

ON-WAFER S-PARAMETER MEASUREMENT USING FOUR-PORT TECHNIQUE AND
INTERMODULATION LINEARITY OF RF CMOS

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

Xiaoyun Wei

Certificated of approval:

Richard C. Jaeger
Distinguished University Professor
Electrical and Computer Engineering

Guofu Niu, Chair
Alumni Professor
Electrical and Computer Engineering

Fa Foster Dai
Professor
Electrical and Computer Engineering

Stuart Wentworth
Associate Professor
Electrical and Computer Engineering

George T. Flowers
Dean
Graduate School

ON-WAFER S-PARAMETER MEASUREMENT USING FOUR-PORT TECHNIQUE AND
INTERMODULATION LINEARITY OF RF CMOS

Xiaoyun Wei

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
December 19, 2008

ON-WAFER S-PARAMETER MEASUREMENT USING FOUR-PORT TECHNIQUE AND
INTERMODULATION LINEARITY OF RF CMOS

Xiaoyun Wei

Permission is granted to Auburn University to make copies of this dissertation at its discretion, upon request of individuals or institutions and at their expense.
The author reserves all publication rights.

Signature of Author

Date of Graduation

VITA

Xiaoyun Wei, daughter of Xinli Wei and Xinye Wang, spouse of Tong Zhang, was born on December 15th, 1978, in Xinxian, Henan Province, P.R. China. She received her BS degree from Huazhong University of Science and Technology in 2000, majoring in Communication Engineering. She received her MS degree from Huazhong University of Science and Technology in 2003, majoring in Circuit and Signal. In Fall 2004, She was accepted into the Electrical and Computer Engineering Department of Auburn University, Auburn, Alabama, where she has pursued her Ph.D. degree.

DISSERTATION ABSTRACT

ON-WAFER S-PARAMETER MEASUREMENT USING FOUR-PORT TECHNIQUE AND
INTERMODULATION LINEARITY OF RF CMOS

Xiaoyun Wei

Doctor of Philosophy, December 19, 2008
(M.S., Huazhong University of Science and Technology, 2003)
(B.S., Huazhong University of Science and Technology, 2000)

236 Typed Pages

Directed by Guofu Niu

Accurate on-wafer characterization of CMOS transistors at extremely high frequencies, e.g. above 60GHz, becomes critical for RFIC designs and CMOS technology development for millimeter wave applications. Traditional two-step error calibration lumps the linear systematic errors as a four-port error adaptor between the perfect VNA receivers and the probe tips, and the distributive on-wafer parasitics as equivalent circuits with shunt and series elements. However, the distributive nature of on-wafer parasitics becomes significant, and the lumped equivalent circuits fail at frequencies above 50GHz.

The distributive on-wafer parasitics is essentially a four-port network between the probe tips and the transistor terminals. This dissertation develops two general four-port techniques that can solve the on-wafer parasitics four-port network, and demonstrates their utility on a 0.13 μ m RF CMOS technology. One is an analytical solution solving

the Y-parameters of the four-port parasitics network. The other one is a numerical solution solving the T-parameters of the four-port parasitics network. Even though the two four-port solutions are developed for on-wafer parasitics de-embedding at the very beginning, the two solutions do not make any reciprocal and symmetric assumptions of the solved four-port network, and can be used for single-step calibration which solves the four-port network between perfect VNA receivers and transistor terminals. In this case, both systematic errors and on-wafer parasitics are included in one four-port network, and can be removed in a single step. With switch error removed, single-step calibration can provide as accurate results as two-step calibration from 2-110GHz.

Another topic that draws the attention of RFIC designers is the linearity (nonlinearity) of CMOS transistors. Experimental IP3 results on a 90nm RF CMOS technology are presented at different biasing voltages, different device width, and different fundamental frequencies. To understand the biasing, device width, and frequency dependence of IP3, a complete IP3 expression is developed using Volterra series analysis and nonlinear current source method. The investigation indicates that not only the 2nd and 3rd order nonlinear output conductance but also the cross terms are important for IP3 sweet spot and high V_{GS} IP3 modeling. Guidelines to identify the IP3 sweet spot for large devices used in RFIC designs are provided.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. Guofu Niu. Without him, this dissertation would not have been possible. His patience and encouragement carried me on through difficult times. I appreciate his vast knowledge and skill in many areas, and his valuable feedback that greatly contributed to this dissertation. I would like to thank the other members of my committee, Dr. Foster Dai, Dr. Stuart M. Wentworth, and Dr. Richard C. Jaeger for the assistance they provided.

Several people deserve special recognition for their contributions to this work. I would like to thank Qingqing Liang for his help with de-embedding techniques. I would like to thank Ying Li for her help with device measurement. I would like to thank Susan Sweeney of IBM Microelectronics for her help with 110GHz S-parameter measurement and noise measurement. I would like to thank Dr. Stewart S. Taylor of Intel Corporation for discussions on device and circuit characterization.

