_G$ was formed by the bond wire and off-chip surface mount inductor. Multiple bond pads were used for ground connections to reduce the ground inductance. The LNA occupies $700\,\mu m \times 500\,\mu m$ active silicon area.

Fig. 101 shows the measured small signal performance of the proposed LNA. The gain (S21) is 17 dB at 960 MHz. The noise figure is 3.4 dB which seems a little higher for GSM application. This is because it includes the output buffer formed by $M_3$, $R_s$ and $L_{D2}$. Bonding pads and inferior quality factors of inductors also makes the NF larger. Simulation shows that the LNA with better inductors and bonding pads removed can have a noise figure about 1.4 dB. The LNA is tested within a plastic package and soldered on a PCB board. The S12 is measured for the whole setup,

so the reverse isolation for the LNA itself should be better than what is shown in Fig. 101. The measured IIP3 plot is illustrated in Fig. 102, which is -5.1 dBm. The LNA draws 5.6 mA from a single 2.3 V power supply.

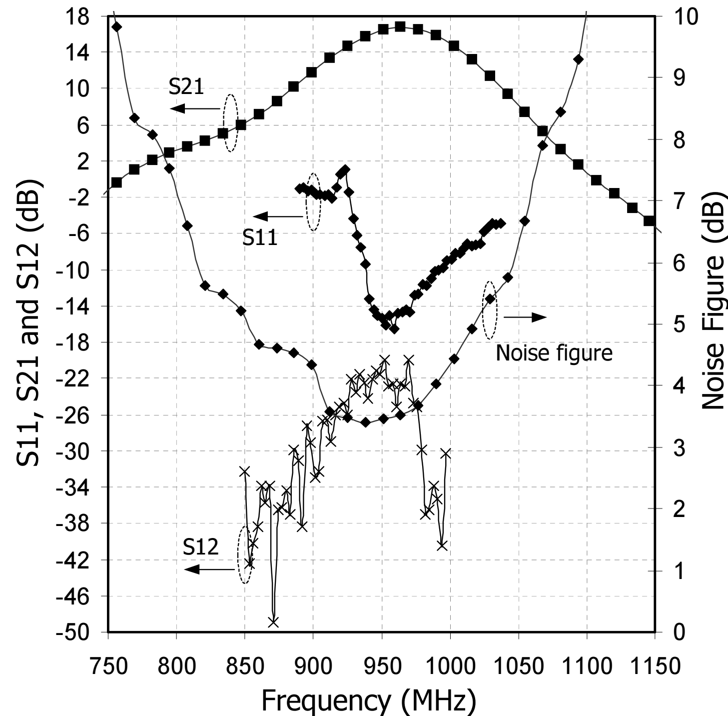



Fig. 101. Measured small signal performance of the mutual-coupled degenerated LNA

Table XVIII summarizes the measurement results together with other reported GSM LNA's performance in the literature.

## C. A Dual-Band Inductive Coupled LNA

Multi-band operation can be implemented by device switching or concurrent impedance match or both. The device switching method can provide separate optimization at different working frequencies [33]. It basically attempts to merge two or more designs into one compact circuit by sharing components. An issue of this method is that the

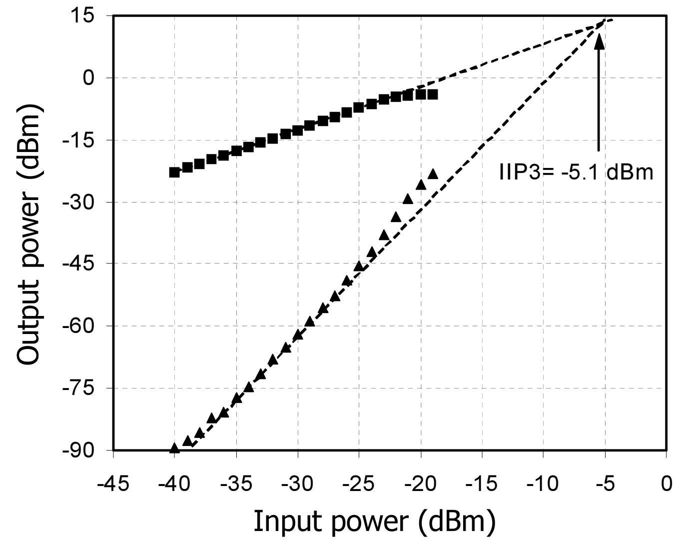



Fig. 102. Measured IIP3 of the mutual-coupled degenerated LNA

Table XVIII. Reported LNA performance for GSM applications

| GSM LNA | Technology | Gain (dB) | NF (dB) | IIP3 (dBm) | S11 (dB) | Power (mW) |
|---|---|---|---|---|---|---|
| [52] | 0.18$\mu m$ CMOS | 22.4 | 0.6 | -5.3 | <-25 | 15.3 |
| [56] | 0.25$\mu m$ CMOS | 15/-5 | 1.9 | -7 | -8 | 25 |
| [29] | 0.25$\mu m$ CMOS | 16.2 | 1.85 | -7.25 | -8 | 27 |
| [57] | 0.35$\mu m$ CMOS | 20 | 1.6 | +2 | -14 | 11.25 |
| This Work | 0.35$\mu m$ CMOS | 17 | 1.4 | -5.1 | -14 | 13 |

shut-off components may affect the active ones due to the parasitics introduced by turned-off devices. The concurrent operation LNA can work at two or more modes at the same time, but usually it is hard to provide optimal working condition for both [51]. Introducing switches at the output of concurrent matched LNA can provide additional degrees of freedom for more frequency bands [52]. Because the switch is at output, it will not affect the input matching condition. But linearity of the switch and parasitics may offset the linearity and output working condition of the circuit.

From (6.4) one can find that without disturbing other factors, using a different $L_G$ value can provide input match around different operation frequency $\omega_o$. A strait forward implementation is to switch in or out more inductance using switches. This approach suffers from the added parasitic capacitance and resistance of the switches and increased noise due to switch resistance. Without using switches, an LC parallel network can be put in series with $L_G$ as shown in Fig. 103. The parallel LC network ($L_B \parallel C_B$) will present inductive or capacitive impedance below or above its intrinsic resonant frequency ($\omega_B = \frac{1}{\sqrt{L_B C_B}}$), thus modify the effective inductance series with the gate. Mutual coupling between $L_G$ and $L_B$ can further modify the impedance and also reduce the required value of inductance thus save area. $L_G^{'}$ represents the bonding wire inductance or additional required inductance for matching.

The equivalent impedance formed by $C_B$, $L_B$ and $L_G$ can be found to be

$$Z_{BG} = j\omega\left(L_G + M_G\right) + j\omega\frac{\left(L_B + M_G\right)\left(1 + \frac{\omega^2}{\omega_M^2}\right)}{1 - \frac{\omega^2}{\omega_B^2}} \tag{6.31}$$

where $\omega_M = \frac{1}{\sqrt{M_G C_B}}$ and $\omega_B = \frac{1}{\sqrt{L_B C_B}}$. The first term in (6.31) shows that mutual inductance $M_G$ enhances the effective inductance of $L_G$, so lower inductance value and size of $L_G$ can be used. The frequency point where the second term of (6.31) changes

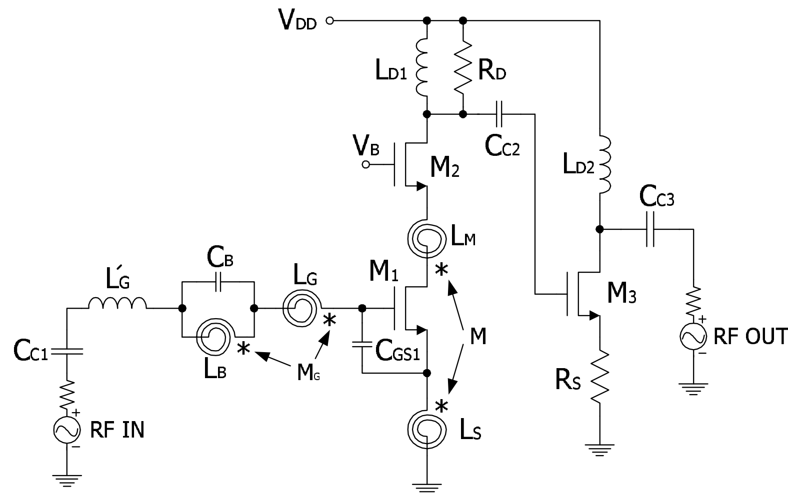



Fig. 103. Dual-band inductive coupled LNA (bias not shown)

from inductive to capacitive does not depends on the mutual coupling. Therefore by changing $C_B$, the impedance property cross-over point can be changed and by adjusting the coupling strength, proper matching at different frequency points can be realized. Fig. 99 can be followed as the design procedure of the dual-band LNA by considering the discussions about input impedance matching.

D. Simulation Results of the Dual-band LNA

The dual-band LNA is design using the TSMC $0.35\mu m$ CMOS technology. Fig. 104(a) is the simulated S11 and S21 plot of the dual-band LNA. The LNA's input is matched to $50\,\Omega$ at 900 MHz GSM band and 1800 MHz DCS-1800 band. The S21 for these two bands are 14 dB and 19 dB respectively.

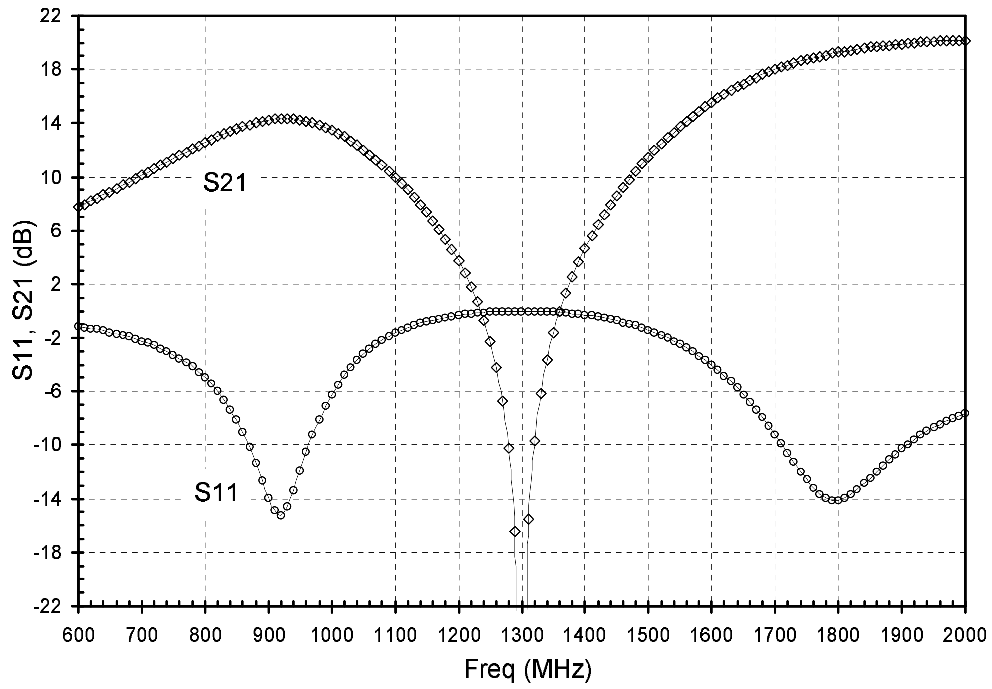
Fig. 104(b) shows the noise performance. The noise figure at the two bands almost equals to the minimum noise figure, which means the input matching network is also optimized for noise matching. The noise figure is 0.8 dB for 900 MHz band and

1.6 dB for 1800 MHz band. Fig. 105 shows the input-referred third-order intercept point (IIP3) plots for both bands. In the 900 MHz GSM band, the IIP3 is -5.5 dBm. For DCS-1800, the IIP3 is -2.5 dBm. A Q value of 10 for the inductors is assumed in the simulation.

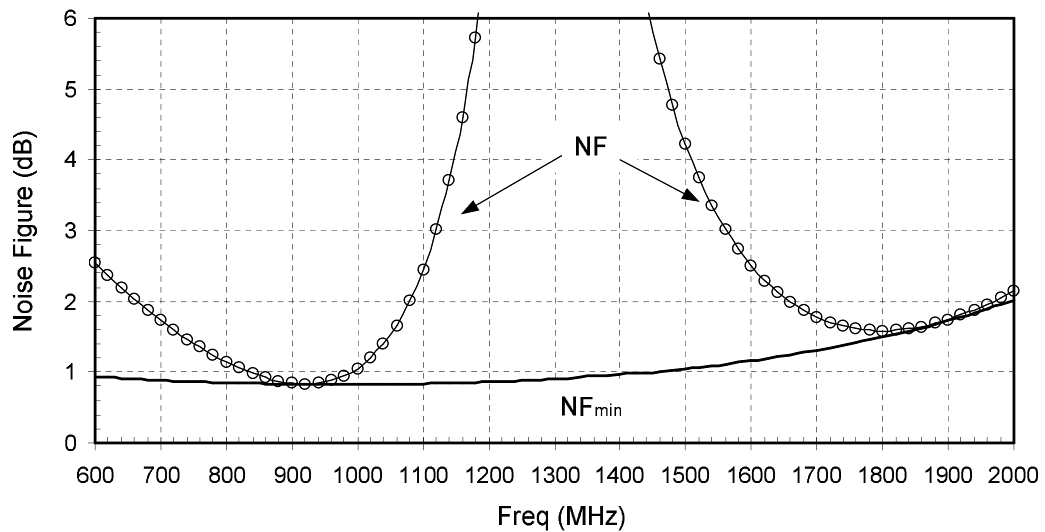
Table XIX summaries the performance of some reported dual/multi-band LNAs and the proposed dual-band LNA. It shows that the inductive-coupled LNA gives comparable results to the literature.

Table XIX. Reported LNA performance in for cellular applications († indicates the value is for the whole front-end)

| Multi-Band LNA | Standard | Technology | Gain (dB) | NF (dB) | IIP3 (dBm) | S11 (dB) | Power (mW) |
|---|---|---|---|---|---|---|---|
| [58] | GSM900 | 0.8 $\mu m$ | 20 | 1.6 | -6.97 | -17 | 15.4 |
| | DCS1800 | BiCMOS | 18 | 1.85 | -3.68 | -21 | |
| [52] | GSM900 | 0.18$\mu m$ | 22.4 | 0.6 | -5.3 | <-25 | 15.3 |
| | GSM1800 | CMOS | 14.5 | 1.0 | -3.8 | | |
| | WCDMA | | 14.1 | 1.4 | -3.1 | | |
| [33] | GSM900 | 0.35$\mu m$ | 39.5† | 2.3† | -19† | <-12 | 11.88 |
| | WCDMA | BiCMOS | 33† | 4.3† | -14.5† | <-18 | 10.98 |
| This Work | GSM900 | 0.35$\mu m$ | 17 | 0.8 | -5.5 | -14 | 13 |
| | DCS1800 | CMOS | 19 | 1.6 | -2.5 | -14 | 13 |

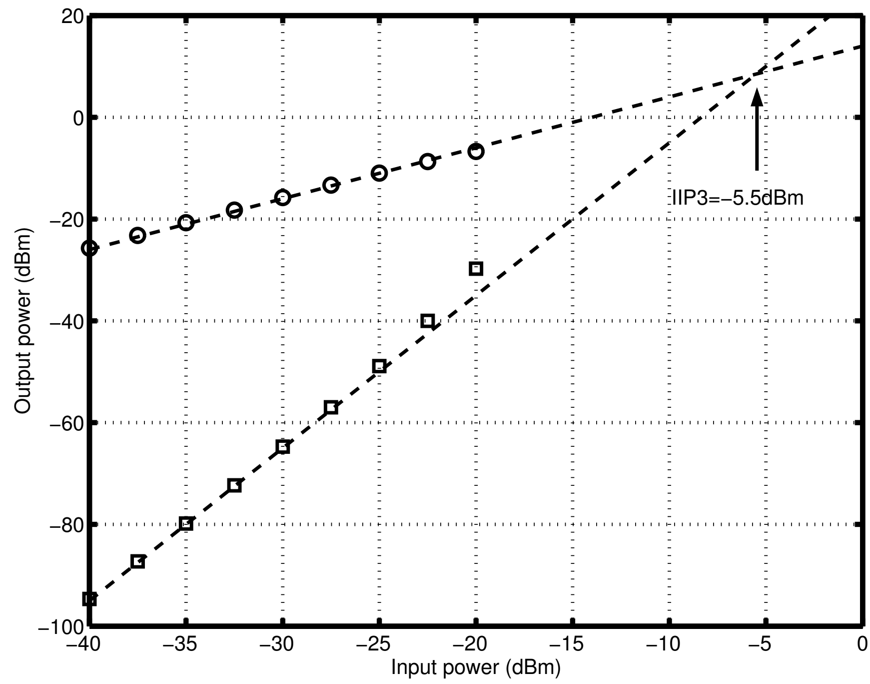



(a)

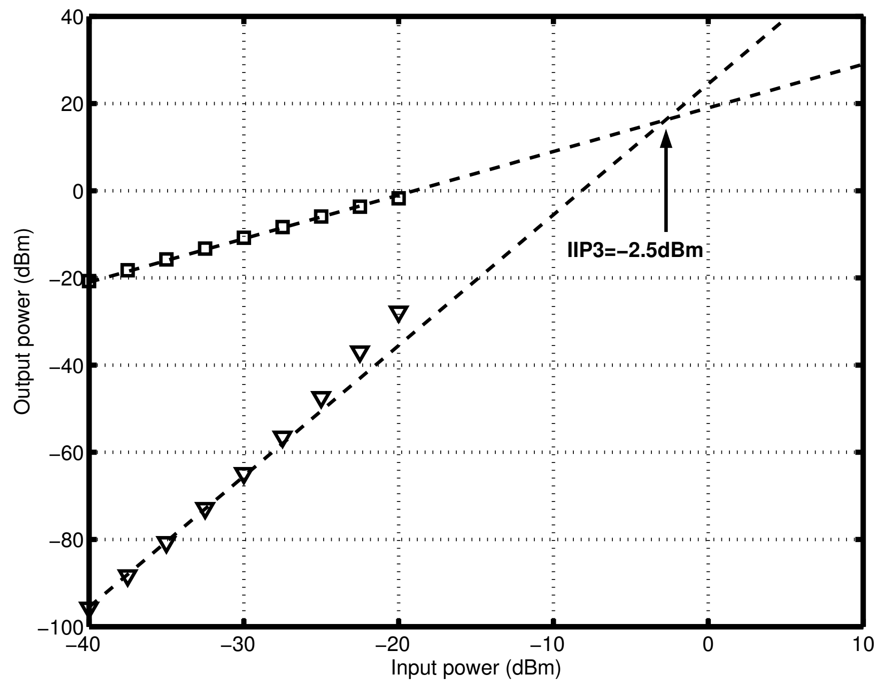



(b)

Fig. 104. Simulated small signal performance of the dual-band LNA (a) S11 and S21 (b) noise figure and minimum noise figure




(a)




(b)

Fig. 105. Simulated IIP3 of the dual-band LNA (a) GSM-900 (b) DCS-1800

# CHAPTER VII

# FRONT-END CIRCUITS FOR WIDE-BAND APPLICATION

Ultra-wide band (UWB) systems are being considered excellent candidates for the future-generation short-range, high-throughput wireless communications. It is emerging as a solution for the IEEE 802.15.3a (TG3a) standard [59] and will complement with Bluetooth and WiFi standards. UWB-based technology is expected to enable personal devices with integrated wireless connectivity. This requires high data rates (110, 200, 480 Mbps) and reasonable low power consumption. Therefore, UWB requires CMOS design in order to achieve low power and low cost integration with other devices, and to fulfill the vision of integrated connectivity [60].