Finally, I am forever indebted to my parents for the support they provided me through my entire life. In particular, I must acknowledge my husband and best friend, Tong Zhang. Without whose love and encouragement, I would not have finished this dissertation.

In conclusion, I recognize that this research would not have been possible without the financial assistance of the Intel Corporation and Vodafone-US foundation. I also would like to thank IBM Microelectronics for fabrication on various technologies.

Style manual or journal used: Transactions of the Institute of Electrical and Electronics Engineers

Computer software used: The document was prepared using Microsoft Word, The plots were generated using Microsoft Visio and Matlab. The Bibliography was generated using EndNote.

TABLE OF CONTENTS

TABLE OF CONTENTS	IX
LIST OF FIGURES	XII
LIST OF TABLES	XIX
CHAPTER 1 INTRODUCTION	1
1.1 Scattering parameter measurement	3
1.2 Intermodulation linearity measurement	10
1.3 Motivation and objectives	14
1.3.1 High-frequency RF CMOS characterization	14
1.3.2 Four-port network for on-wafer parasitics	15
1.3.3 General four-port solution	17
1.3.4 Single-step calibration	19
1.3.5 Validity of BSIM4 model for nonlinear RF modeling	21
1.3.6 Third order intercept point modeling	21
1.3.7 Third order intermodulation distortion characterization	22
1.4 Outline of Contributions	23
CHAPTER 2 ON-WAFER TEST STRUCTURE	25
2.1 Typical on-wafer transistor test structure	26
2.2 Probing pad design considerations	28
2.3 CMOS transistor design considerations	32
2.3.1 Gate pattern and multiplier factor	35
2.3.2 Gate finger configuration	39
2.4 Summary	41
CHAPTER 3 ERROR MODELS FOR TWO-PORT S-PARAMETER MEASUREMENT	42
3.1 Two-port S-parameter measurement	44
3.2 Error adaptor concept	49
3.3 The simplest 8-term error model	50
3.4 The classical 12-term error model	51
3.4.1 Forward mode	52
3.4.2 Reverse mode	53
3.4.3 12-term model	55
3.4.4 SOLT calibration	58
3.5 The most complete 16-term error model	61
3.6 Error adaptor for single-step calibration	66
3.7 Summary	70

CHAPTER 4	GENERIC ANALYTICAL FOUR-PORT SOLUTION.....	71
4.1	Four-port network in Y-parameters	73
4.2	General four-port Solution.....	74
4.2.1	Relationship between open-short and four-port.....	74
4.2.2	Open-short de-embedded LEFT, RIGHT, and THRU.....	77
4.2.3	Analytical solution of A and B	78
4.2.4	Summary of general four-port de-embedding.....	83
4.2.5	Impact of non-ideal load in LEFT and RIGHT	84
4.2.6	Quantifying errors of open-short	86
4.2.7	Reciprocity and symmetry of the four-port parasitics	87
4.3	Reciprocal four-port solution and pad-open-short.....	88
4.4	Summary	90
CHAPTER 5	NUMERICAL FOUR-PORT SOLUTION.....	92
5.1	Four-port parasitic network in T-parameters	93
5.2	SVD based four-port Solution	94
5.3	Experimental results for on-wafer parasitics de-embedding	97
5.4	Reduction of Error Terms and Number of Standards	100
5.4.1	Quantify error terms for four-port on-wafer parasitics	100
5.4.2	8-term solution using three on-wafer standards.....	102
5.5	Summary	106
CHAPTER 6	SINGLE-STEP CALIBRATION	107
6.1	Analytical four-port single-step calibration.....	108
6.2	Numerical four-port single-step calibration.....	111
6.3	Impact of switch errors	113
6.3.1	Quantify error terms using S-parameters.....	115
6.4	Summary	117
CHAPTER 7	VALIDITY OF BSIM4 MODEL FOR NONLINEAR RF MODELING	118
7.1	Linearity measurement and simulation.....	119
7.2	DC and linear characteristics	121
7.3	Nonlinear characteristics.....	125
7.4	Summary	131
CHAPTER 8	MODELING OF INTERMODULATION LINEARITY	132
8.1	First order IP3 theory	133
8.2	Complete IP3 expression	135
8.2.1	Two dimension nonlinear current source.....	136
8.2.2	Input IP3 expression	138
8.3	Impact of the additional terms	140
8.4	Device width scaling.....	142
8.5	DIBL effect	144
8.6	Summary	147
CHAPTER 9	CHARACTERIZATION OF RF INTERMODULATION LINEARITY	149
9.1	Power gain measurement	150

9.2	Linearity Characteristics	151
9.2.1	Drain voltage dependence	151
9.2.2	Finger number dependence	153
9.2.3	Frequency dependence	154
9.2.4	Large signal linearity	157
9.3	Summary	158
BIBLIOGRAPHY		160
APPENDICES		
APPENDIX A ABBREVIATIONS AND SYMBOLS		168
A.1	Abbreviations	168
A.2	Matrix symbols and matrix index	168
APPENDIX B TWO PORT NETWORK REPRESENTATIONS		170
APPENDIX C REVIEW OF ON-WAFER DE-EMBEDDING METHODS		172
C.1	Open-Short de-embedding	172
C.2	Pad-open-Short de-embedding	174
C.3	Three-step de-embedding	176
C.4	Transmission line de-embedding	178
APPENDIX D SWITCH ERROR REMOVAL		182
D.1	Switch error removal equations	182
D.2	Step-by-step guide to measure the switch errors	184
APPENDIX E CALIBRATION KIT SETUP		188
APPENDIX F THE RELATIONSHIP BETWEEN OPEN-SHORT AND FOUR-PORT		192
APPENDIX G SINGULARITY OF LINEAR EQUATION SET		196
G.1	Typical calibration standards	196
G.2	Singularity of on-wafer standards	197
APPENDIX H ONE-PORT ERROR CORRECTION		201
H.1	Error adaptor for one-port system	202
H.2	Relationship between Γ^M and Γ^{DUT}	203
H.3	A generalized interpretation	204
APPENDIX I DERIVATION OF FIRST ORDER INPUT IP3		207
APPENDIX J DERIVATION OF INPUT IP3 BASED ON VOLTERRA SERIES		210
J.1	First order kernels	211
J.2	Second order kernels	212
J.3	Third order kernels	214
J.4	Input IP3	215

LIST OF FIGURES

Fig. 1.1	The power spectrum at the drain of a single transistor under a two-tone excitation, measured by a 50 Ω spectrum analyzer.	3
Fig. 1.2	A typical two-port system for on-wafer S-parameter measurement.	4
Fig. 1.3	On-wafer parasitics and reference planes for system error calibration and on-wafer parasitics de-embedding.	5
Fig. 1.4	The lumped equivalent circuits for (a) open-short, (b) pad-open-short, and (c) three-step de-embedding.	7
Fig. 1.5	The equivalent input resistance and capacitance extracted from open-short, pad-open-short, and improved three-step de-embedded results.	9
Fig. 1.6	An on-wafer intermodulation linearity measurement system.	11
Fig. 1.7	The fundamental and IM3 output products versus input power for a two-tone excited system.	12
Fig. 1.8	IIP3 versus V_{GS} from first order IP3 theory, linearity simulation, and two-tone measurement.	14
Fig. 1.9	(a) The four-port network for systematic errors. (b) The four-port network for on-wafer parasitics.	17
Fig. 2.1	The top view of an on-wafer test structure for transistors. (a) The whole test structure including probing pads. (b) The MOS transistor under test only. The dimension is not to scale.	27
Fig. 2.2.	Chip pictures of the fabricated transistor structures on three RF CMOS technologies. (b) and (c) are fabricated at different foundries.	28
Fig. 2.3	Cross section view of an advanced RF CMOS technology. The dimension is to scale.	29
Fig. 2.4	The cross section view of GSG pads and MOS transistor along three cuts.	31
Fig. 2.5.	Layout for one cell of the desired transistor.	32
Fig. 2.6.	Layout for the desired transistor, NMOS, and the on-wafer standards, OPEN, SHORT, LEFT, RIGHT and THRU.	34
Fig. 2.7.	Layout for NMOS transistors with different gate patterns and multiplier factors.	36
Fig. 2.8.	An example for f_T extraction.	37