## A. Introduction to Ultra-Wide Band System

Ultra-wide band systems transmit signals that demonstrate extremely low power-spectral-density and occupy very wide bandwidth. The UWB signal bandwidth is greater than 20% of its center frequency and must have a minimum value of 500 MHz as required by US Federal Communication Commission (FCC). The ultra-wide nature of UWB signals in the frequency domain leads to ultra-fine multi-path resolution in the time domain, which enables a path diversity gain by using a RAKE receiver. Therefore UWB signals are immune to multi-path fading. According to Shannon channel capacity theorem, the information a channel can carry [61] (channel capacity, C) is

$$C = B\log_2\left(1+\frac{P}{BN_0}\right) \tag{7.1}$$

where B is the channel bandwidth in hertz (Hz), P is the received signal power in watt (W), $N_0$ is the noise power spectral density in watt per hertz (W/Hz). In order to increase the channel capacity, one can increase the transmitted power or the signal bandwidth. But the channel capacity has a linear increase in bandwidth while a logarithmic increase in signal power. So it is more beneficial to have a wider bandwidth than higher signal power. With increased bandwidth, the transmitted power can be reduced thus reducing the interference to other systems. As an overall consequence, UWB can provide very high data rates at limited range using very low signal power. Due to its extremely low power spectral density, UWB can co-exist with other standards operating in its frequency bands. Fig. 106 shows the UWB spectrum compared with the 802.11a signal spectrum.

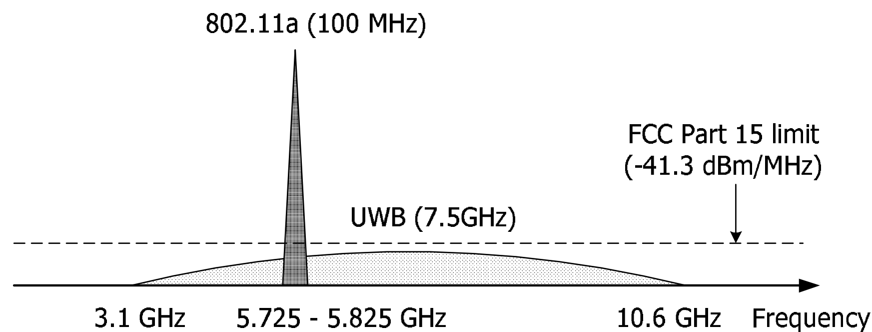



Fig. 106. UWB signal spectrum

The bandwidth resources available for UWB can be used two different ways. Impulse radio was the original approach to UWB realization. It communicates with baseband pulses of very short durations, typically on the order of a nanosecond, thereby spreading the energy of the radio signal very thinly from near DC to a few gigahertz. Data could be modulated using either pulse amplitude modulation (PAM) or pulse-position modulation (PPM). Multiple-access could be supported by utilizing

the time-hopping format [62].

A more recent approach to UWB is a multi-banded system where the UWB frequency band from 3.1 GHz to 10.6 GHz is divided into several smaller bands. Each of these bands has a bandwidth greater than 500MHz to comply with the FCC definition of UWB. Several companies like Intel, Texas Instruments and Time Domain support this approach. The multi-banded approach has a much greater flexibility in coexistence with other wireless systems and is based on more conventional technologies. The official UWB standard is still under development by several companies that have already provided their proposals to the 802.15.3a task group. As an example, Texas Instruments proposed a multi-band OFDM system. It divides the UWB spectrum into several 528 MHz bands. Information is transmitted using OFDM modulation on each band. The OFDM carries are generated using an 128-point IFFT/FFT.

Due to the wide bandwidth of UWB signal, direct conversion may be the best receiver architecture. If an IF is to be used, it should be at least higher than 250 MHz. This high IF will increase the complexity, power consumption and cost of the baseband circuit. The wide band nature of UWB signal makes the DC offset and low frequency noise less problematic than that in a narrow band system. Fig. 107 depicts a conceptual UWB receiver block diagram.

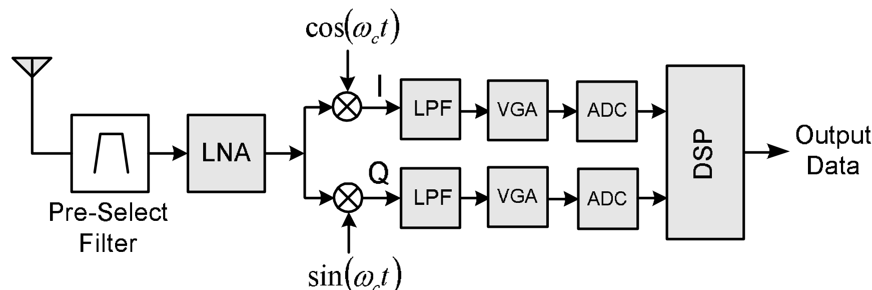



Fig. 107. Direct conversion UWB receiver

The wide spread spectrum of UWB signal makes its dynamic range (10-12 dB) much smaller than that of its narrow band counterpart (50 dB for Bluetooth and 66 dB for IEEE 802.11b). It may be suitable to employ the so called software-defined radio architecture where the ADC is put directly after the LNA. But for current technology, a high sampling rate ADC is still a big challenge, so the frequency conversion scheme will still be needed in the near future.

## B. Distributed RF Front-End Circuits

In conventional circuit design, it is well known that there exists a fundamental limitation of the gain-bandwidth product. The gain-bandwidth product is proportional to $\frac{g_m}{C}$, and is intrinsic to the devices used. In other words, gain will trade with bandwidth. Conventional circuit design treats the circuit elements as lumped-elements, i.e. each element only accounts for one dominant physical or electrical effect. Resistor represents the heat dissipation, inductor represents energy storage of magnetic field and capacitor represents energy storage of electric field. But in the real world, especially at high frequency, when the dimensions of a circuit is comparable to the wavelength, no single lumped-elements can be identified. For example, a 10 mm long lossy metal line at 30 GHz has all of the effects of heat dissipation, magnetic and electric energy storage. No single one of the effects overwhelms others. Thus the circuit will be treated as distributed. Distributed circuits have the potential of large bandwidth and the gain will not trade with bandwidth but rather with time delay.

### 1. Transmission Line Properties and Characterization

Transmission line (T-line) is the key element of distributed circuits. Fig. 108 is the equivalent circuit model of an infinitesimal small segment of a transmission line. The

series inductance is due to the magnetic field effects and the shunt capacitance is due to electric field coupling between the signal and ground lines. The loss in the transmission media is modeled by series and shunt resistor. The R, L and C constants are defined as per unit length circuit parameters. Note that this model only applies to TEM mode transmission lines exactly [2].

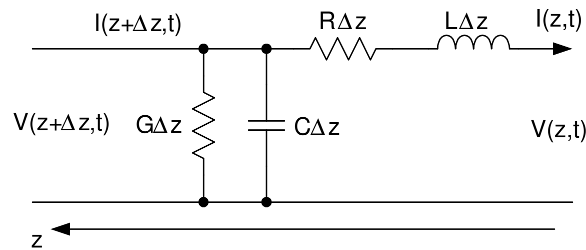



Fig. 108. Distributed model of transmission line

Assuming steady state operation, $V(z,t) = V(z)\,e^{j\omega t}$, $I(z,t) = I(z)\,e^{j\omega t}$. $V(z)$ and $I(z)$ are voltage phasor and current phasor respectively. It can be shown that the voltage and current along the line fulfill the following equations

$$\frac{dV(z)}{dz} = (R + j\omega L)\,I(z) \tag{7.2}$$

$$\frac{dI(z)}{dz} = (G + j\omega C)\,V(z) \tag{7.3}$$

or

$$\frac{d^2V(z)}{dz} = \gamma^2 V(z) \tag{7.4}$$

$$\frac{d^2I(z)}{dz} = \gamma^2 I(z) \tag{7.5}$$

where $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta$. $\gamma$ is known as propagation constant, $\alpha$ is known as attenuation constant, $\beta$ is usually called the wave number, and $\beta = \frac{2\pi}{\lambda}$.

The general solutions for (7.4) and (7.5) are

$$V = V_0^+ e^{\gamma z} + V_0^- e^{-\gamma z} \tag{7.6}$$

$$I = I_0^+ e^{\gamma z} + I_0^- e^{-\gamma z} \tag{7.7}$$

The current phasor can also be written as

$$I = \frac{V_0^+}{Z_c} e^{\gamma z} - \frac{V_0^-}{Z_c} e^{-\gamma z} \tag{7.8}$$

where $Z_c = \sqrt{\frac{R+j\omega L}{G+j\omega C}}$ is the characteristic impedance of the transmission line.

For lossless transmission line, $R = 0$ and $G = 0$, the characteristic impedance becomes

$$Z_c = \sqrt{\frac{L}{C}} \tag{7.9}$$

and the propagation constant is $\gamma = j\omega\sqrt{LC}$, so $\alpha = 0$ and

$$\beta = \omega\sqrt{LC} \tag{7.10}$$

The phase velocity of lossless transmission line is

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{7.11}$$

For low loss transmission line, it is generally assumed that $R \ll \omega L$ and $G \ll \omega C$, so $Z_c$ is approximately to be

$$Z_c \cong \sqrt{\frac{L}{C}} \tag{7.12}$$

which is the same as the characteristic impedance of a lossless line. The propagation constant can be written as

$$\gamma = \sqrt{RG - \omega^2 LC + j\omega\left(RC + LG\right)}$$

Due to the low loss assumption, the $RG$ term can be dropped off from the above equa-

tion. Then applying the binomial expansion, $(a-b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \cdots$ to the above equation it reads

$$\gamma \cong j\omega\sqrt{LC} + \frac{1}{2}\sqrt{LC}\left(\frac{R}{L} + \frac{G}{C}\right) + \frac{j}{8}\omega\sqrt{LC}\left(\frac{R}{\omega L} + \frac{G}{\omega C}\right)^2 + \cdots \tag{7.13}$$

Based on low loss assumption, only keep the first two terms of (7.13),

$$\gamma \cong \alpha + j\beta = \frac{1}{2}\sqrt{LC}\left(\frac{R}{L} + \frac{G}{C}\right) + j\omega\sqrt{LC} \tag{7.14}$$

The phase constant $\beta$ is the same as the wave number in the lossless case. But the attenuation constant is not zero any more:

$$\alpha = \frac{R}{2Z_c} + \frac{GZ_c}{2} \tag{7.15}$$

The phase velocity of the signal in the low loss line is

$$v_p = \frac{j\omega}{\gamma} \cong \frac{j\omega}{j\omega\sqrt{LC} + \frac{1}{2}\sqrt{LC}\left(\frac{R}{L} + \frac{G}{C}\right)} = \frac{1}{\sqrt{LC} + \frac{1}{2j}\left(\frac{R}{\omega L} + \frac{G}{\omega C}\right)\sqrt{LC}} \cong \frac{1}{\sqrt{LC}} \tag{7.16}$$

which is the same as the lossless line.

Now the formula of the loss per unit length for low loss line will be derived. For a matched line, there is no reflected signal, the voltage along the line at point $z + \Delta z$ can be expressed as

$$V(z + \Delta z) = V_0^+ e^{\gamma(z+\Delta z)}$$

When the voltage propagates to point $z$, it is

$$V(z) = V_0^+ e^{\gamma z}$$

So the attenuation is

$$\frac{P(z + \Delta z)}{P(z)} = \left|\frac{V(z + \Delta z)}{V(z)}\right|^2 = e^{2\alpha\Delta z}$$

Expressed in dB form, it is

$$\frac{\Delta P_{\text{dB}}}{\Delta z} = 20\alpha \log_{10} e = \frac{10}{\ln(10)} \left( \frac{R}{Z_c} + G Z_c \right) \tag{7.17}$$

Table XX summarizes the basic properties of lossless and low-loss transmission lines.

Table XX. Basic properties of lossless and low loss T-line

| | Lossless T-line | Low-loss T-line |
|---|---|---|
| $Z_c$ | $\sqrt{\frac{L}{C}}$ | $\sqrt{\frac{L}{C}}$ |
| $\alpha$ | 0 | $\frac{R}{2Z_c} + \frac{GZ_c}{2}$ |
| $\beta$ | $\omega\sqrt{LC}$ | $\omega\sqrt{LC}$ |
| $v_p$ | $\frac{1}{\sqrt{LC}}$ | $\frac{1}{\sqrt{LC}}$ |

For a certain length of transmission line, it is characterized by its characteristic impedance $Z_c$ and propagation constant $\gamma$. A S-parameter-base method can be used to measure these values [63]. The S-parameter matrix for a lossy unmatched transmission line with characteristic impedance $Z_c$ and propagation constant $\gamma$ in a $Z_o$ impedance system is [63]

$$[S] = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \frac{1}{D_s} \begin{bmatrix} (Z_c^2 - Z_o^2)\sinh \gamma l & 2Z_c Z_o \\ 2Z_c Z_o & (Z_c^2 - Z_o^2)\sinh \gamma l \end{bmatrix} \tag{7.18}$$

where $l$ is the physical length of the line, $D_s = 2Z_c Z_o \cosh \gamma l + (Z_c^2 + Z_o^2)\sinh \gamma l$. It can be shown that

$$Z_c = Z_o \sqrt{\frac{(1 + s_{11})^2 - s_{21}^2}{(1 - s_{11})^2 - s_{21}^2}} \tag{7.19}$$

$$e^{-\gamma l} = \left( \frac{1 - s_{11}^2 + s_{21}^2}{2s_{21}} \pm K \right)^{-1} \tag{7.20}$$

where $K = \frac{1}{2s_{21}}\sqrt{\left(s_{11}^2 - s_{21}^2 + 1\right)^2 - 4s_{11}^2}$.

After obtaining $\gamma$ and $Z_c$, the distributed parameters R, G, L and C can be calculated by

$$R = \Re\left\{\gamma Z_c\right\} \tag{7.21}$$

$$G = \Re\left\{\frac{\gamma}{Z_c}\right\} \tag{7.22}$$

$$L = \Im\left\{\gamma Z_c\right\}\omega^{-1} \tag{7.23}$$

$$C = \Im\left\{\frac{\gamma}{Z_c}\right\}\omega^{-1} \tag{7.24}$$

## 2. Transmission Lines on Silicon Substrate

In distributed circuits, interconnection lines become an integrated part of the circuits. The key to successful design of distributed circuits is the modeling of transmission line used in the circuits. One can have two ways to arrive at the desired circuit. One is have a well modeled transmission line, then design the circuit based on that line. The other way is to build a general macro model of transmission lines, design the circuit using that model then implement the line to map the macro model. Both methods require a good modeling. There are two types of transmission lines which are suitable to be integrated on silicon, they are coplanar stripeline and micro-strip line. Among them, coplanar stripline seems more suitable for the purpose. The following sessions investigate these two different lines.

The structure of the silicon and coplanar stripline is demonstrated in Fig. 109 and Fig. 110. They are the structure dimensional drawing and the rendered 3D drawing in HFSS respectively. The substrate structure of the TSMC 0.18 $\mu m$ CMOS process is very complicated, the structure shown here is a simplified version. Some of the layers are lumped together and their parameters are averaged over their thickness. The spacing and width of the lines (M6) are chosen to be 5 $\mu m$ [64]. The total length

of the line is 500 $\mu m$. The dielectric thickness between the bottom of metal 6 and substrate under field oxide (FOX) is 8.15 $\mu m$. Using the line width comparable to the thickness of the dielectric, the electric-field (E-field) will only penetrate the surface of the lossy silicon substrate and at the same time large metal surface area is used to conducting current. So loss due to lossy silicon substrate and metal skin effect is reduced [65]. The skin depth of a transmission line can be calculated by

$$\delta_s = \sqrt{\frac{2}{\omega\mu\sigma}} \tag{7.25}$$

For metal 1 and metal 6 the skin depth is about 0.6 $\mu m$ at 30 GHz.

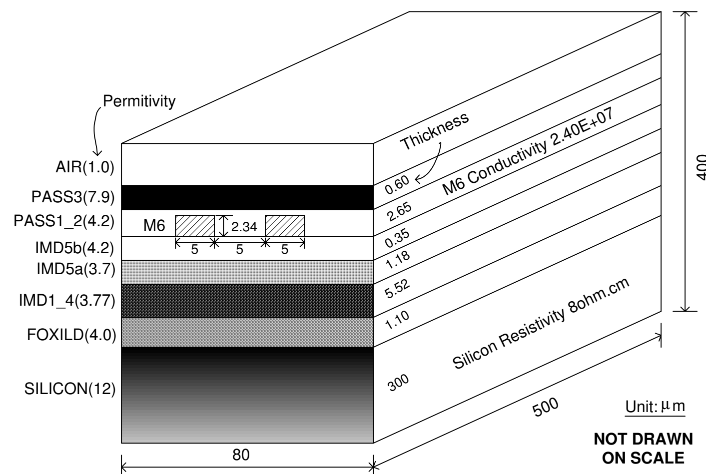



Fig. 109. Coplanar stripline formed by metal 6 (dimensional drawing)

There exist two modes of operation for the coplanar stripline, even-mode and odd-mode. Fig. 111 depicts the field distribution of these two operation modes. The desired propagation mode is the odd-mode. It has better field confinement than that of even mode, therefore less loss. Simulation shows that at 30 GHz the loss of the line is within 0.8 dB per millimeter (dB/mm).

The micro-stripline structure is depicted in Fig. 112. The ground plane is formed

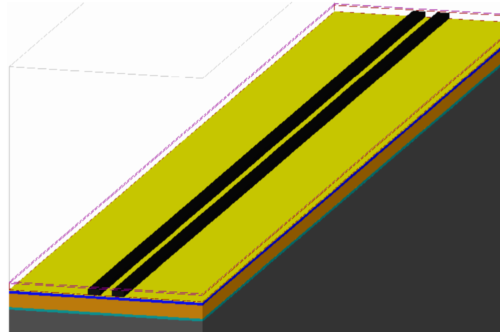
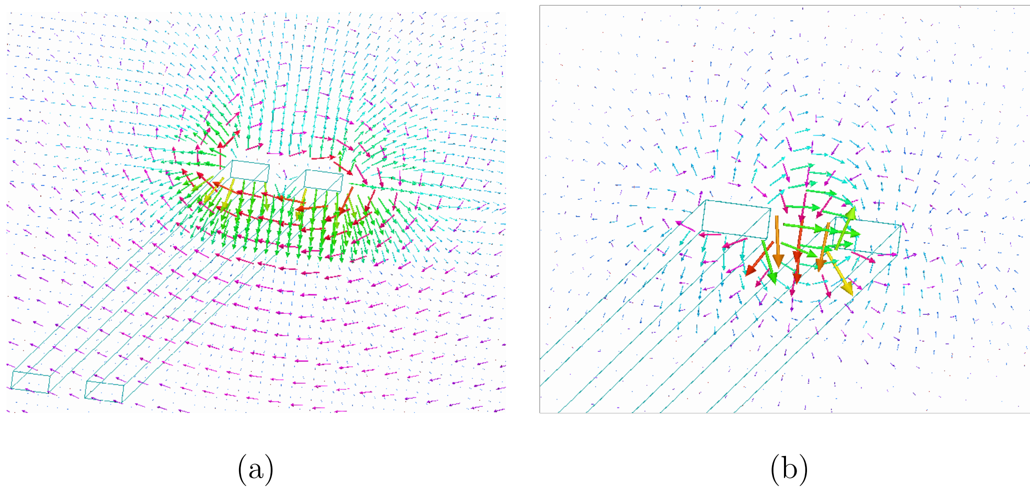


Fig. 110. Coplanar stripline rendered by HFSS



(a) (b)

Fig. 111. Coplanar stripline operation mode (a) Even mode (b) Odd mode

by metal 1. The signal line is formed by metal 6, which is the highest level and thickest metal available in this technology. There is only one operation mode for the micro-stripline. Fig. 113 shows the field distribution. For line width from 5 $\mu m$ to 25 $\mu m$, the loss is approximately 0.44 dB/mm at 30 GHz. The smaller loss is due to better field confinement than the coplanar stripline.

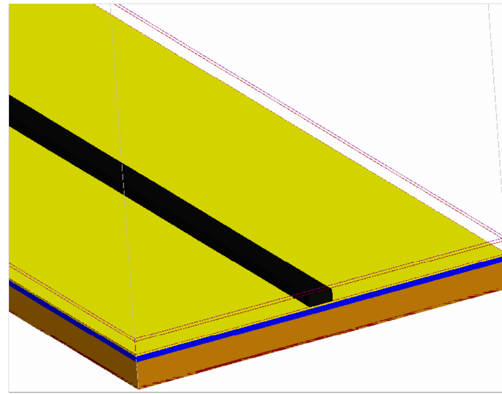


Fig. 112. Micro-stripline HFSS render

Now comes the question: Which type of transmission line is more suitable for silicon integration? Note that the transmission line will be loaded by transistors, thus lowering its effective characteristic impedance. If the loaded input and/or output line is required to be 50 Ω, then the line itself should be greater than 50 Ω. Fig. 114 shows the characteristic impedance of different dimensions of coplanar stripline and micro-stripline. It shows that for reasonable line width and spacing, the impedance of a coplanar stripline can be controlled from around 66 Ω to 124 Ω. On the other hand, the line width of micro-stripline is the only changeable variable by the designer. It is observed that the wider the line, the smaller the impedance. In order to keep the impedance greater than 50 Ω before loading, the line width should be smaller than 7 $\mu m$. For a larger impedance value, the line width will be even smaller. This will

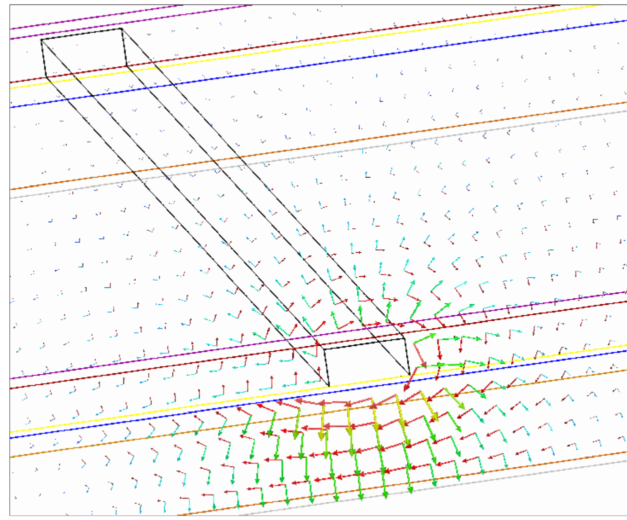
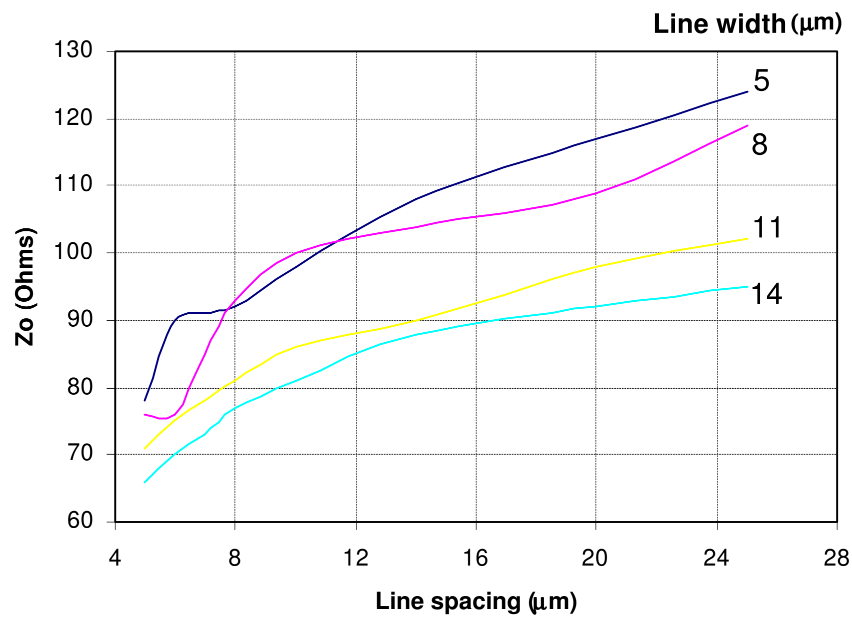


Fig. 113. Field distribution of micro-stripline

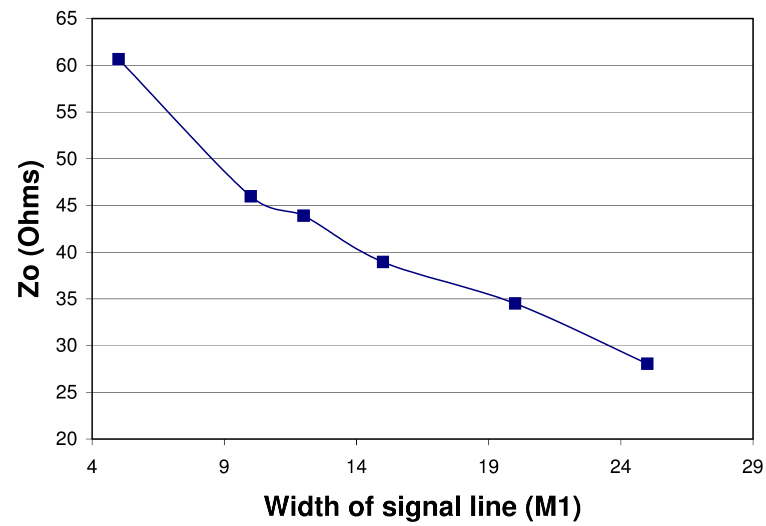
increase the line loss due to its high resistance. So the coplanar stripline is preferred for distributed circuit design on silicon.

## 3. Distributed Amplifier as LNA

Distributed circuits can be realized by using either artificial transmission lines or planar transmission lines. Artificial line uses lumped inductors and capacitors to emulate the behavior of a real transmission lines. It generally requires high quality inductors and capacitors which are usually not readily available in regular silicon process, especially for inductors. Artificial line also has intrinsic cut-off characteristics which will limit the operation bandwidth. Planar transmission lines discussed in the previous section may be a better choice for large bandwidth and high operation frequency. The distributed LNA is implemented by periodically loading planar transmission lines with active devices usually MOSFET transistors. Fig. 115 is the schematic representation of a distributed LNA. Two transmission lines, the gate line




(a)




(b)

Fig. 114. $Z_c$ of (a) coplanar stripline and (b) micro-stripline at 30 GHz

and drain line are coupled unilaterally by MOS transistors. The portion between the dashed lines is a unit section or cell. The amplifier shown in this figure has four sections. Fig. 116 is the equivalent circuit for the unit section. For simplicity, the T-lines are assumed to be lossless, the MOS transistor is modeled by its gate-source capacitance, drain-source capacitance and a transconductance.

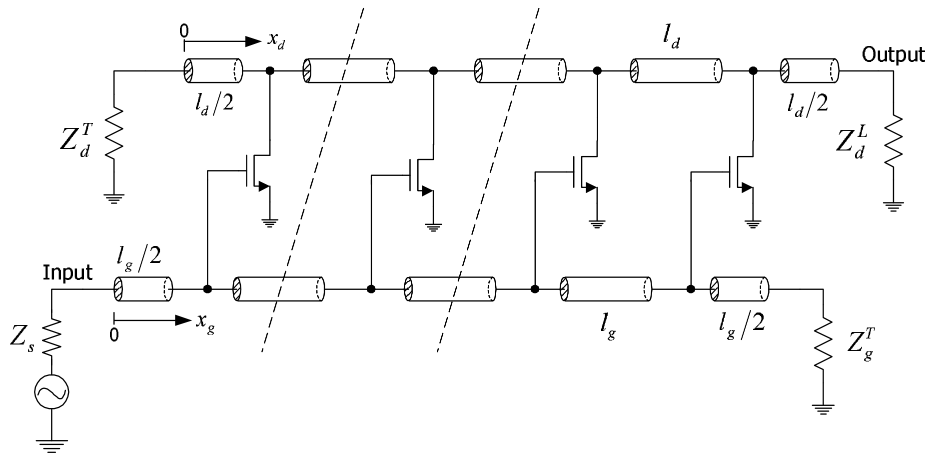



Fig. 115. Schematic representation of a distributed LNA

Due to the periodic loading nature of the T-lines, $C_{gs}$ and $C_{ds}$ can be treated as uniformly distributed along the length ($l_g$ for gate line, $l_d$ for drain line) of the unit cell. If the intrinsic gate line and drain have $L_g$, $C_g$ and $L_d$, $C_d$ as their distributed parameters, the propagation constants and characteristic impedances of the two loaded lines can be obtained as

$$\gamma_g = j\beta_g \approx j\omega\sqrt{L_g\left(C_g + \frac{C_{gs}}{l_g}\right)} \tag{7.26}$$

$$Z_c^g \approx \sqrt{\frac{L_g}{C_g + C_{gs}/l_g}} \tag{7.27}$$

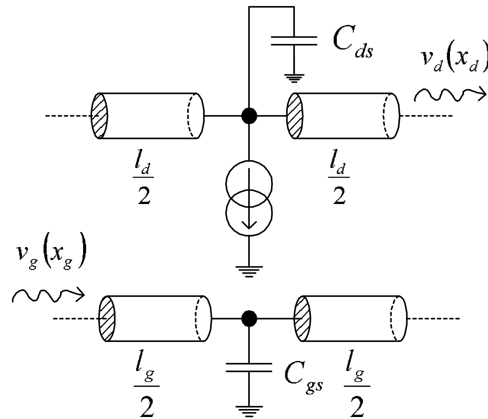



Fig. 116. Unit section of a distributed LNA

for the gate line and

$$\gamma_d = j\beta_d \approx j\omega\sqrt{L_d\left(C_d + \frac{C_{ds}}{l_d}\right)} \tag{7.28}$$

$$Z_c^d \approx \sqrt{\frac{L_d}{C_d + C_{ds}/l_d}} \tag{7.29}$$

for the drain line.

The current generated by the transconductance is also deemed as spread over the length of the drain line. Under matched conditions both lines are terminated by their characteristic impedances:

$$Z_s = Z_c^g = Z_g^T \tag{7.30}$$

$$Z_d^T = Z_c^d = Z_d^L \tag{7.31}$$

and assume the phase synchronization condition

$$\beta_d l_d = \beta_g l_g = \theta \tag{7.32}$$

The voltage gain of the amplifier [66] can be shown to be

$$A_v = -\frac{N g_m Z_c^d}{2} e^{-jN\theta} \tag{7.33}$$

where N is the number of unit sections. It is observed that within the valid frequency range of uniform loading assumptions made in (7.26) through (7.29), the amplifier has a gain which is increasing linearly with the total number of sections and a constant group delay which is also a linear function of N. So increase gain by adding more sections will increase the time delay between input and output. The power gain can be written

$$G = \frac{N^2 g_m^2 Z_c^d Z_c^g}{4} \tag{7.34}$$

In practice, losses exist both in the transmission lines and the active loading MOS transistors. The gate line is then loaded by $C_{gs}$ and $R_{gs}$ in series, and the drain line is loaded by $C_{ds}$and $R_{ds}$ in parallel. In silicon implementation, losses come from the T-lines may also need to be considered. Therefore the propagation constants of loaded gate line and drain line will have a real part $\alpha_g$ and $\alpha_d$ respectively. Imaginary components will appear in the loaded characteristic impedances but their contribution is usually small within the useful frequency range. Under the phase synchronization condition of (7.32), the power gain [66] of the amplifier is

$$G \approx \frac{g_m^2 Z_c^d Z_c^g}{4} \left| \frac{e^{-N\alpha_g l_g} - e^{-N\alpha_d l_d}}{\alpha_g l_g - \alpha_d l_d} \right|^2 \tag{7.35}$$

Thus the gain does not increase monotonically with N and at a particular frequency, the optimal value of N is given by

$$N_{opt,\,LNA} = \frac{\ln(\alpha_d l_d) - \ln(\alpha_g l_g)}{\alpha_d l_d - \alpha_g l_g} \tag{7.36}$$

In a practical amplifier, the major losses is due to the gate loading, an upper

bound for the total number of sections can be found as [66]

$$N \leq \frac{2}{R_{gs}\omega^2 C_{gs} Z_c^d} \tag{7.37}$$

A five-section distributed LNA is implemented using TSMC 0.18 $\mu m$ CMOS process. The T-lines are formed by coplanar striplines. This circuit draws 45 mA from 1.2 V power supply. Fig. 117 is the layout view of the distributed LNA. Fig. 118 is the S-parameters plot. The flat gain bandwidth is from DC to 18 GHz with 2 dB ripple. From DC to 14 GHz, the input and output return losses are better than 10 dB.