Fig. 2.9.	An example for f_{\max} extraction.....	37
Fig. 2.10.	Extracted parameters for three NMOS transistors with different gate patterns and multiplier factors.....	39
Fig. 2.11.	Layout for three NMOS transistors with same total channel width but different finger width and finger number. $W_{\text{total}}=40\mu\text{m}$	40
Fig. 2.12.	Extracted parameters for three NMOS transistors with same total channel width but different finger width and finger number. $W_{\text{total}}=40\mu\text{m}$	41
Fig. 3.1	Block diagram for two-port S-parameter measurement using Agilent 8510C system.	45
Fig. 3.2	A two-port VNA system with four receivers.	46
Fig. 3.3	The magnitude of the measured S_{11} of an ideal (a) LOAD and (b) SHORT.....	49
Fig. 3.4	The four-port system error adaptor for two-port S-parameter measurement.....	50
Fig. 3.5	Signal flow graph of 8-term error model for a two-port system.	51
Fig. 3.6	The modified 10-term error model with two leakage errors added.....	51
Fig. 3.7	A two-port S-parameter measurement system configured for forward mode.	52
Fig. 3.8	Forward mode signal flow graph for two-port system including non-ideal Z_0 termination.	53
Fig. 3.9	Simplified forward mode signal flow graph.....	53
Fig. 3.10	A two-port S-parameter measurement system configured for reverse mode.	54
Fig. 3.11	Reverse mode signal flow graph for two-port system including non-ideal Z_0 termination.	54
Fig. 3.12	Simplified reverse mode signal flow graph for two-port system.	54
Fig. 3.13	Forward mode signal flow graph for two-port system.	56
Fig. 3.14	Normalized 6-term error model for forward mode.....	56
Fig. 3.15	Normalized 6-term error model for reverse mode.....	57
Fig. 3.16	(a) OPEN, (b) SHORT, (c) LOAD, and (d) THRU standards for SOLT calibration on Cascade ISS 101-190.....	59
Fig. 3.17	Raw and corrected data for S_{11} and S_{21} of a $0.13\mu\text{m}$ NMOS transistor.	61
Fig. 3.18	Signal flow graph of the 16-term model for a two-port system.	62

Fig. 3.19	The combined four-port network including system errors and on-wafer parasitics.	67
Fig. 4.1.	Block diagram of the on-wafer parasitics four-port network using I-V representation.	73
Fig. 4.2.	The equivalent two-port network of the intrinsic NMOS transistor and the five on-wafer standards OPEN, SHORT, LEFT, RIGHT and THRU.	75
Fig. 4.3.	The real part of four-port de-embedded y_{21} using different a_{21}/a_{11} choices. The a_{21}/a_{11} defined from M_{22}/M_{12} is clearly noisier, and should not be used.	82
Fig. 4.4.	The four-port de-embedded transistor Y-parameters with and without including parasitic capacitance in Y_L and Y_R . For comparison, open-short de-embedded results are also plotted. No reciprocal assumptions are made for four-port parasitics de-embedding.	85
Fig. 4.5.	Effective gate resistance and capacitance extracted from four-port de-embedded results with and without parasitic capacitance included in LEFT and RIGHT. Open-short de-embedded results are also shown for comparison. No reciprocal assumptions are made.	86
Fig. 4.6.	The elements of A' and B' versus frequency.	87
Fig. 4.7.	Reciprocal four-port de-embedded transistor Y-parameters versus the results using open-short and pad-open-short de-embedding.	90
Fig. 5.1	The four-port error adaptor for on-wafer parasitics in wave representation.	93
Fig. 5.2	Comparison of Y-parameters between open-short, pad-open-short, SVD based numerical four-port solution, and analytical four-port solution.	98
Fig. 5.3	Comparison of effective gate resistance and capacitance between open-short, pad-open-short, SVD based four-port solution, and analytical four-port solution.	99
Fig. 5.4	The magnitude of the S-parameters for the four-port on-wafer parasitics.	102
Fig. 5.5	Comparison of Y-parameters between open-short, SVD based 16-term solution, and SVD based 8-term solution.	105
Fig. 5.6	Comparison of Y-parameters between open-short, SVD based 16-term solution, and SVD based 8-term solution.	105
Fig. 6.1.	Single-step versus two-step four-port using the analytical four-port solution with data measured using a HP 8510XF system from 2 GHz to 110 GHz.	110

Fig. 6.2.	Single-step versus two-step four-port using the analytical four-port solution with data measured using a HP 8510C system from 2 GHz to 26.5 GHz.....	110
Fig. 6.3	Comparison between two-step open-short and four-port on-wafer parasitics de-embedding results with ISS calibration and single-step four-port calibration results without any ISS calibration.	112
Fig. 6.4	Condition numbers of the coefficient matrix in on-wafer parasitics de-embedding and single-step calibration.	113
Fig. 6.5	Comparison of the single-step four-port calibrated results with and without switching error correction. The analytical four-port solution in Section 4.2 is applied.....	114
Fig. 6.6	Comparison of the single-step four-port calibrated results with and without switching error correction. The SVD based numerical four-port solution in Section 5.2 is applied.....	115
Fig. 6.7	The magnitude of the solved 16 error terms of the combined four-port network.	116
Fig. 7.1	Block diagram for two-tone intermodulation linearity measurement.	119
Fig. 7.2	Schematic for two-tone intermodulation linearity simulation in Cadence.	120
Fig. 7.3	Measured and simulated $I_{DS} - V_{GS}$ for $V_{DS} = 0.6, 0.8, \text{ and } 1.0\text{V}$	122
Fig. 7.4	Measured and simulated $I_{DS} - V_{DS}$ for $V_{GS} = 0.4\text{V}$ and 0.8V	122
Fig. 7.5	(a) S_{21} in dB versus frequency at $V_{GS} = 0.4\text{V}$ and $V_{DS} = 1.0\text{V}$. (b) S_{21} in dB versus V_{GS} at 5GHz and $V_{DS} = 1.0\text{V}$	123
Fig. 7.6	f_T extracted from measured and simulated S-parameters.	123
Fig. 7.7	Y-parameters versus frequency at $V_{GS} = 0.4\text{V}$ and $V_{DS} = 1.0\text{V}$. \Re and \Im stand for real and imaginary parts.	124
Fig. 7.8	Y-parameters at 5GHz versus V_{GS} . $V_{DS} = 1.0\text{V}$. \Re and \Im stand for real and imaginary parts.	125
Fig. 7.9	The amplitude of the fundamental output signal versus input power level at $V_{GS} = 0.4\text{V}$, $V_{DS} = 0.8\text{V}$	126
Fig. 7.10	The amplitude of the fundamental output signal and the third order intermodulation product versus J_{DS}	127
Fig. 7.11	Measured and simulated IIP3 versus V_{GS} at multiple V_{DS}	128

Fig. 7.12	Measured and simulated IIP3 versus J_{DS} for devices with $N_f=10, 20$, and 64.	129
Fig. 7.13	Measured and simulated IIP3 versus V_{GS} at multiple frequencies for $N_f=10$ ($W=20\mu m$).	130
Fig. 7.14	Measured and simulated IIP3 versus V_{GS} at multiple frequencies for $N_f=64$ ($W=128\mu m$).	131
Fig. 8.1	The small signal equivalent circuit used for IP3 analysis.	134
Fig. 8.2	First order IP3 with a sweet spot at $K_{3g_m}=0$	135
Fig. 8.3	The nonlinear coefficients versus V_{GS}	137
Fig. 8.4	(a) The denominator in (8.5) versus V_{GS} . (b) Each term in the denominator of (8.5) versus V_{GS} . $V_{DS}=0.8V$	141
Fig. 8.5	IIP3 versus V_{GS} from simulation, first order IP3 expression in (8.3), and complete IP3 expression in (8.5) with different nonlinearities included. $V_{DS}=0.8V$	142
Fig. 8.6	IIP3 calculated using (8.5) and (8.3) versus J_{DS} for devices with multiple finger numbers.	143
Fig. 8.7	(a) I_{DS} , (b) K_{3g_m} versus V_{GS} at multiple V_{DS} for simulation with and without V_{th} shift due to ΔV_{th} (DIBL).	145
Fig. 8.8	IIP3 calculated using (8.5) versus V_{GS} at multiple V_{DS} for simulation with and without ΔV_{th} (DIBL).	145
Fig. 8.9	(a) $K_{3g_m} / g_m + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$, (b) K_{3g_m} / g_m , and (c) $\Delta_1 + \Delta_2$ versus V_{GS} at multiple V_{DS} for Cadence simulation with and without ΔV_{th} (DIBL). ..	147
Fig. 9.1	Gain from linearity measurement ($P_{out,1st} - P_{in}$) and gains-parameter measurement (S_{21}) versus V_{GS}	150
Fig. 9.2	Measured and analytical IIP3 versus V_{GS} at multiple V_{DS} . Analytical IIP3 is calculated using (8.5).	152
Fig. 9.3	K_{3g_m} / g_m and $K_{3g_m} / g_m + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$ versus V_{GS} at multiple V_{DS}	153
Fig. 9.4	Measured and analytical IIP3 versus J_{DS} for devices with $N_f=10, 20$, and 64. Analytical IIP3 is calculated using (8.5).	154

Fig. 9.5	Measured IIP3 versus J_{DS} at multiple frequencies for $N_f=10$ and 64 ($W=20\mu m$ and $128\mu m$).....	155
Fig. 9.6	Analytical IIP3 (a) without C_{gd} and (b) with C_{gd} at multiple frequencies for $N_f=64$ ($W=128\mu m$). Analytical IIP3 without C_{gd} is calculated using (8.5).....	156
Fig. 9.7	Analytical IIP3 with and without C_{gd} at multiple frequencies for $N_f=10$ ($W=20\mu m$). Analytical IIP3 without C_{gd} is calculated using (8.5).	156
Fig. 9.8	The output power amplitude for fundamental and 3 rd order intermodulation products versus input power.	157
Fig. 9.9	Contour of 3 rd order intermodulation output power with sweeping gate bias and input power.....	158
Fig. C.1	Equivalent circuit of on-wafer parasitics for open-short de-embedding. ..	173
Fig. C.2	Equivalent circuits and layouts of (a) OPEN, and (b) SHORT standards.	173
Fig. C.3	Equivalent circuit for pad-open-short de-embedding.....	175
Fig. C.4	Equivalent circuits and layouts of PAD, OPEN and SHORT standards for pad-open-short.	175
Fig. C.5	Equivalent circuit for improved three step de-embedding.	177
Fig. C.6	Equivalent circuits and layouts of OPEN, SHORT1, SHORT2, and THRU standards for improved three step.....	177
Fig. C.7	Equivalent circuit for transmission line de-embedding.....	180
Fig. C.8	Equivalent circuits and layouts of THRU1 and THRU2 for transmission line de-embedding. The length of transmission line is not to scale.	181
Fig. D.1	A two-port S-parameter measurement system with four receivers.	182
Fig. F.1.	Block diagram of the 4-port network for on-wafer parasitics using I-V representation.	193
Fig. G.1	Condition number, minimum and maximum singular value for four standards.	198
Fig. G.2	Condition number, minimum and maximum singular value for five standards.	200
Fig. G.3	Condition number for multiple number of standards.	200
Fig. H.1	The block diagram for a one-port measurement.....	201