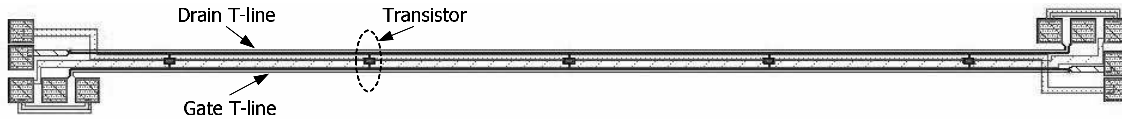



Fig. 117. Layout of the five-section distributed LNA

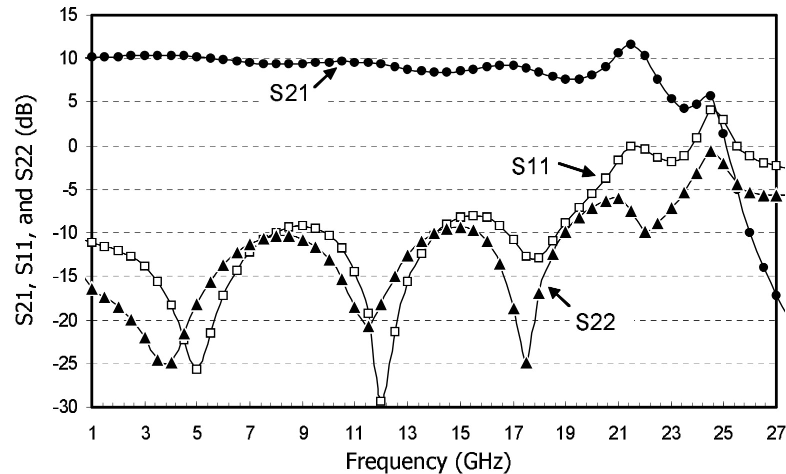



Fig. 118. S-parameters of the five-section distributed LNA

Fig. 119 is the noise figure plot of this 5-section distributed LNA. From 8 GHz to

20 GHz, the noise figure is better than 6 dB. A detailed noise analysis will be carried out in the following text to show the noise performance. The loss due to T-lines loading and T-line loss itself will be ignored for simplicity. Before the noise figure can be calculated, one must obtain the noise contribution due to MOS transistors and termination resistors ( $Z_s^T$ and $Z_d^T$ ) as well as the noise power available from the source generator. The noise of load impedance $Z_d^L$ belongs to the next stage, so it is not included in the noise figure calculation.

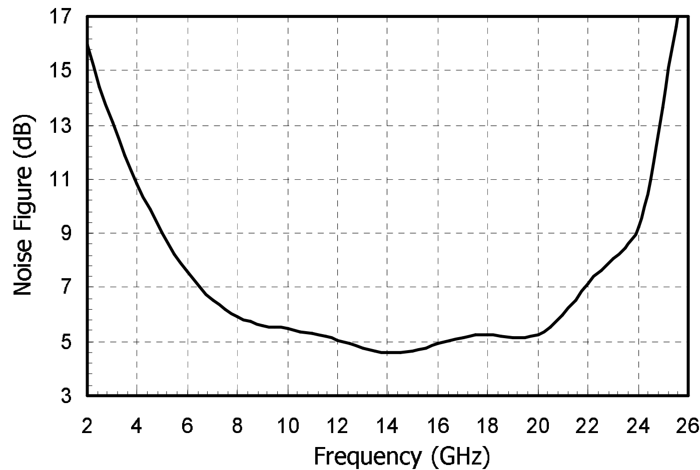



Fig. 119. Noise figure of the five-section distributed LNA

The noise contributed from the transistors in different unit sections is assumed to be uncorrelated with each other. Thus one can first calculate the noise power delivered into the load due to the transistor in one unit section, then sum the noise power over the whole N sections to obtain the total noise power generated by transistors in the distributed circuit. For this purpose, Fig. 120 shows the noise model of a unit section of the distributed LNA.

The major noise sources are the MOS transistor's drain noise current $i_{nd}$ and

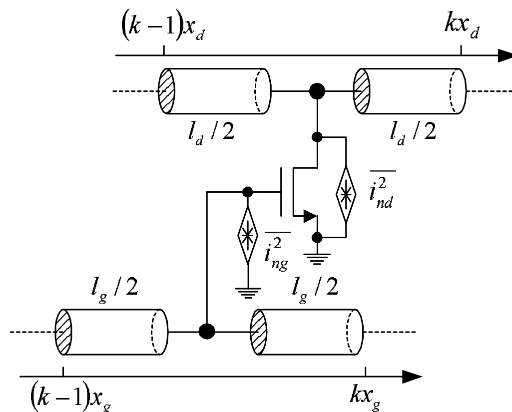



Fig. 120. Noise model of a distributed LNA unit section

induced gate noise current $i_{ng}$. Their mean-square values are

$$\overline{i_{nd}^2} = 4kT\frac{\gamma}{\alpha}g_m\Delta f \tag{7.38}$$

and

$$\overline{i_{ng}^2} = 4kT\delta\alpha\frac{\omega^2 C_{gs}^2}{5g_m}\Delta f \tag{7.39}$$

respectively. The drain noise sources will be treated as uniformly distributed over the unit section. That is to say the drain T-line is periodically driven by noise current $\frac{i_{nd}}{l_d}$ . The noise voltage developed across the load $Z_d^L$ due to the drain noise source of the k-th section can be found by the following integral:

$$\begin{aligned} v_{nd,k} &= \int_{(k-1)l_d}^{kl_d} \frac{i_{nd}}{l_d}\frac{Z_c^d}{2}e^{-j\beta_d(Nl_d - x_d)}dx_d \\ &= \frac{1}{2}i_{nd}Z_c^d\frac{\sin\frac{\theta}{2}}{\frac{\theta}{2}}e^{-j\theta\left(N-k+\frac{1}{2}\right)} \end{aligned} \tag{7.40}$$

When the gate noise current is injected into the k-th section gate T-line at position $(k-1)\,l_g$, the voltage generated across the load $Z_d^L$ of the drain T-line due

to the forward traveling signal is

$$v_{ng,\,k}^{+}=\frac{1}{4}g_m Z_c^g Z_c^d i_{ng}\left(N-k+\frac{1}{2}\right)e^{-j\left(N-k+\frac{1}{2}\right)\theta} \tag{7.41}$$

and due to the reverse propagation is

$$v_{ng,\,k}^{-}=\frac{1}{4}g_m Z_c^g Z_c^d i_{ng}\frac{\sin\left(k-\frac{1}{2}\right)\theta}{\theta}e^{-j\left(k-\frac{1}{2}\right)\theta}e^{-j\left(N-k+\frac{1}{2}\right)\theta} \tag{7.42}$$

The noise contributed from the k-th section gate noise is the sum of the forward and reverse voltage:

$$\begin{aligned} v_{ng,\,k} &= v_{ng,\,k}^{+}+v_{ng,\,k}^{-} \\ &= \tfrac{1}{4}g_m Z_c^g Z_c^d i_{ng}\left[\left(N-k+\tfrac{1}{2}\right)+\tfrac{\sin\left(k-\frac{1}{2}\right)\theta}{\theta}e^{-j\left(k-\frac{1}{2}\right)\theta}\right]e^{-j\left(N-k+\frac{1}{2}\right)\theta} \end{aligned} \tag{7.43}$$

Adding (7.40) and (7.43) together, the output noise voltage due to transistors in the k-th section will be obtained as

$$v_{no,\,k}=\left[\frac{\sin\frac{\theta}{2}}{\theta}i_{nd}+\frac{1}{4}M(k)g_m Z_c^g i_{ng}\right]Z_c^d e^{-j\theta\left(N-k+\frac{1}{2}\right)} \tag{7.44}$$

where

$$M\left(k\right)=\left(N-k+\frac{1}{2}\right)+\frac{\sin\left(k-\frac{1}{2}\right)\theta}{\theta}e^{-j\left(k-\frac{1}{2}\right)\theta} \tag{7.45}$$

It is known that $i_{ng}$ and $i_{nd}$ are correlated with correlation coefficient $c$ which is purely imaginary and equals $j0.395$ for long channel devices. $i_{ng}$ can be decomposed into a component $i_{ngc}$ which is fully correlated with $i_{nd}$ and another component $i_{ngu}$ which is uncorrelated with $i_{nd}$ :

$$i_{ng}=i_{ngc}+i_{ngu} \tag{7.46}$$

where

$$\overline{i_{ngu}^2}=4kTG_u\Delta f \tag{7.47}$$

$$i_{ngc} = F_c i_{nd} \tag{7.48}$$

and

$$G_u = \delta\alpha \frac{\omega^2 C_{gs}^2}{5g_m}\left(1 - |c|^2\right) \tag{7.49}$$

$$\begin{aligned} F_c &\equiv \frac{i_{ngc}}{i_{nd}} = \frac{\overline{i_{ngc} i_{nd}^*}}{\overline{i_{nd} i_{nd}^*}} = \frac{\overline{i_{ng} i_{nd}^*}}{\sqrt{\overline{i_{nd}^2}}\sqrt{\overline{i_{ng}^2}}} \frac{\sqrt{\overline{i_{ng}^2}}}{\sqrt{\overline{i_{nd}^2}}} = c\frac{\sqrt{\overline{i_{ng}^2}}}{\sqrt{\overline{i_{nd}^2}}} \\ &= j\,|c|\,\alpha\sqrt{\frac{\delta}{5\gamma}}\frac{\omega C_{gs}}{g_m} \end{aligned} \tag{7.50}$$

Using the above notations, the noise power delivered to the load by transistors in the distributed circuit is

$$\begin{aligned} P_{no} &= 4kTg_m Z_c^d \tfrac{\gamma}{\alpha}\Delta f \textstyle\sum_{k=1}^{N} \left| \frac{\sin\frac{\theta}{2}}{\theta} + \frac{1}{4} g_m Z_c^g F_c M(k) \right|^2 \\ &+ \tfrac{1}{4} kT g_m^2 G_u \left(Z_c^g\right)^2 Z_c^d \Delta f \textstyle\sum_{k=1}^{N} |M(k)|^2 \end{aligned} \tag{7.51}$$

The output noise power contributed by the drain T-line termination impedance $Z_d^T$ is

$$P_{ndT} = kT\Delta f \tag{7.52}$$

The gate T-line termination impedance $Z_g^T$ contributes to the output noise power in the form of

$$P_{nsT} = kTG \left| \frac{\sin N\theta}{N\theta} \right|^2 \Delta f$$

The noise power available from the signal source is

$$P_{ns} = kT\Delta f \tag{7.53}$$

The noise figure can then be calculated from

$$F = 1 + \frac{P_{no} + P_{ndT} + P_{nsT}}{P_{ns} G} \tag{7.54}$$

where G is the power gain as expressed in (7.34). Written explicitly

$$\begin{aligned} F &= 1+\left|\tfrac{\sin N\theta}{N\theta}\right|^2+\tfrac{4}{N^2 g_m^2 Z_c^d Z_c^g} \\ &+ \tfrac{16}{N^2 g_m Z_c^g}\tfrac{\gamma}{\alpha}\textstyle\sum_{k=1}^{N}\left|\tfrac{\sin\frac{\theta}{2}}{\theta}+\tfrac{1}{4}g_m Z_c^g F_c M\left(k\right)\right|^2+\tfrac{1}{N^2}G_u Z_c^g\textstyle\sum_{k=1}^{N}\left|M\left(k\right)\right|^2 \end{aligned} \tag{7.55}$$

In order to gain insights, a simplified version of (7.55) is needed. For large number of sections, one can assume that the first term in (7.45) dominates, so two summations in (7.55) will be written as

$$\sum_{k=1}^{N}\left|\frac{\sin\frac{\theta}{2}}{\theta}+\frac{1}{4}g_m Z_c^g F_c M\left(k\right)\right|^2 \sim N\left(\frac{\sin\frac{\theta}{2}}{\theta}\right)^2+\frac{1}{16}g_m^2\left(Z_c^g\right)^2\left|F_c\right|^2\frac{N^3}{3}$$

$$\Sigma_{k=1}^{N}\left|M\left(k\right)\right|^2 \sim \frac{N^3}{3}$$

This essentially removes the correlation between the gate and drain noise current sources. The second and third term in (7.55) are small quantities for a large N, therefore are ignored. So the noise figure can be simplified as

$$F=1+\frac{1}{NZ_c^g}\frac{4\gamma}{\alpha}\frac{1}{g_m}\left(\frac{\sin\frac{\theta}{2}}{\frac{\theta}{2}}\right)^2+NZ_c^g\frac{\alpha\delta}{3}\frac{\omega^2 C_{gs}^2}{5g_m} \tag{7.56}$$

It is observed that there is an optimal value of $NZ_c^g$to minimize noise figure. It is easy to show that at a specific frequency when

$$NZ_c^g=\frac{2}{\alpha}\sqrt{\frac{15\gamma}{\delta}}\left(\frac{\sin\frac{\theta}{2}}{\frac{\theta}{2}}\right)\frac{1}{\omega C_{gs}} \tag{7.57}$$

the noise figure has a minimum value of

$$F_{min}=1+4\sqrt{\frac{\delta\gamma}{15}}\frac{\omega}{\omega_T}\left(\frac{\sin\frac{\theta}{2}}{\frac{\theta}{2}}\right) \tag{7.58}$$

where $\omega_T=\frac{g_m}{C_{gs}}$.

The sinc function in (7.56) dominates the low frequency portion of the noise figure. Between DC and the first null, the sinc is a descending function of frequency.

Thus the noise figure decreases with frequency increasing as shown in Fig. 119. When frequency increases the gate noise contribution denoted by the $\omega^2$ term in (7.56) becomes more significant and at the same time the gain drops from its ideal value, therefore noise figure goes up again with frequency increasing. [67] gives the noise figure analysis for distributed amplifiers using artificial transmission lines. The simplified form of noise figure is almost the same as derived here except that there is no sinc function in [67]. For the exact noise figure calculation, [67] did not consider the fact that the forward and reverse propagated gate noise currents are actually fully correlated.

### 4. Analysis of Distributed Mixer

Distributed mixers are constructed the same way as distributed amplifiers. Three types of distributed mixers are studied in [68]. Cascoded mixer mimics the dual-gate FET mixer. Matrix mixer utilizes three tiers of transistor to implement RF amplification, mixing and IF amplification. Balanced mixers use CG-CS FET pairs. By injecting signals from different terminals of a FET, different mixing modes can be obtained, [69] studied four different distributed mixing modes of FET. It demonstrated that LO-Gate/RF-Drain and LO-Source/RF-Gate mixing can provide conversion gain and good port isolation (RF/LO to IF 40 dB).

Fig. 121 is a typical distributed mixer block diagram. Ignore losses introduced by the T-lines and resistive loading by the transistors, assume the loaded characteristic impedances and propagation constants of the three transmission lines as followings: $Z_c^{RF}$ and $\beta_{RF}$ for RF line, $Z_c^{LO}$ and $\beta_{LO}$ for LO line, $Z_c^{IF}$ and $\beta_{IF}$ for IF line. And further assume all the T-lines are terminated by their characteristic impedances. The

RF signal traveling along the RF line can be written as

$$v_{RF}\left(x_{RF}\right) = V_{RF} e^{-j\beta_{RF} x_{RF}} \tag{7.59}$$

where $V_{RF}$ is the signal incidence into the RF port. The conversion transconductance $g_{mc}$ of a unit section is uniformly spread over the unit length $l_{IF}$ of IF line. The LO signal traveling along the LO T-line can be modeled by the phase-shift of conversion transconductance. Therefore the distributed conversion transconductance $g_m^{dist}\left(x_{LO}\right)$ is given by

$$g_m^{dist}\left(x_{LO}\right) = \frac{g_{mc}}{l_{IF}} e^{-j\beta_{LO} x_{LO}} \tag{7.60}$$

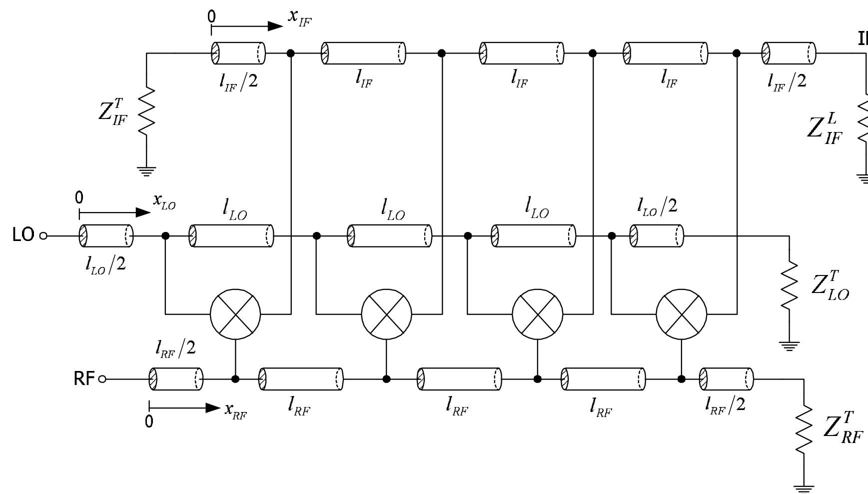



Fig. 121. A distributed mixer block diagram

The total voltage developed across the IF load impedance $Z_{IF}^L$ can be calculated by

$$V_{IF} = \frac{1}{2}\int_0^{Nl_{IF}} g_m^{dist}\left(x_{LO}\right) v_{RF}^*\left(x_{RF}\right) Z_c^{IF} e^{-j\beta_{IF}\left(Nl_{IF}-x_{IF}\right)} dx_{IF} \tag{7.61}$$

where down-conversion is assumed, N is the number of sections, and $v_{RF}^*\left(x_{RF}\right)$ rep-

resents the complex conjugate of $v_{RF}(x_{RF})$.