Fig. H.2	The combined two-port error adaptor for one-port S-parameter measurement.....	202
Fig. H.3	Signal flow graph of the two-port error adaptor in one-port measurement.....	203
Fig. H.4	The three error terms solved using OPEN, SHORT and LOAD standards.....	204
Fig. I.1	The small signal equivalent circuit used for IP3 analysis.	208
Fig. J.1	The small signal equivalent circuit used for IP3 analysis.	211
Fig. J.2	The linearized equivalent circuit for solving first order kernels.	212
Fig. J.3	The equivalent circuit for solving the second order kernels.....	213
Fig. J.4	The equivalent circuit for solving the third order kernels	215

LIST OF TABLES

Table 8.1	Definition of nonlinearity coefficients of nonlinear drain current.	136
Table B.1	Transformation between two port H, Y, Z, and ABCD representations ...	171
Table E.1	Calibration Kit Coefficients	191
Table G.1	Nonsingular combinations of five two-port calibration standards for 16 term error model. Assuming one standard is a zero length THRU.	199

CHAPTER 1

INTRODUCTION

The growth of wire-line and wireless communication demands RF integrated circuits (RFIC) on CMOS technologies because of the low cost and the eligibility for high volume integration. As well known, the RF section is the biggest challenge in CMOS transceiver designs due to the lack of accurate RF CMOS models. This demands reliable RF measurements, which are mainly done on-wafer with the advent of coplanar probes. The measured data must reflect the intrinsic transistor without the effects of the surrounding environment.

The notable available models for a bulk MOSFET (Metal Oxide Silicon Field Effect Transistor) are BSIM3V3 [1], BSIM4 [2], MODEL 11 [3], PSP [4]. BSIM3V3, BSIM4 are charge-based models, while MODEL 11 and PSP are surface-potential-based models [5] [6]. Usually, a set of DC, CV, and S-parameter measurements are carefully designed to evaluate the performance of a technology, and extract the unknown model parameters [7] [8]. For example, from DC measurement, one can have an idea of the mathematical relationship between the voltages and currents at each terminal, and the operating limits of the transistor, e.g. threshold voltage, breakdown voltage. The accuracy of DC measurement is determined by the DC probes and the equipments. Essential to obtaining a good RF model is the accuracy of on-wafer scattering parameter (S-parameter) measurements. S-parameter measurement gives an

idea of the RF performance of the transistor, e.g. cut-off frequency, power gain. The accuracy of measured S-parameters directly affects high frequency model parameters, e.g. gate-source capacitance. The accuracy of the model determines the time to market of any RFIC designs [1]. The system setup and the techniques to remove errors in S-parameter measurement will be detailed later in Section 1.1.

However, S-parameter describes the RF performance of transistors in linear mode only, because VNA is operated in linear mode, and the measured S-parameters only include small-signal information of the transistor at the excitation frequency [9]. The real-world transistor characteristics are nonlinear that the transistor will generate harmonics and intermodulation products in addition to the stimulus signal [9] [8]. The higher-order harmonics and intermodulation products become apparent when the input power is significant. The 1dB compression point and the two-tone third order intermodulation (IM3) distortion are the most widely used figure of merit to evaluate the linearity of transistors.