It is easy to show that

$$x_{LO} = \frac{l_{LO}}{l_{IF}} x_{IF} \tag{7.62}$$

and

$$x_{RF} = \frac{l_{RF}}{l_{IF}} x_{IF} \tag{7.63}$$

Substitute (7.59), (7.60), (7.62) and (7.63) into (7.61), the voltage conversion gain can be found to be

$$A_{vc} = \frac{V_{IF}}{V_{RF}} = \frac{N}{2} g_{mc} Z_c^{IF} e^{-jN\theta_{IF}} \left[ \frac{\sin \frac{N}{2}\Theta}{\frac{N}{2}\Theta} e^{-j\frac{N}{2}\Theta} \right] \tag{7.64}$$

where $\Theta = \theta_{LO} - \theta_{RF} - \theta_{IF}$. If $(L_{LO},\ C_{LO})$, $(L_{RF},\ C_{RF})$ and $(L_{IF},\ C_{IF})$ are the distributed parameters of the loaded LO, RF and IF T-lines respectively, then one can write:

$$\theta_{LO} = \beta_{LO} l_{LO} = \omega_{LO} l_{LO} \sqrt{L_{LO} C_{LO}} \tag{7.65}$$

$$\theta_{RF} = \beta_{RF} l_{RF} = \omega_{RF} l_{RF} \sqrt{L_{RF} C_{RF}} \tag{7.66}$$

$$\theta_{IF} = \beta_{IF} l_{IF} = \omega_{IF} l_{IF} \sqrt{L_{IF} C_{IF}} \tag{7.67}$$

The phase synchronization can be established by arranging the T-line parameters such that

$$l_{LO} \sqrt{L_{LO} C_{LO}} = l_{RF} \sqrt{L_{RF} C_{RF}} = l_{IF} \sqrt{L_{IF} C_{IF}} \tag{7.68}$$

Notice that $\omega_{IF} = \omega_{LO} - \omega_{RF}$ and combined with (7.68), then $\Theta = 0$ and the voltage conversion gain has constant amplitude and group delay over a valid frequency range:

$$A_{vc} = \frac{N}{2} g_{mc} Z_c^{IF} e^{-jN\theta_{IF}} \tag{7.69}$$

Similar to the distributed LNA voltage gain, the conversion gain and group delay is linearly proportional to the number of sections.

Consider the loss in the circuit, $\alpha_{RF}$, $\alpha_{LO}$ and $\alpha_{IF}$ are the attenuation constants of loaded RF, LO and IF T-lines. Furthermore, assume that the loss in the LO signal will introduce the corresponding reduction in the conversion transconductance $g_{mc}$. Under the phase synchronization condition of (7.68), the conversion power gain is approximately

$$G_c \approx \frac{g_{mc}^2 Z_c^{IF} Z_c^{RF}}{4} \left| \frac{e^{-N(\alpha_{RF} l_{RF} + \alpha_{LO} l_{LO})} - e^{-N(\alpha_{IF} l_{IF})}}{\alpha_{RF} l_{RF} + \alpha_{LO} l_{LO} - \alpha_{IF} l_{IF}} \right|^2 \tag{7.70}$$

If the LO signal is large enough such that its attenuation along the line will not affect the conversion transconductance, then the power conversion gain is not dependent on $\alpha_{LO}$:

$$G_c \approx \frac{g_{mc}^2 Z_c^{IF} Z_c^{RF}}{4} \left| \frac{e^{-N\alpha_{RF} l_{RF}} - e^{-N\alpha_{IF} l_{IF}}}{\alpha_{RF} l_{RF} - \alpha_{IF} l_{IF}} \right|^2 \tag{7.71}$$

This resembles the distributed LNA power gain equation (7.35), so the optimal number of sections is

$$N_{opt,\,mixer} = \frac{\ln(\alpha_{RF} l_{RF}) - \ln(\alpha_{IF} l_{IF})}{\alpha_{RF} l_{RF} - \alpha_{IF} l_{IF}} \tag{7.72}$$

and if the loss is dominated by the RF T-line loading due to the MOS transistor's gate resistance, (7.37) can also be applied to determine the upper bound of N by changing $Z_c^d$ to $Z_c^{IF}$.

### C. Wide-Band Impedance Match Using Lumped Components

In the wide-band LNA design, the input impedance matching is another major problem additional to bandwidth. There are several ways to provide matching. One solution is to use common gate or common base as input stage. Another way is to use feedback. These two methods can only work for a moderately wide bandwidth. For a wider bandwidth new topologies must be developed. Distributed structures

using artificial or real transmission lines have good wide band matching property. It provides 50 Ohm input and output intrinsically. Even for non-distributed circuit, transmission line matching network can also achieve relatively large bandwidth for prescribed matching requirement. For on-chip implementation, especially at lower frequency range (several GHz) transmission line's dimension will be prohibitively large. So using lumped components will be more appropriate.

### 1. Impedance Matching Procedure Using Lumped Components

Fig. 122 shows the detailed procedure to match a specific input impedance ($s_{11}$) over a specific frequency band to around the center of Smith chart using lumped components.

First, for a start point, the input impedance of a unilateral MOS transistor is obtained. This impedance can be modeled approximately by a capacitor and resistor in series. The capacitance comes from the gate-source capacitance

$$C_{gs} = C_{ox} W L_{ov} + \frac{2}{3} C_{ox} W L_{eff} \tag{7.73}$$

where $L_{ov}$ is the gate-source overlap length, $L_{eff}$ is the effective gate length which equals $L - 2L_{ov}$. $L$ is the drawn length of the gate. The resistance is due to poly resistance $R_{poly}$ and non-quasi-state (NQS) gate resistance $R_{nqs}$. The poly resistance comes from the physical resistance of gate poly material and will contributes noise. In order to reduce the gate poly resistance, multiple transistors usually connect in parallel instead of using just one big transistor. Suppose a transistor comprises number of $n$ smaller transistors with channel length $L$ and channel width $W$, then the gate poly resistance can be calculated from

$$R_{g,poly} = \frac{R_{sh,poly}}{12n^2} \frac{W}{L} \tag{7.74}$$

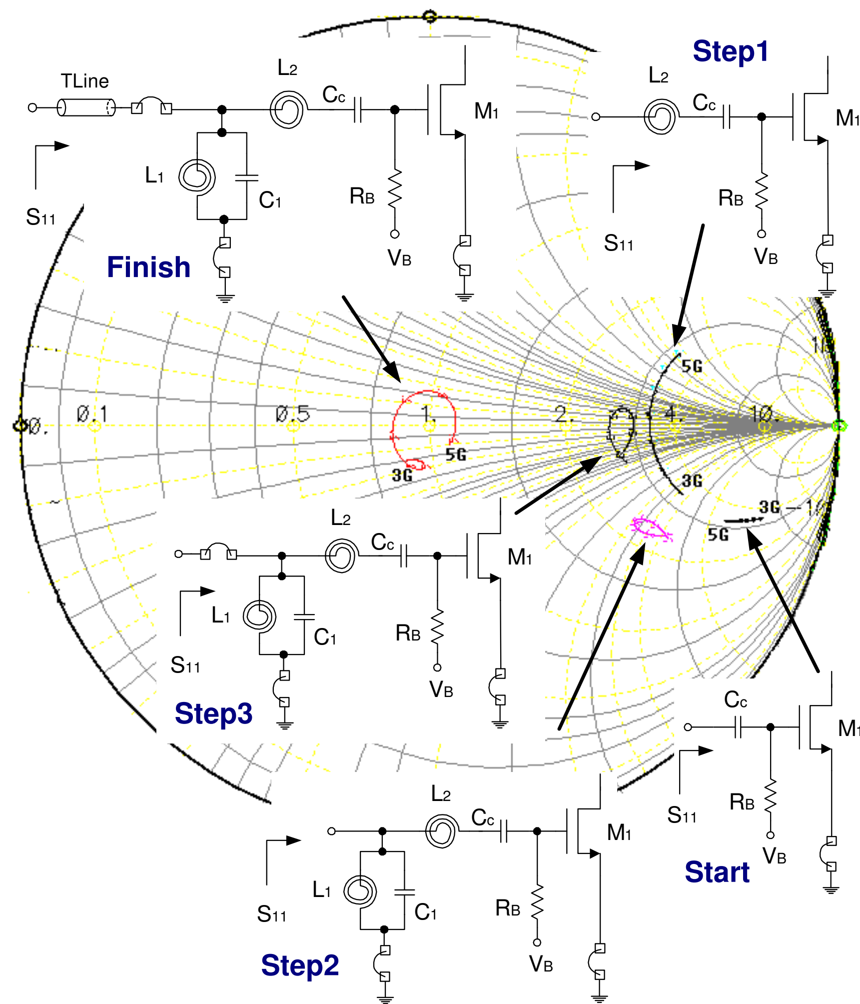



Fig. 122. Wide band matching procedure using lumped components

where $R_{sh,poly}$ is the gate poly sheet resistance. Here it is assumed the both ends of the gate are connected together in parallel.

For operation frequency comparable to transistor's cut-off frequency $f_T$, quasi-state (QS) assumption for the charge behavior under the gate dose not hold anymore. The NQS effect can be modeled by a resistor $R_{nqs}$ in series with the gate capacitor $C_{gs}$ [70]. Note that resistor $R_{nqs}$does not contribute to noise, because it is not a physical resistor. The value of $R_{nqs}$ is approximately $\frac{1}{5g_m}$ [70].

The start point impedance is plotted in the Smith chart over the desired frequency range. The impedance curve is located at position *Start* in Fig. 122. Then the following steps are carried out:

1. Series a proper value of inductance such that the conductance at the frequency edges have the same real part. The impedance curve should move to the position *Step1* in Fig. 122.

2. Parallel an inductor and capacitor tank to bring the frequency edges close to each other. The impedance contour resembles a circle and moves to the position *Step2* in Fig. 122.

3. Series an inductor again to move the center of the impedance contour to a pure resistance point as at the position *Step3* in Fig. 122. This inductor can be implemented using bond wire and/or off-chip inductors.

Finally, a quarter-wave transmission line with a proper characteristic impedance is used to rotate the impedance contour to the center of the Smith Chart, as shown at the position *Finish* in Fig. 122. The output impedance matching can be achieved similarly if required.

2. Wide-Band LNA with Lumped-Matched Network

An implementation of wide-band LNA using the impedance matching technique discussed in the previous section is shown in Fig. 123. This circuit is simulated using TSMC 0.18 $\mu m$ CMOS technology. It draws 5 mA from a 1.8 V power supply.

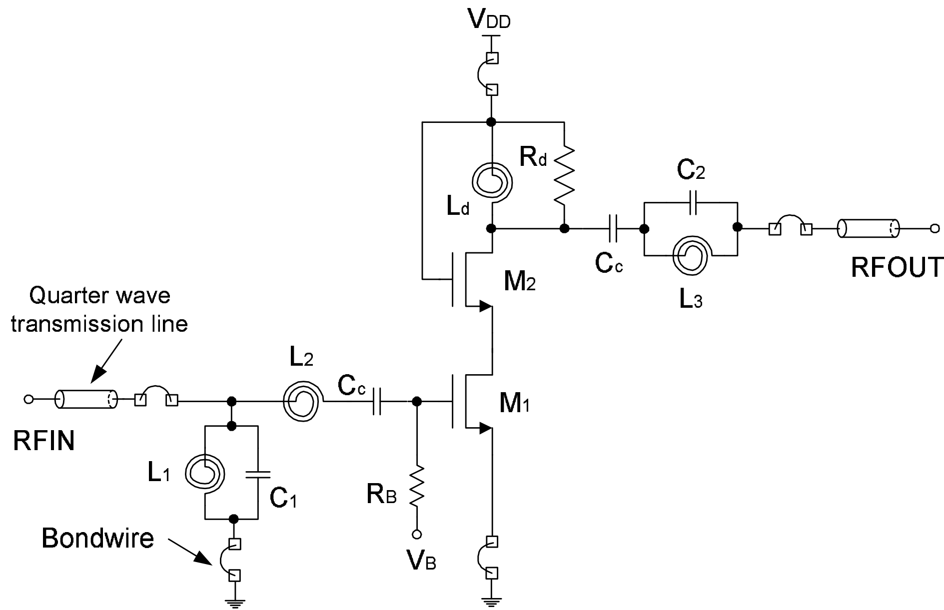



Fig. 123. Wide-band LNA with lumped-matched network

Fig. 124 gives the S-parameters and noise figure plots from 2.8 GHz to 5.0 GHz. Table XXI summarizes the post-layout simulation results of this wide-band LNA. This LNA can cover the first 3 bands (from 3.168 GHz to 4.752 GHz) of Texas Instruments OFDM UWB proposal and can be used in its Mode 1 UWB devices [71].

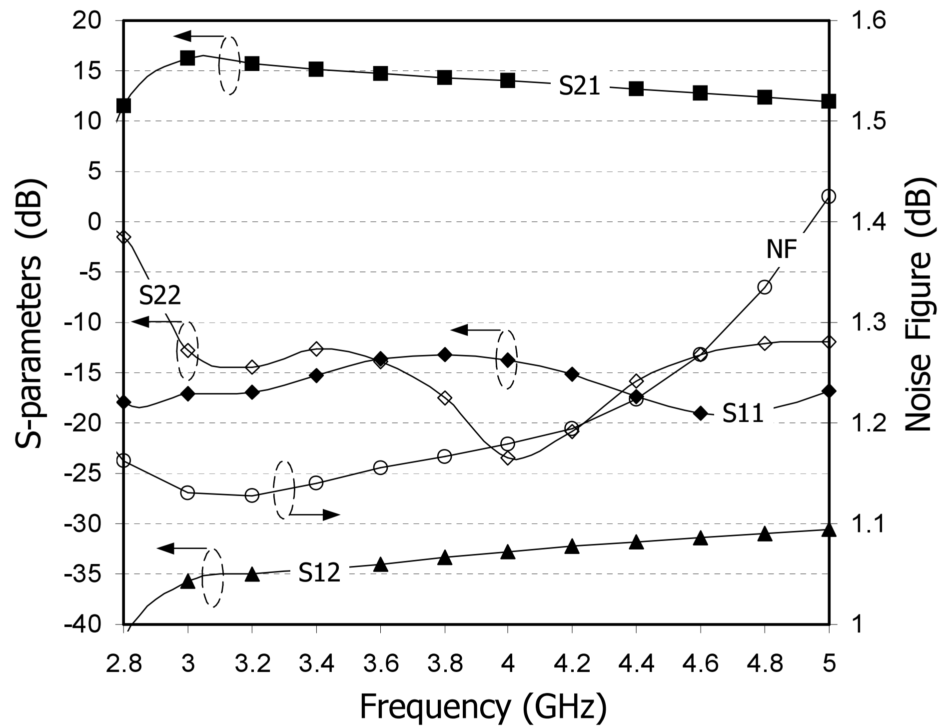



Fig. 124. Wide-band LNA S-parameters and noise figure

Table XXI. Simulation results of lumped-matched wide-band LNA

| Parameter | Value | Unit |
|---|---|---|
| Bandwidth | 2.8-5.0 | GHz |
| $S_{11}$ | $<$ -13 | dB |
| $S_{21}$ | $>$ 12 | dB |
| Noise Figure | $<$ 1.9 | dB |
| $P_{1\text{dB}}$@ input | -4.4 | dBm |
| Power | 9 | mW |

# CHAPTER VIII

# SUMMARY AND CONCLUSION

In the dissertation, LNA design aspects and the major design considerations were first reviewed. In order to implement a good design, a lot of information from the process capability to system requirements are needed to give access to RF designers. This makes the seemingly simple circuit of a LNA actually the trickiest part in the front end of a receiver system. Two design approaches exist. One is the microwave approach, which first selects a specific microwave transistor according to the data sheet obtained from manufacture. The choice of device relies on its gain, noise and linearity capability. Then trade-offs are made for gain, impedance match and noise match. The matching network at input and output will finally be implemented considering the above merits and stability. The way RF IC designers are adopted is more complicated. In IC design, one has the freedom to change the geometry of the devices as well as bias condition. The microwave design target is to find an optimal matching network for a specific device, but the IC designers want to find an optimal structure for all possible devices. Of course, this is not always possible, but the freedom of varying device sizes really gives more room for IC designers. Another consideration is that if the LNA does not require driving off-chip filters, it is not necessary to match its output to 50 ohm or even there is no need to realize matching network between the output of the LNA and the input of the next stage (usually mixers). Of course, proper interface between them still needs to be implemented.

As an important frequency translation functional block, the mixer design deals with three different frequencies. Mixer translates the amplified signal form the LNA to a frequency band suitable for the baseband circuit to process. Conversion gain

shows the mixer's frequency translation efficiency. A mixer is more noisy than a LNA. Especially for low-IF applications, low frequency flicker noise dominates the mixer's noise performance. It is vital to choose a proper IF frequency such that the MOS transistor's flicker noise corner is lower than the IF frequency. Mixers in silicon implementation usually use double-balanced structure to improve port isolations and suppress common-mode noise and interferences. A low-IF Bluetooth receiver is designed and tested. The mixer for this receiver using the current-bleeding technique to reduce the flicker noise effect of the current switching pairs. Measurements show that the mixer worked properly with other blocks in the whole system.

A detailed implementation was discussed for a Blutooth/WiFi dual-mode direct conversion receiver RF front-end. The LNA features inductive degeneration, gate-induced noise reduction and bipolar cascoded techniques and the down-conversion mixer has combined RF driver and bipolar current switching pairs. A PTAT biasing circuit is also included for stable bias condition setup. The dual-mode receiver with this front-end has sensitivity for Bluetooth mode -91 dBm and for WiFi mode (11 Mb/s) -86 dBm.

More research effort was exerted on LNA design techniques. With the development of the IC fabrication, bipolar transistor will be readily available in CMOS process. For the LNA design, the possibility of using both MOS transistor and bipolar transistor was explored. This hybrid structure was inspired by the multi-gated MOS configuration for linearization purpose. It is the first structure that MOS transistor and BJT are working together at the signal path. A mutual inductive degenerated LNA designed for dual-band application was studied to show the benefits of mutual inductive coupling in the LNA design.

UWB technology can provide costumers more bandwidth, high speed and versatile services. It has enormous potential applications, especially in the wireless PAN

standards. Distributed circuits are applicable in the RF front-end of UWB systems due to their inherent wide-band nature. For lower frequency bands, lumped-matched wide-band LNA occupies less silicon area and is more power efficient. A successful design of a UWB system will require multi-disciplinary knowledge, such as channel measurements, antenna theory, RF and high speed analog design, software development and system simulation. For the RF front-end portion, the LNA/antenna co-design may be needed for optimal performance.

To conclude, the contributions of the dissertation include: design RF front-end for a low-IF Bluetooth receiver and a direct conversion Bluetooth/WiFi dual mode receiver (Chameleon); LNA linearization technique using MOS/BJT hybrid; mutual inductive degeneration for a dual-band cellular LNA; analysis of distributed LNA and mixer, and the proposed procedure to achieve wide band impedance match using lumped components.

# REFERENCES


[1] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design.* Upper Saddle River, N.J.: Prentice Hall, 2nd ed., 1997.

[2] R. Ludwig and P. Bretchko, *RF Circuit Design Theory and Applications.* Upper Saddle River, N.J.: Prentice Hall, book and CD-ROM ed., Nov. 1999.

[3] T. H. Lee, *The Design of CMOS Radio Frequency Integrated Circuits.* New York: Cambridge University Press, 2000.

[4] B. Razavi, *RF Microelectronics.* Upper Saddle River, NJ: Prentice Hall, 1998.

[5] J. Rogers and C. Plett, *Radio Frequency Integrated Circuit Design.* Boston, MA: Artech House Microwave Library, 2003.

[6] A. N. Karanicolas, "A 2.7-V 900-MHz CMOS LNA and mixer," *IEEE J. of Solid-State Circuits*, vol. 31, pp. 1939–1944, Dec. 1996.

[7] Y. Ge and K. Mayaram, "A comparative analysis of CMOS low noise amplifiers for RF applications," *IEEE Int. Symp. on Circuits and Systems*, vol. 4, pp. 349–352, May 1998.

[8] F. Bruccoleri, E. Klumperink, and B. Nauta, "Noise cancelling in wideband CMOS LNAs," *IEEE Int. Solid-State Circuits Conf.*, vol. 1, pp. 406–407, Feb. 2002.

[9] F. Bruccoleri, E. Klumperink, and B. Nauta, "Generating all two-MOS-transistor amplifiers leads to new wide-band LNAs," *IEEE J. of Solid-State Circuits*, vol. 36, pp. 1032–1040, Jul. 2001.

[10] D. K. Shaeffer and T. H. Lee, "A 1.5V, 1.5GHz CMOS low noise amplifier," *IEEE J. of Solid-State Circuits*, vol. 32, pp. 745–759, May 1997.

[11] P. Andreani and H. Sjöland, "Noise optimization of an inductively degenerated CMOS low noise amplifier," *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 48, pp. 835–841, Sept. 2001.

[12] G. Gramegna, M. Paparo, P. G. Erratico, and P. D. Vita, "A sub-1-dB NF ±2.3-kV ESD-protected 900-MHz CMOS LNA," *IEEE J. of Solid-State Circuits*, vol. 36, pp. 1010–1017, Jul. 2001.

[13] V. Geffroy, G. D. Astis, and E. Bergeault, "RF mixers using standard digital CMOS 0.35$\mu$m process," *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, pp. 83–86, May 2001.

[14] T. Yamaji and H. Tanimoto, "A 2 GHz balanced harmonic mixer for direct-conversion receivers," *IEEE Custom Integrated Circuits Conf.*, pp. 193–196, May 1997.

[15] W. Namgoong and T. H. Meng, "Direct-conversion RF receiver design," *IEEE Trans. on Communications*, vol. 49, pp. 518–529, Mar. 2001.

[16] J. Crols and M. S. Steyaert, "Low-IF topologies for high-performance analog front ends of fully integrated receivers," *IEEE Trans. On Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 45, pp. 269–282, Mar. 1998.

[17] W. Sheng, B. Xia, A. E. Emira, C. Xin, A. Y. Valero-López, S. T. Moon, and E. Sánchez-Sinencio, "A 3-v, 0.35$\mu$m CMOS Bluetooth receiver IC," *IEEE J. of Solid-State Circuits*, vol. 38, pp. 30–42, Jan. 2003.

[18] S.-G. Lee and J.-K. Choi, "Current-reuse bleeding mixer," *Electronics Letters*, vol. 36, pp. 696–697, Apr. 2000.

[19] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. New York: John Wiley & Sons, Inc., 4th ed., 2001.

[20] H. Aoki and M. Shimasue, "Channel width and length dependent flicker noise characterization for n-MOSFETs," *Proc. of the 2001 Int. Conf. on Microelectronic Test Structures*, vol. 14, pp. 257–261, 2001.

[21] L. A. NacEachern and T. Manku, "A charge-injection method for Gilbert cell biasing," *IEEE Canadian Conf. on Electrical and Computer Engineering*, vol. 1, pp. 365–368, May 1998.

[22] W. Sansen, "Distortion in elementary transistor circuits," *IEEE Trans. On Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 46, pp. 315–325, March 1999.

[23] M. T. Terrovitis and R. G. Meyer, "Noise in current-commutating CMOS mixers," *IEEE J. of Solid-State Circuits*, vol. 34, pp. 772–783, June 1999.

[24] R. Shepherd, "Bluetooth wireless technology in the home," *Electronics and Communication Engineering Journal*, vol. 13, pp. 195–203, Oct. 2001.

[25] P. S. Henry and H. Luo, "WiFi: what's next?," *IEEE Communications Magazine*, vol. 40, pp. 66–72, Dec. 2002.

[26] D. Coursey, "Bluetooth vs. WiFi: Why it's NOT a death match." http://techupdate.zdnet.com/techupdate/stories/main/0,14179,2868374,00.html, May 2002.

[27] A. E. Emira, A. Valdes-Garcia, B. Xia, A. Mohieldin, A. Y. Valero-López, S. T. Moon, C. Xin, and E. Sánchez-Sinencio, "A dual-mode 802.11b/Bluetooth receiver in 0.25$\mu$m BiCMOS," *IEEE Int. Solid-State Circuits Conf.*, vol. 47, pp. 270–271, 527, Feb. 2004.

[28] X. Li, T. Brogan, M. Esposito, B. Myers, and K. K. O, "A comparison of CMOS and SiGe LNA's and mixers for wireless LAN application," *IEEE Conf. on Custom Integrated Circuits*, pp. 531–534, May 2001.

[29] Q. Huang, P. Orsatti, and F. Piazza, "Broadband, 0.25 $\mu$m CMOS LNAs with sub-2dB NF for GSM applications," *Proc. of the IEEE Custom Integrated Circuits Conf.*, pp. 67–70, May 1998.

[30] Y. Cheng, M. Chan, K. Hui, M. chie Jeng, Z. Liu, J. Huang, K. Chen, J. Chen, R. Tu, P. K. Ko, and C. Liu, *BSIM3v3 Manual.* Dept. of Electrical Engineering and Computer Sciences, University of California, Berkeley, 1996.

[31] A. Hastings, *The Art of Analog Layout.* Upper Saddle River, NJ: Prentice-Hall, Inc., 2001.

[32] A. M. Niknejad, "ASITIC: Analysis and simulation of spiral inductors and transformers for ICs." http://rfic.eecs.berkeley.edu/∼niknejad/asitic.html.

[33] J. Ryynänen, K. Kivekäs, J. Jussila, A. Pärssinen, and K. A. I. Halonen, "A dual-band RF front-end for WCDMA and GSM applications," *IEEE J. of Solid-State Circuits*, vol. 36, pp. 1198–1204, Aug. 2001.

[34] H. Sjoland, A. Karimi-Sanjaani, and A. A. Abidi, "A merged CMOS LNA and mixer for a WCDMA receiver," *IEEE J. of Solid-State Circuits*, vol. 38, pp. 1045–1050, Jun. 2003.

[35] D. Johns and K. Martin, *Analog Integrated Circuit Design.* New York: John Wiley & Sons, Inc., 1997.

[36] L. E. Larson, “Analog/RF design in SiGe BiCMOS processes,” *Int. Solid-State Circuits Conf. Short Course*, Feb. 2004.

[37] B. Leung, *VLSI for Wireless Communication.* Upper Saddle River, N.J.: Prentice Hall, 2002.

[38] K. L. Fong and R. G. Meyer, “High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages,” *IEEE J. of Solid-State Circuits*, vol. 33, pp. 548–555, Apr. 1998.

[39] V. Aparin and C. Persico, “Effect of out-of-band terminations on intermodulation distortion in common-emitter circuits,” *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 3, pp. 977–980, Jun. 1999.

[40] B. Kim, J.-S. Ko, and K. Lee, “Highly linear CMOS RF MMIC amplifier using multiple gated transistors and its volterra series analysis,” *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, pp. 515–518, May 2001.

[41] T. W. Kim, B. Kim, and K. Lee, “Highly linear receiver front-end adopting MOSFET transconductance linearization by multiple gated transistors,” *IEEE J. of Solid-State Circuits*, vol. 39, pp. 223–229, Jan. 2004.

[42] B. Kim, J.-S. Ko, and K. Lee, “A new linearization technique for MOSFET RF amplifier using multiple gated transistors,” *IEEE Microwave and Guided Wave Letters*, vol. 10, pp. 371–373, Sept. 2000.

[43] T. W. Kim, B. Kim, I. Nam, B. Ko, and K. Lee, “A low-power highly linear cascoded multiple-gated transistor CMOS RF amplifier with 10dB IP3 improve-

ment (revised)," *IEEE Microwave and Wireless Components Letters*, vol. 13, pp. 420–422, Sept. 2003.

[44] S. S. Weng, L. K. Chong, C. K. Vai, C. W. Wa, K. W. Tam, and R. P. Martins, "A analytical linearization method for CMOS MMIC power amplifier using multiple gated transistors," *Proc. of 4th Int. Conf. on ASIC*, pp. 670–672, Oct. 2001.

[45] M. T. Terrovitis and R. G. Meyer, "Intermodulation distortion in current-commutating CMOS mixers," *IEEE J. of Solid-State Circuits*, vol. 35, pp. 1461–1473, Oct. 2000.

[46] T. W. Kim, B. Kim, and K. Lee, "Highly linear RF CMOS amplifier and mixer adopting MOSFET transconductance linearization by multiple gated transistors," *IEEE Radio Frequency Integrated Circuits Symp.*, pp. 107–110, Jun. 2003.

[47] V. Aparin and L. E. Larson, "Modified derivative superposition method for linearizing fet Low noise amplifiers," *IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, June 2004.

[48] S. Yan and E. Sánchez-Sinencio, "Low voltage analog circuit design techniques: A tutorial," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E83-A, pp. 179–195, Feb. 2000.

[49] Y. Ding and R. Harjani, "A +18dBm IIP3 LNA in 0.35$\mu$m CMOS," *IEEE ISSCC Dig. of Tech. Papers*, pp. 162–163, 443, Feb. 2001.

[50] V. Aparin, E. Zeisel, and P. Gazzerro, "Highly linearSiGe BiCMOS LNA and mixer for cellular CDMA/AMPS applications," *IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, pp. 129–132, June 2002.

[51] H. Hashemi and A. Hajimiri, “Concurrent multiband low-noise amplifiers-theory, design, and applications,” *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, pp. 288–301, Jan. 2002.

[52] S. H. M. Lavasani, B. Chaudhuri, and S. Kiaei, “A pseudo-concurrent 0.18$\mu$m multi-band CMOS LNA,” *IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, pp. 695–698, Jun. 2003.

[53] J. Long, N. Badr, and R. Weber, “A 2.4GHz sub-1 dB CMOS low noise amplifier with on-chip interstage inductor and parallel intrinsic capacitor,” *IEEE Radio and Wireless Conf.*, pp. 165–168, Aug. 2002.

[54] W.-S. Kim, X. Li, and M. Ismail, “A 2.4 GHz CMOS low noise amplifier using an inter-stage matching inductor,” *42nd Midwest Symp. on Circuits and Systems*, vol. 2, pp. 1040–1043, Aug. 1999.

[55] W. Guo and D. Huang, “The noise and linearity optimization for a 1.9-GHz CMOS low noise amplifier,” *IEEE Asia-Pacific Conf. on ASIC*, pp. 253–257, Aug. 2002.

[56] P. Orsatti, F. Piazza, Q. Huang, and T. Morimoto, “A 20 mA-receive 55 mA-transmit GSM transceiver in 0.25$\mu$m CMOS,” *IEEE Int. Solid-State Circuits Conf.*, pp. 232–233, Feb. 1999.

[57] S. Tadjpour, E. Cijvat, E. Hegazi, and A. A. Abidi, “A 900-MHz dual-conversion low-IF GSM receiver in 0.35-$\mu$m CMOS,” *IEEE J. of Solid-State Circuits*, vol. 36, pp. 1992–2002, Dec. 2001.

[58] K. Sharaf and H. ElHak, “A compact approach for the design of a dual-band low-noise amplifier,” *Proc. of the 44th IEEE 2001 Midwest Symp. on Circuits*

*and Systems*, vol. 2, pp. 890–893, Aug. 2001.

[59] "IEEE 802.15WPAN high rate alternative PHY task group 3a (TG3a)." http://www.ieee802.org/15/pub/TG3a.html.

[60] G. R. Aiello, "Challenges for ultra-wideband (UWB) CMOS integration," *IEEE Radio Frequency Integrated Circuits (RFIC) Symp.*, pp. 497–500, Jun. 2003.

[61] J. G. Proakis, *Digital Communications.* New York: The McGraw-Hill Companies, Inc., fourth ed., 2001.

[62] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works," *IEEE Communications Letters*, vol. 2, pp. 36–38, Feb. 1998.

[63] W. R. Eisenstadt and Y. Eo, "S-parameter-based IC interconnect transmission line characterization," *IEEE Trans. on Components, Hybrids, and Manufacturing Technology*, vol. 15, pp. 483–490, Aug. 1992.

[64] E. A. M. Klumperink, R. Kreienkamp, T. Ellermeyer, and U. Langmann, "Transmission lines in CMOS: An explorative study," *Proc. of ProRISC*, Nov. 2001.

[65] B. Kleveland, C. H. Diaz, D. Voook, L. Madden, T. H. Lee, and S. S. Wong, "Exploiting CMOS reverse interconnect scaling in multigigahertz amplifier and oscillator design," *IEEE J. of Solid-State Circuits*, vol. 36, pp. 1480–1488, Oct. 2001.

[66] T. T. Wong, *Fundamentals of Distributed Amplification.* Boston, MA: Artech House, 1993.

[67] C. S. Aitchison, "The intrinsic noise figure of the MESFET distributed amplifier," *IEEE Trans. on Microwave Theory and Techniques*, vol. 33, pp. 460–466, Jun. 1985.

[68] I. D. Robertson and A. H. Aghvami, “Multi-octave MMIC distributed mixers,” *IEE Colloquium on Multi-Octave Microwave Circuits*, pp. 4/1–4/7, Nov. 1991.

[69] M. Fairburn, B. J. Minnis, and J. Neale, “A novel monolithic distributed mixer design,” *IEE Colloquium on Microwave and Millimetre Wave Monolithic Integrated Circuits*, pp. 13/1–13/6, Nov. 1988.

[70] Y. Tsividis, *Operation and modeling of the MOS transistor*. New York: McGraw Hill, second ed., 1999.

[71] A. Batra, Texas Instruments et al., “Multi-band OFDM physical layer proposal,” *TG3a Multi-band OFDM CFP Presentation*, Sept. 2003.

[72] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: John Wiley & Sons, 1980.

# APPENDIX A

# CONSTANT CIRCLES

## Circle Equation in the Complex Plane

A circle in complex plane can be represented by

$$|z|^2 - c * z - cz* = b \tag{A.1}$$

where $z$ is the point on the circle, $c$ is a complex constant number, $b$ is a real constant number and $b + |c|^2 > 0$. The proof is given as follows.

Adding $|c|^2$ to both sides of (A.1) leads to $|z|^2 - c * z - cz * + |c|^2 = b + |c|^2$. Since $|z|^2 - c * z - cz * + |c|^2 = |z - c|^2$, then $|z - c|^2 = b + |c|^2$ and $|z - c| = \sqrt{b + |c|^2}$. Because $b + |c|^2 > 0$, it is the equation for a circle with center located at point $c$ and having radius $r = \sqrt{b + |c|^2}$.

## Constant Gain Circle

Taking the input mismatch factor (**??**) for example. We will prove that for fixed value of $s_{11}$ and $G_s$, $\Gamma_s$ resides on a circle and find its center and radius.