For a nonlinear system, the IM3 products are the remixed products when the input signal contains two adjacent channels. Fig. 1.1 illustrates the impact of the IM3 product on the desired signals. The spacing between the two-tone input signals, f_1 and f_2 , is Δf . The two components at $2f_1 - f_2$ and $2f_2 - f_1$, are the IM3 products induced by the nonlinear drain current to gate bias function, which are Δf away from the two-tone signals. Since the frequency step for mobile communication channels ranges from 30KHz to 200KHz, $\Delta f = 100\text{KHz}$ is chosen for the two-tone intermodulation measurement in Fig. 1.1. If the transistor is not very linear, the amplitude of the two IM3 products can be comparable to the amplitude of the desired signals. And thus the information you received can be

way off if the filter's roll-off is not narrow enough. The third order intercept point (IP3) is usually used to quantify the third order intermodulation distortion [10] [11]. The details of IM3 measurement and IP3 extraction are presented in Section 1.2. The 1dB compression point can be simultaneously extracted while extracting IP3.

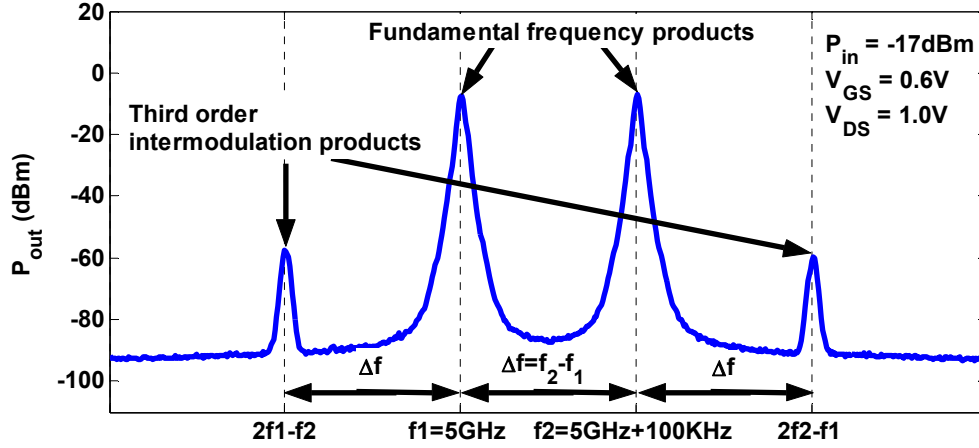


Fig. 1.1 The power spectrum at the drain of a single transistor under a two-tone excitation, measured by a 50Ω spectrum analyzer.

1.1 Scattering parameter measurement

Fig. 1.2 illustrates a typical two-port system for on-wafer S-parameter measurement. It includes a two-port vector network analyzer (VNA), several RF cables and connectors, two RF probes, and a probe station. The Agilent VNA8510C system in Fig. 1.2 consists of four equipments, and can work up to 50GHz with proper configuration. The VNA8510C system is mainly used to measure 26.5GHz and 40GHz S-parameters in this dissertation due to the limitation of RF cables and connectors. The 110GHz data is measured by an Agilent VNA 8510XF system with helps from IBM, Essex Junction. One of the most accurate coplanar ground-signal-ground (GSG) probes,

the Cascade RF infinity probe is used to contact the on-wafer structures. An Alessi manual probe station with a round 6" chuck is used to provide mechanical support and motorization controls of the wafer. Two magnetic positioners are stuck to the metal top plate of the probe station to support the RF probes and provide motorization controls of the probes.

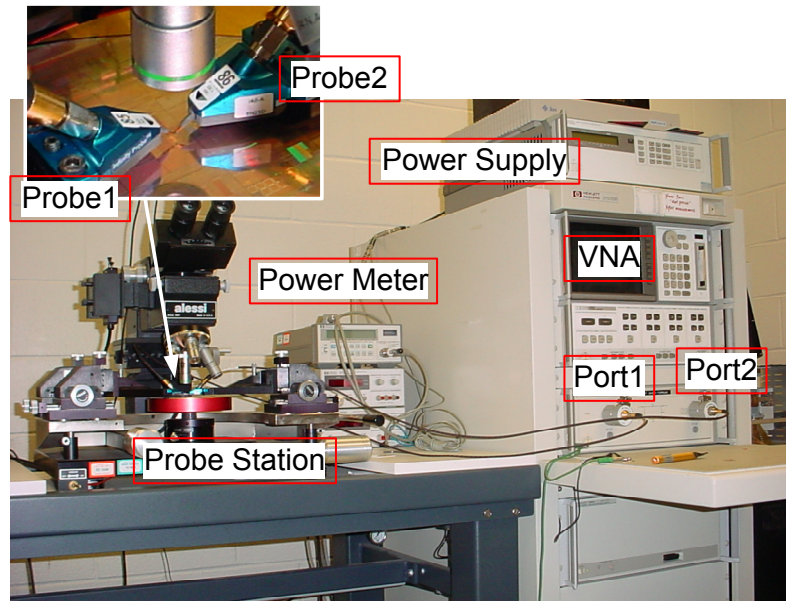


Fig. 1.2 A typical two-port system for on-wafer S-parameter measurement.