Let us define $g_s = G_s \left(1 - |s_{11}|^2\right) = \frac{G_s}{G_s \max}$. Substituting (**??**) in to $g_s$ gives

$$g_s = \frac{\left(1 - |s_{11}|^2\right)\left(1 - |\Gamma_s|^2\right)}{|1 - s_{11}\Gamma_s|^2}$$

$$g_s (1 - s_{11}\Gamma_s)(1 - s_{11}^*\Gamma_s^*) = 1 - |s_{11}|^2 - |\Gamma_s|^2 + |s_{11}|^2 |\Gamma_s|^2$$

$$g_s \left(1 - s_{11}\Gamma_s - s_{11}^*\Gamma_s^* + |s_{11}|^2 |\Gamma_s|^2\right) = 1 - |s_{11}|^2 - |\Gamma_s|^2 + |s_{11}|^2 |\Gamma_s|^2$$

Finally,

$$|\Gamma_s|^2 - \frac{g_s s_{11}}{1-|s_{11}|^2(1-g_s)}\Gamma_s - \frac{g_s s_{11}^*}{1-|s_{11}|^2(1-g_s)}\Gamma_s^* = \frac{(1-g_s)-|s_{11}|^2}{1-|s_{11}|^2(1-g_s)} \tag{A.2}$$

Notice that $g_s$ is a real number and $0 < g_s < 1$, according to (A.1), (A.2) represents a circle with

$$c = \frac{g_s s_{11}^*}{1-|s_{11}|^2(1-g_s)} \tag{A.3}$$

which is the center of the circle, and

$$b = \frac{(1-g_s)-|s_{11}|^2}{1-|s_{11}|^2(1-g_s)}$$

The radius of the circle is

$$r = \sqrt{b+|c|^2} = \frac{\sqrt{1-g_s}\left(1-|s_{11}|^2\right)}{1-|s_{11}|^2(1-g_s)}$$

The proof for output constance circle can be carried out similarly.

**Constant Noise Circle**

We want to show that (2.66) is a circle equation for variable $\Gamma_s$. Rewriting (2.66) as

$$(\Gamma_s - \Gamma_{opt})\left(\Gamma_s^* - \Gamma_{opt}^*\right) = N\left(1-|\Gamma_s|^2\right)$$

and then

$$(1+N)|\Gamma_s|^2 - \Gamma_{opt}^*\Gamma_s - \Gamma_{opt}\Gamma_s^* = N - |\Gamma_{opt}|^2$$

at last

$$|\Gamma_s|^2 - \left(\frac{\Gamma_{opt}}{1+N}\right)^*\Gamma_s - \left(\frac{\Gamma_{opt}}{1+N}\right)\Gamma_s^* = \frac{N-|\Gamma_{opt}|^2}{1+N} \tag{A.4}$$

According to (A.1), (A.4) describes a circle with

$$c = \frac{\Gamma_{opt}}{1+N}$$

which is center location, and

$$b = \frac{N - |\Gamma_{opt}|^2}{1+N}$$

The radius of the circle is

$$r = \sqrt{b + |c|^2} = \frac{N}{1+N}\sqrt{1 + \frac{1}{N}\left(1 - |\Gamma_{opt}|^2\right)}$$

# APPENDIX B

# VOLTERRA SERIES AND EXAMPLES

Vito Volterra first studied the functional series named after him in the 1880s as a generalization of the Taylor series expansion of a non-linear function. In 1942, N. Wiener applied the Volterra theory to the analysis of a series RLC circuit with a non-linear resistor to a white Gaussian excitation. The systematic study of the application of the Volterra series to non-linear system was conducted by J. F. Barrett in 1957. D. A. George extended Barrett's work and developed a system algebra and used the multidimensional Laplace transformation to study Volterra operators and their application to non-linear system. The development of the Volterra theory has led to an extensive study of its application to practical problems in may fields including the calculation of small, but nevertheless troublesome, distortion terms in transistor amplifiers and systems.

## Volterra Series Representation of Non-Linear Systems

A linear, causal and time-invariant system with memory can be described by the convolution integral

$$y\left(t\right)=\int_{-\infty}^{\infty}h\left(\tau\right)x\left(t-\tau\right)d\tau \tag{B.1}$$

where $h\left(t\right)$ is the impulse response of the system. A non-linear system without memory can be represented using a Taylor series

$$y\left(t\right)=K_1x\left(t\right)+K_2\left[x\left(t\right)\right]^2+K_3\left[x\left(t\right)\right]^3+\cdots+K_n\left[x\left(t\right)\right]^n+\cdots \tag{B.2}$$

A Volterra series combines (B.1) and (B.2) to describe a non-linear system with

memory [72]

$$y(t) = \mathbf{H}_1[x(t)] + \mathbf{H}_2[x(t)] + \mathbf{H}_3[x(t)] + \cdots + \mathbf{H}_n[x(t)] + \cdots \tag{B.3}$$

in which

$$\mathbf{H}_n[x(t)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n)\, x(t-\tau_1) \cdots x(t-\tau_n)\, d\tau_1 \cdots d\tau_n \tag{B.4}$$

Therefore, Volterra series is an infinite sum of n-fold convolution integrals. $\mathbf{H}_n[\cdot]$ is called the n-th order Volterra operator. $h_n(\tau_1, \cdots, \tau_n)$is the n-th order Volterra kernel and for $n = 1, 2, \cdots$,

$$h_n(\tau_1, \cdots, \tau_n) = 0 \qquad \text{for any}\, \tau_j < 0,\ j = 1, 2, \cdots, n \tag{B.5}$$

In general, kernel $h_n(\tau_1, \cdots, \tau_n)$ is not symmetrical to its variables. However, it can be shown that it is always possible to construct a symmetrical kernel from an asymmetrical form of the kernel $h_n^{(a)}(\tau_1, \cdots, \tau_n)$:

$$h_n^{(s)}(\tau_1, \cdots, \tau_n) = \frac{1}{n!} \sum_P h_n^a(\tau_1, \cdots, \tau_n) \tag{B.6}$$

where the summation runs through all possible permutations of the n $\tau$'s.

To gain more insights, one would like to study the non-linear system in frequency domain. The basic integral transform, Fourier transform can be extended to the Volterra series representation. For an n-th order kernel $h_n(\tau_1, \cdots, \tau_n)$, its n-dimensional Fourier transform is

$$H_n(j\omega_1, \cdots, j\omega_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n)\, e^{j(\omega_1\tau_1 + \cdots + \omega_n\tau_n)} d\tau_1 \cdots d\tau_n \tag{B.7}$$

If the input's Fourier transform is $X(j\omega)$, then the Fourier transform of the n-th

order system (B.4) can be shown to be

$$\begin{aligned} Y_n(j\omega) &= \tfrac{1}{(2\pi)^{n-1}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(j\omega - j\nu_1,\, j\nu_1 - j\nu_2,\, j\nu_2 - j\nu_3,\, \cdots,\, j\nu_{n-1}) \\ &\times\ X(j\omega - j\nu_1)\, X(j\nu_1 - j\nu_2)\, X(j\nu_2 - j\nu_3) \cdots X(j\nu_{n-1})\, d\nu_1 \cdots d\nu_n \end{aligned} \tag{B.8}$$

The Fourier transform of the whole Volterra series will be the sum of different orders. Written the first three terms explicitly in terms of frequency $f$:

$$\begin{aligned} Y(f) &= H_1(f)\, X(f) \\ &+\ \int_{-\infty}^{\infty} H_2(f - f_1,\, f_1)\, X(f - f_1)\, X(f_1)\, df_1 \\ &+\ \int_{-\infty}^{\infty} H_3(f - f_1,\, f_1 - f_2,\, f_2)\, X(f - f_1)\, X(f_1 - f_2)\, X(f_2)\, df_1 df_2 \\ &+\ \cdots \end{aligned} \tag{B.9}$$

**Circuit Model of a Non-linear Time-Invariant System**

In a general non-linear system or circuit, the memory effect and non-linearity are not distinctly separated, which makes the circuit representation rather complicated even if possible. An important simplification assumes that frequency dependent signal paths are multiplied so that the multiplication does not affect the time constant of the following filters. For example, a lot of circuits can be approximated as a non-linear memoryless gain stage with input and output filters, as shown in Fig. 125. The non-linear gain stage is represented by the Taylor series as in (B.1). The first order response of this system is simply

$$H_1(j\omega) = F(j\omega)\, K_1 G(j\omega) \tag{B.10}$$

The second and third order responses are depicted in Fig. 126. For the second order response, the input filter is evaluated at two different frequencies, then the response are multiplied and further scaled by the second order gain and output filter response

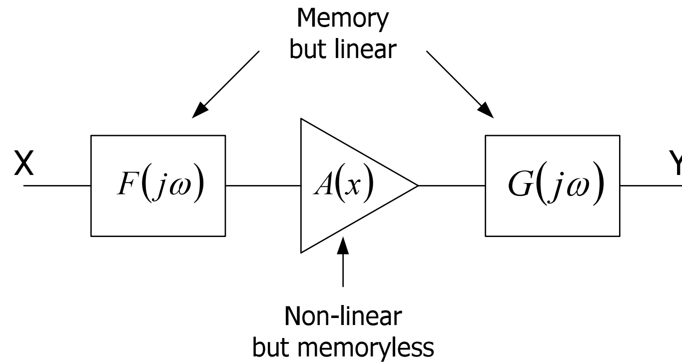



Fig. 125. Circuit model of a non-linear time-invariant system

at the output frequency:

$$H_2\left(j\omega_1,\, j\omega_2\right) = F\left(j\omega_1\right) F\left(j\omega_2\right) K_2 G\left[j\left(\omega_1 + \omega_2\right)\right] \tag{B.11}$$

Similarly, the third order response is

$$H_3\left(j\omega_1,\, j\omega_2,\, j\omega_3\right) = F\left(j\omega_1\right) F\left(j\omega_2\right) F\left(j\omega_3\right) K_3 G\left[j\left(\omega_1 + \omega_2 + \omega_3\right)\right] \tag{B.12}$$

For sinusoidal inputs, a short-hand notation is used to represent the output frequency components of the non-linear system. Suppose the input has m frequency components

$$X = X_1 \cos\omega_1 t + X_2 \cos\omega_2 t + \cdots + X_m \cos\omega_m t \tag{B.13}$$

The output of the system can be denoted as

$$Y = H_1\left(j\omega_{p1}\right) \circ X + H_2\left(j\omega_{p1},\, j\omega_{p2}\right) \circ X^2 + H_3\left(j\omega_{p1},\, j\omega_{p2},\, j\omega_{p3}\right) \circ X^3 + \cdots \tag{B.14}$$

where $\omega_{p1}$, $\omega_{p1}$, …, $\omega_{pn}$ choose from all the possible n out of m permutations of $\pm\omega_1$, $\pm\omega_2$, …, $\pm\omega_m$. $X^n$ represents the corresponding amplitude $X_{p1}X_{p2}\cdots X_{pn}$. The operator $\circ$ generates all the items $\cos\left(\omega_{p1} + \omega_{p2} + \cdots + \omega_{pn}\right)$ and scale their magnitudes

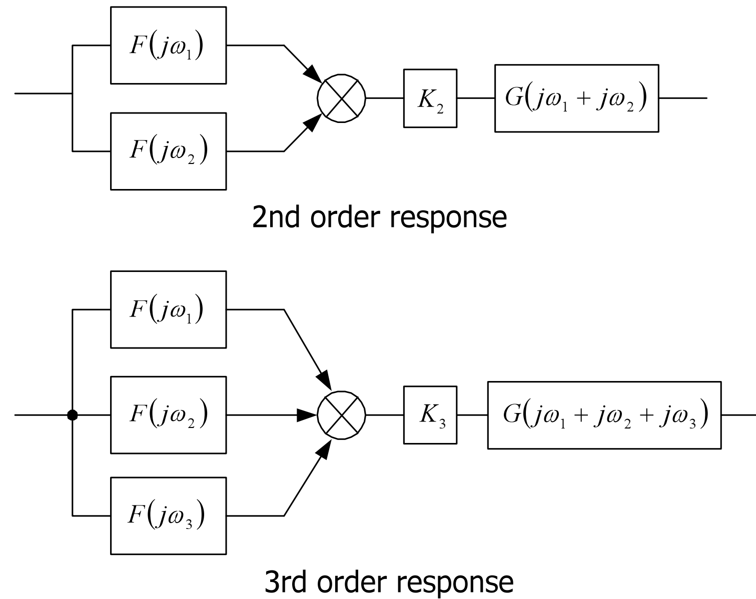



Fig. 126. Second and third order response block diagrams

with $|H_n (j\omega_{p1}, j\omega_{p2}, \cdots j\omega_{pn})|$ and modify their phases with $\angle H_n (j\omega_{p1}, j\omega_{p2}, \cdots j\omega_{pn})$. The short-hand notation (B.14) is usually used for actual circuit non-linear analysis.

**An Example: Volterra Series of a MOS Differential Pair**

In this section, a detailed procedure for deriving a MOS differential pair's Volterra series is given as an example [37]. The MOS differential under study is shown in Fig. 127. This seemingly simple circuit can not be represented using the circuit model representation given in Fig. 125, so full Volterra expansion has to be employed. The small signal input to the differential pair is assumed to have only differential voltage $v_d$. The input common-mode voltage provides DC bias. The small signal voltages applied to the gates of $M_1$ and $M_2$ are $\frac{v_d}{2}$ and $-\frac{v_d}{2}$ respectively. The tail current source is assumed to be ideal and provides DC bias current $I_{ss}$. The small signal voltage at

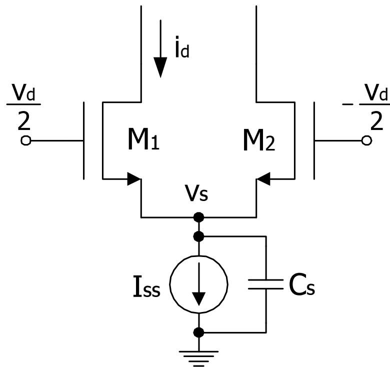



Fig. 127. A MOS differential pair for Volterra analysis

the common-source node is $v_s$. The only memory effect is introduced by the parasitic capacitance $C_s$ at node $v_s$. All other capacitances are ignored for simplicity. The MOS transistors are modeled as square law devices and

$$I_{ss} = K\left(V_{GS} - V_t\right)^2 \tag{B.15}$$

The goal is to obtain the Volterra series expansion for the small signal drain current $i_d$ of $M_1$ (or $M_2$) in terms of input differential voltage $v_d$ up to 3rd order. To this end, the Volterra series of the common-source voltage $v_s$ will be determined first, i.e. $v_s$ will be expanded into

$$\begin{aligned} v_s &= G_1 \circ v_d + G_2 \circ v_d^2 + G_3 \circ v_d^3 + \cdots \\ &= v_{s1} + v_{s2} + v_{s3} + \cdots \end{aligned} \tag{B.16}$$

where $v_{sk} = G_k \circ v_d^k$, is the k-th order term of the Volterra series of $v_s$. Applying KCL at the common-source node

$$C_s \frac{dV_s}{dt} + I_{ss} = \frac{K}{2}\left[\left(V_{gs1} - V_t\right)^2 + \left(V_{gs2} - V_t\right)^2\right] \tag{B.17}$$

where $V_s$, $V_{gs1}$ and $V_{gs2}$ can be written into DC and signal small signal terms

$$V_s = V_S + v_s \tag{B.18}$$

$$V_{gs1} = V_{GS} + v_{gs1} \tag{B.19}$$

$$V_{gs2} = V_{GS} + v_{gs2} \tag{B.20}$$

and further express $v_{gs1}$ and $v_{gs2}$ in terms of differential input voltage $v_d$ and common-source voltage $v_s$

$$v_{gs1} = \frac{v_d}{2} - v_s \tag{B.21}$$

$$v_{gs2} = -\frac{v_d}{2} - v_s \tag{B.22}$$

Substituting from (B.18) to (B.22) into (B.17) and simplifying it,

$$C_s \frac{dv_s}{dt} + 2g_m v_s - K v_s^2 = \frac{K}{4} v_d^2 \tag{B.23}$$

where $g_m = K\left(V_{GS} - V_t\right)$ is the MOS transistor's transconductance. Substituting (B.16) into (B.23) and taking the phasor form of $\frac{dv_s}{dt}$

$$\left(j\omega C_s + 2g_m\right)\left(v_{s1} + v_{s2} + v_{s3} + \cdots\right) - K\left(v_{s1} + v_{s2} + v_{s3} + \cdots\right)^2 = \frac{K}{4} v_d^2 \tag{B.24}$$

The Volterra kernel $G_k$ can be obtained by equating the same order term of $v_d$ at both sides of (B.24). Keeping only the first order terms

$$\left(j\omega C_s + 2g_m\right) G_1\left(\omega\right) \circ v_d = 0 \tag{B.25}$$

Hence

$$G_1\left(\omega\right) = 0 \tag{B.26}$$

This means $v_{s1} = 0$. Substituting it into (B.24)

$$\left(j\omega C_s + 2g_m\right)\left(v_{s2} + v_{s3} + \cdots\right) - K\left(v_{s2} + v_{s3} + \cdots\right)^2 = \frac{K}{4} v_d^2 \tag{B.27}$$

Keeping only the second order terms

$$[j(\omega_1+\omega_2)C_s+2g_m]\,G_2(\omega_1,\,\omega_2)\circ v_d^2=\frac{K}{4}v_d^2 \tag{B.28}$$

Thus

$$G_2(\omega_1,\,\omega_2)=\frac{\frac{K}{4}}{j(\omega_1+\omega_2)C_s+2g_m} \tag{B.29}$$

$C_s$ is parasitic capacitance, so usually $(\omega_1+\omega_2)\,C_s \ll 2g_m$. (B.29) can be simplified by taking its Taylor expansion and only retain the first two terms

$$G_2(\omega_1,\,\omega_2)\approx\frac{K}{8g_m}\left[1-j(\omega_1+\omega_2)\frac{C_s}{2g_m}\right] \tag{B.30}$$

Now factoring out the third order terms from (B.27)

$$[j(\omega_1+\omega_2+\omega_3)C_s+2g_m]\,G_3(\omega_1,\,\omega_2,\,\omega_3)\circ v_d^3=0 \tag{B.31}$$

This means that

$$G_3(\omega_1,\,\omega_2,\,\omega_3)=0 \tag{B.32}$$

Higher order kernels can be calculated by repeating the above procedures.

Now, $v_s$ has been expanded into Volterra series up to the third order. Note that $G_1$ and $G_3$ are all zero, so

$$v_s=v_{s2}+\cdots=G_2(\omega_1,\,\omega_2)\circ v_d^2+\cdots \tag{B.33}$$

Because of the operator $\circ$, (B.33) actually consists of four terms corresponding four different cases: $\omega_1,\,\omega_2=\pm\omega_a,\,\pm\omega_b$.

Remember that the final goal is to obtain the Volterra series expansion for $i_d$ in terms of $v_d$, i.e.

$$\begin{aligned} i_d &= H_1\circ v_d+H_2\circ v_d^2+H_3\circ v_d^3+\cdots \\ &= i_{d1}+i_{d2}+i_{d3}+\cdots \end{aligned} \tag{B.34}$$

Using the MOS device equation, the small-signal output current $i_d$ can be related to the input differential voltage and common-source voltage $v_s$ as

$$i_d = \frac{K}{2}\left(\frac{v_d}{2} - v_s\right)^2 + g_m\left(\frac{v_d}{2} - v_s\right) \tag{B.35}$$

Substituting (B.33) and (B.34) into (B.35)

$$i_{d1} + i_{d2} + i_{d3} + \cdots = \frac{K}{8}v_d^2 + \frac{K}{2}\left(v_{s2} + \cdots\right)^2 - \frac{K}{2}v_d\left(v_{s2} + \cdots\right) + g_m\left(\frac{1}{2}v_d - v_{s2} - \cdots\right) \tag{B.36}$$

Keeping the first order terms in (B.36)

$$i_{d1} = \frac{1}{2}g_m v_d \tag{B.37}$$

Thus

$$H_1(\omega) = \frac{1}{2}g_m \tag{B.38}$$

This is the transconductance of the differential pair for a single-ended output.

Isolating the second order terms in (B.36)

$$i_{d2}(\omega_1, \omega_2) = \frac{K}{8}v_d^2 - g_m G_2(\omega_1, \omega_2) \circ v_d^2 \tag{B.39}$$

So

$$H_2(\omega_1, \omega_2) = \frac{K}{8}\left[1 - \frac{1}{1 + j(\omega_1 + \omega_2)\frac{C_s}{2g_m}}\right] \approx j(\omega_1 + \omega_2)\frac{K}{16}\frac{C_s}{g_m} \tag{B.40}$$

To obtain the third order kernel, separating the 3rd order terms from (B.36)

$$\begin{aligned} i_{d3}(\omega_1, \omega_2, \omega_3) &= -\tfrac{K}{2}v_d v_{s2} = -\tfrac{K}{2}\overline{G_2(\omega_1, \omega_2)} \circ v_d^3 \\ &= -\tfrac{K}{2}\left[\tfrac{G_2(\omega_1,\omega_2)+G_2(\omega_1,\omega_3)+G_2(\omega_2,\omega_3)}{3}\right] \circ v_d^3 \end{aligned} \tag{B.41}$$

The summation of permutations of $G_2$ notated by $\overline{G_2}$ is used to obtain a symmetrical

kernel. Substituting (B.30) into the above equation

$$H_3(\omega_1, \omega_2, \omega_3) \approx -\frac{K^2}{16g_m}\left[1 - j(\omega_1+\omega_2+\omega_3)\frac{C_s}{3g_m}\right] \tag{B.42}$$

This leads to the end of the Volterra series expansion of drain current $i_d$ in terms of differential input $v_s$. Table XXII compares the results obtained from the above analysis with low frequency Taylor series analysis. It is seen that Taylor series does not reveal the high frequency second-order nonlinearity for a differential pair.

Table XXII. Volterra series versus Taylor series

| | Volterra Kernel | Taylor Coefficient |
|---|---|---|
| 1st order | $\frac{1}{2}g_m$ | $\frac{1}{2}g_m$ |
| 2nd order | $j(\omega_1+\omega_2)\frac{K}{16}\frac{C_s}{g_m}$ | 0 |
| 3rd order | $-\frac{K^2}{16g_m}\left[1 - j(\omega_1+\omega_2+\omega_3)\frac{C_s}{3g_m}\right]$ | $-\frac{K^2}{16g_m}$ |

**Another Example: Resistive-Degenerated BJT**

Here the detailed derivation of the Volterra series of a resistive-degenerated bipolar transistor (see Fig. 75) is given. The results were used in Chapter V. We start from the equation (5.67) in Chapter V. This equation is rewritten as following by using $H_k$ instead of $B_k$ as the kernel notation:

$$\begin{aligned} &H_0 + H_1\circ v_i + H_2\circ v_i^2 + H_3\circ v_i^3 + \cdots = \\ &\quad I_Q\Big\{1 + \tfrac{1}{I_t}\left[-H_0 + (g_e - H_1)\circ v_i - H_2\circ v_i^2 - H_3\circ v_i^3 + \cdots\right] \\ &\quad + \tfrac{1}{2I_t^2}\left[-H_0 + (g_e - H_1)\circ v_i - H_2\circ v_i^2 - H_3\circ v_i^3 + \cdots\right]^2 \\ &\quad + \tfrac{1}{6I_t^3}\left[-H_0 + (g_e - H_1)\circ v_i - H_2\circ v_i^2 - H_3\circ v_i^3 + \cdots\right]^3 + \cdots\Big\} \end{aligned} \tag{B.43}$$

In the above expression, the frequency variables of $H_k$ are dropped for conciseness. $H_k$ actually means $H_k(\omega_1, \omega_2, \cdots, \omega_k)$, $k = 0, 1, 2, \cdots$. Also the following equations hold:

$$I_t(\omega) = I_{t0}\left(1 + j\frac{\omega}{\omega_\pi}\right) \tag{B.44}$$

$$I_{t0} = \phi_t g_e \tag{B.45}$$

$$\omega_\pi = \frac{g_e}{C_\pi} \tag{B.46}$$

$$R_e = \frac{1}{g_e} \tag{B.47}$$

The Volterra kernels can be calculated by equating the same order terms from both sides of (B.43). The zero-th order or DC term will be calculated first, then the first order, second order and third order, etc.

To calculate the zero-th order, isolating $H_0$ from both sides of (B.43)

$$\frac{H_0}{I_Q} = 1 - \frac{H_0}{I_{t0}} + \frac{H_0^2}{2I_{t0}^2} - \frac{H_0^3}{6I_{t0}^3} + \cdots = e^{-\frac{H_0}{I_{to}}} \tag{B.48}$$

$\frac{H_0}{I_{t0}}$ and $\frac{I_Q}{I_{t0}}$ are much less than unity values, high order terms in (B.48) can be ignored and

$$H_0 \approx \frac{I_Q}{1 + \frac{I_Q}{I_{to}}} \approx I_Q \tag{B.49}$$

For the 1st order kernel $H_1(\omega)$, we have

$$\begin{aligned} \frac{H_1}{I_Q} &= \frac{1}{I_{t1}}(g_e - H_1) + \frac{1}{I_{t1}^2}\left[-H_0(g_e - H_1)\right] + \frac{1}{2I_{t1}^3}H_0^2(g_e - H_1) + \cdots \\ &= \frac{1}{I_{t1}}(g_e - H_1)\left[1 + \frac{-H_0}{I_{t1}} + \frac{H_0^2}{2I_{t1}^2} + \cdots\right] \\ &= \frac{1}{I_{t1}}(g_e - H_1)\, e^{-\frac{H_0}{I_{t1}}} \end{aligned} \tag{B.50}$$

where $I_{t1} = I_t(\omega) = I_{t0}\left(1 + j\frac{\omega}{\omega_\pi}\right)$. Solving for $H_1$ from (B.50)

$$H_1 = \frac{I_Q g_e e^{-\frac{H_0}{I_{t1}}}}{I_{t1} + I_Q e^{-\frac{H_0}{I_{t1}}}} \tag{B.51}$$

Note that

$$\frac{H_0}{I_{t0}} \approx \frac{I_Q}{I_{t0}} = \frac{I_Q}{\phi_t g_e} = g_m R_e \tag{B.52}$$

For intermediate frequency of operation $\omega \ll \omega_\pi$ and also notice $e^{-\frac{H_0}{I_{t0}}} \approx 1$,

$$e^{-\frac{H_0}{I_{t1}}} = e^{-\frac{H_0}{I_{t0}\left(1+j\frac{\omega}{\omega_\pi}\right)}} \approx e^{-\frac{H_0}{I_{t0}}} \cdot e^{j\frac{H_0}{I_{t0}}\frac{\omega}{\omega_\pi}} \approx e^{j\frac{\omega}{\omega_\pi} g_m R_e} \tag{B.53}$$

Substituting (B.52) and (B.53) into (B.51), an intermediate frequency approximation can be obtained:

$$\begin{aligned} H_1 &\approx \frac{g_m\left(1+j\frac{\omega}{\omega_\pi} g_m R_e\right)}{(1+g_m R_e)+j\frac{\omega}{\omega_\pi}(1+g_m^2 R_e^2)} \\ &\approx \frac{g_m}{1+g_m R_e}\left(1+j\frac{\omega}{\omega_\pi} g_m R_e\right)\left[1-j\frac{\omega}{\omega_\pi}(1-g_m R_e)\right] \\ &\approx \frac{g_m}{1+g_m R_e}\left[1-j\frac{\omega}{\omega_\pi}(1-2g_m R_e)\right] \\ &\approx \frac{g_m}{1+g_m R_e}\left(1-j\frac{\omega}{\omega_\pi}\right) \end{aligned} \tag{B.54}$$

For DC and low frequency operations $\frac{\omega}{\omega_\pi} \sim 0$,

$$H_1 \approx \frac{g_m}{1+g_m R_e} \tag{B.55}$$

This is the same as using Taylor series expansion for the circuit's transconductance.

The 2nd order kernel $H_2(\omega_1, \omega_2)$ satisfies

$$\begin{aligned} \frac{H_2}{I_Q} &= \frac{-H_2}{I_{t2}} + \frac{1}{2I_{t2}^2}\overline{(g_e - H_1)^2} + \frac{1}{I_{t2}^2} H_0 H_2 \\ & \frac{-H_0}{2I_{t2}^3}\overline{(g_e - H_1)^2} + \frac{-H_0^2}{2I_{t2}^3} H_2 + \cdots \end{aligned} \tag{B.56}$$

where $I_{t2} = I_t(\omega_1+\omega_2) = I_{t0}\left(1+j\frac{\omega_1+\omega_2}{\omega_\pi}\right)$, and

$$\overline{(g_e - H_1)^2} = \frac{1}{2}\left\{[g_e - H_1(\omega_1)]^2 + [g_e - H_1(\omega_2)]^2\right\} \tag{B.57}$$

Rearranging the terms in (B.56) yields

$$\left(\frac{I_{t2}}{I_Q} + 1 - \frac{H_0}{I_{t2}} + \frac{1}{2}\frac{H_0^2}{I_{t2}^2} - \cdots\right) H_2 = \frac{1}{2I_{t2}}\overline{(g_e - H_1)^2}\left(1 - \frac{H_0}{I_{t2}} + \cdots\right) \tag{B.58}$$

or equivalently

$$\left(\frac{I_{t2}}{I_Q}+e^{-\frac{H_0}{I_{t2}}}\right)H_2=\frac{1}{2I_{t2}}\overline{(g_e-H_1)^2}e^{-\frac{H_0}{I_{t2}}} \tag{B.59}$$

and solving for $H_2$

$$H_2=\frac{I_Q}{2I_{t2}}\frac{\overline{(g_e-H_1)^2}e^{-\frac{H_0}{I_{t2}}}}{I_{t2}+I_Qe^{-\frac{H_0}{I_{t2}}}} \tag{B.60}$$

An intermediate frequency approximation can be obtained easily by going through a procedure similar to that of $H_1$, and using the $H_1$'s low frequency approximation. The result is given below:

$$H_2\approx\frac{g_m^2}{2I_Q}\frac{1}{(1+g_mR_e)^3}\left[1-j\frac{2(\omega_1+\omega_2)}{\omega_\pi}\right] \tag{B.61}$$

The DC and low frequency approximation is

$$H_2\approx\frac{g_m^2}{2I_Q}\frac{1}{(1+g_mR_e)^3} \tag{B.62}$$

The 3rd order kernel fulfills

$$\begin{aligned}\tfrac{H_3}{I_Q} &= \tfrac{-H_3}{I_{t3}}+\tfrac{1}{I_{t3}^2}H_0H_3-\tfrac{1}{I_{t3}^2}\overline{(g_e-H_1)H_2}\\ &+ \tfrac{1}{6I_{t3}^3}\overline{(g_e-H_1)^3}-\tfrac{1}{2I_{t3}^3}H_0^2H_3+\cdots\end{aligned} \tag{B.63}$$

where $I_{t3}=I_t(\omega_1+\omega_2+\omega_3)=I_{t0}\left(1+j\frac{\omega_1+\omega_2+\omega_3}{\omega_\pi}\right)$, and

$$\begin{aligned}\overline{(g_e-H_1)H_2} &= \tfrac{1}{3}\{[g_e-H_1(\omega_1)]H_2(\omega_2,\omega_3)\\ &\quad+[g_e-H_1(\omega_2)]H_2(\omega_1,\omega_3)\\ &\quad+[g_e-H_1(\omega_3)]H_2(\omega_2,\omega_1)\}\end{aligned} \tag{B.64}$$

$$\overline{(g_e-H_1)^3}=\frac{1}{3}\left\{[g_e-H_1(\omega_1)]^3+[g_e-H_1(\omega_2)]^3+[g_e-H_1(\omega_3)]^3\right\} \tag{B.65}$$

Rearranging the terms in (B.63) yields

$$\left(\frac{I_{t3}}{I_Q}+1-\frac{H_0}{I_{t3}}+\frac{H_0^2}{2I_{t3}^2}-\cdots\right)H_3\approx\frac{1}{6I_{t3}^2}\overline{(g_e-H_1)^3}-\frac{1}{I_{t3}}\overline{(g_e-H_1)H_2} \tag{B.66}$$

or equivalently

$$\left(\frac{I_{t3}}{I_Q} + e^{-\frac{H_0}{I_{t3}}}\right) H_3 \approx \frac{1}{6I_{t3}^2}\overline{(g_e - H_1)^3} - \frac{1}{I_{t3}}\overline{(g_e - H_1)\,H_2} \tag{B.67}$$

Solving for $H_3$

$$H_3 \approx \frac{I_Q}{6I_{t3}^2}\frac{\overline{(g_e - H_1)^3} - 6I_{t3}\overline{(g_e - H_1)\,H_2}}{I_{t3} + I_Q e^{-\frac{H_0}{I_{t3}}}} \tag{B.68}$$

Applying for intermediate frequency approximation $\omega \ll \omega_\pi$, and using low frequency approximation for $H_1$ and $H_2$

$$\begin{aligned} H_3 &\approx \tfrac{g_m^3}{6I_Q^2(1+g_mR_e)^5}\,(1 - 2g_mR_e)\left[1 - j\tfrac{\omega_1+\omega_2+\omega_3}{\omega_\pi}\,(3 - g_mR_e)\right] \\ &\approx \tfrac{g_m^3}{6I_Q^2(1+g_mR_e)^5}\,(1 - 2g_mR_e)\left[1 - j\tfrac{3(\omega_1+\omega_2+\omega_3)}{\omega_\pi}\right] \end{aligned} \tag{B.69}$$

A low frequency approximation is given by dropping the frequency term in (B.69)

$$H_3 \approx \frac{g_m^3}{6I_Q^2\,(1 + g_mR_e)^5}\,(1 - 2g_mR_e) \tag{B.70}$$

Higher order kernels can be calculated by repeating the above procedures. For quick reference, the results in this example are summarized in Table XXIII.

Table XXIII. Volterra kernels of degenerated BJT

| | Exact expression | Intermediate frequency |
|---|---|---|
| $H_0$ | $I_Q$ | $I_Q$ |
| $H_1$ | $\frac{I_Q g_e e^{-\frac{H_0}{I_{t1}}}}{I_{t1}+I_Q e^{-\frac{H_0}{I_{t1}}}}$ | $\frac{g_m}{1+g_mR_e}\left(1 - j\frac{\omega}{\omega_\pi}\right)$ |
| $H_2$ | $\frac{I_Q}{2I_{t2}}\frac{\overline{(g_e-H_1)^2 e^{-\frac{H_0}{I_{t2}}}}}{I_{t2}+I_Q e^{-\frac{H_0}{I_{t2}}}}$ | $\frac{g_m^2}{2I_Q}\frac{1}{(1+g_mR_e)^3}\left[1 - j\frac{2(\omega_1+\omega_2)}{\omega_\pi}\right]$ |
| $H_3$ | $\frac{I_Q}{6I_{t3}^2}\frac{\overline{(g_e-H_1)^3}-6I_{t3}\overline{(g_e-H_1)H_2}}{I_{t3}+I_Q e^{-\frac{H_0}{I_{t3}}}}$ | $\frac{g_m^3}{6I_Q^2(1+g_mR_e)^5}\,(1 - 2g_mR_e)\left[1 - j\frac{3(\omega_1+\omega_2+\omega_3)}{\omega_\pi}\right]$ |

# VITA

Chunyu Xin was born in Tianjin, China. He received his B.S. and M.S. degrees from Nankai University, Tianjin, China in 1997 and 1999 respectively. He is presently working toward his Ph.D. degree under the supervision of Dr. Edgar Sánchez-Sinencio in the Analog and Mixed Signal Center, Texas A&M University. From September 2001 to August 2002 he worked for Motorola in Austin, TX for his internship. His research interests are focused on integrated radio frequency circuits design for wireless receivers. He held a Motorola fellowship in 2000. He has been an IEEE student member since 2001. He can be reached through the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843.

The typist for this dissertation was Chunyu Xin.