However, the system is not perfect. Random and systematic measurement errors are involved in the measured S-parameters [12]. The random errors, e.g. thermal drift, cannot be removed systematically, but the systematic errors can. VNA usually provides several standard techniques for correcting systematic errors, e.g. short-open-load-thru (SOLT). These techniques utilize accurate standards on an impedance standard substrate (ISS) to solve the error terms between the probe tips and the perfect ports inside VNA, a step called “system error calibration.” After system error calibration, the

test system ends at the probe tips, which is then defined as the reference plane for systematic error removal. Reference plane is a factitious separation which defines where the test system ends and the device under test (DUT) begins [13]. Fig. 1.3 illustrates the reference planes defined for on-wafer S-parameter measurement. The reference plane at the probe tips is the reference plan defined for system error calibration.

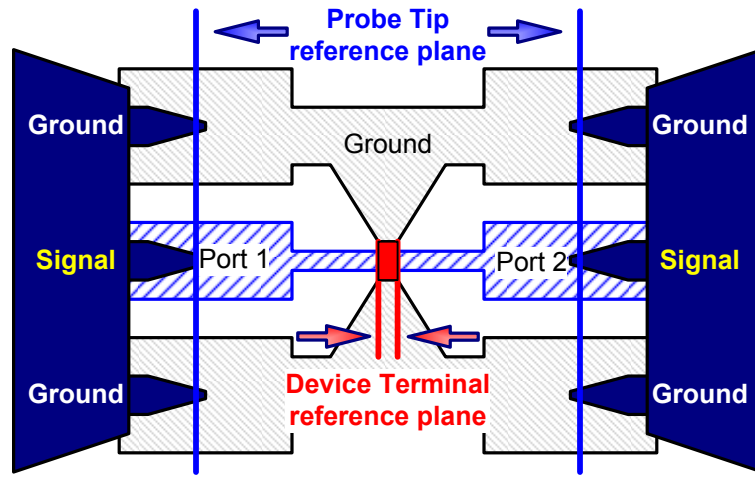


Fig. 1.3 On-wafer parasitics and reference planes for system error calibration and on-wafer parasitics de-embedding.

Besides the systematic errors, on-wafer parasitics including the probing pads and the interconnections need to be removed secondly, a process called “on-wafer de-embedding.” As shown in Fig. 1.3, the probing pads and interconnections often have much larger dimensions when compared with the intrinsic transistor due to the size limitations of RF probes. Thus, a second reference plane is defined at the very end of the interconnections from probing pads to device terminals, which is the device terminal reference plane in Fig. 1.3. The standards used to solve error terms are fabricated on the same wafer as the desired device. The same probing pads and interconnections are

shared by the desired device and the on-wafer standards to keep the reference plane consistent. Since systematic errors and on-wafer parasitics are removed in two steps, this approach is identified as “two-step calibration” in the dissertation. Open-short developed in 1991 lumps on-wafer parasitics as three shunt and three series elements, which is still the industrial standard on-wafer de-embedding technique until now. Fig. 1.4 (a) shows the equivalent circuit for open-short. Two on-wafer standards, an OPEN and a SHORT, are necessary to remove the six lumped elements [14].

Fig. 1.4 (b) and (c) give the equivalent circuits for two alternatives to open-short, three-step and pad-open-short, which make different assumptions of on-wafer parallel parasitics. Open-short assumes that the large probing pads are the only source of parallel parasitics, and thus the three shunt elements are representing the parasitics at the pads [14]. Three-step also lumps the parallel parasitics as three shunt elements, but the third one is between the two series elements instead of the two parallel elements [15] [16]. This assumes that the parasitics between the two pads can be ignored, while the parasitics between the ends of the two interconnect lines are considerable, because of the smaller distance between the two ends when compared with the distance between the two pads. Four on-wafer standards, an OPEN, a THRU, a SHORT1 and a SHORT2, are necessary for three-step de-embedding [15] [16]. Pad-open-short lumps the parallel parasitics at the pads and the interconnect lines separately. Three shunt elements are used to represent the parallel parasitics at the pads, which can be evaluated from a PAD standard without any interconnect lines. The distributive parallel parasitics along the interconnect lines is lumped as three series elements and three shunt elements at the end of interconnect lines. Although, pad-open-short lumps on-wafer parasitics as nine

elements, it only need three standards, a PAD, an OPEN, and a SHORT [17]. Pad-open-short was shown to be better than open-short for on-wafer inductor structures measured above 10GHz. However, this improvement, to a large extent, depends on the layout design [17]. For on-wafer transistor structures, the interconnect lines are not as long and wide as the interconnect lines for the conductor structures in [17], and the parallel parasitics along the interconnect lines is not comparable to the pad parasitics. In this case, pad-open-short will not show great advantage over open-short.

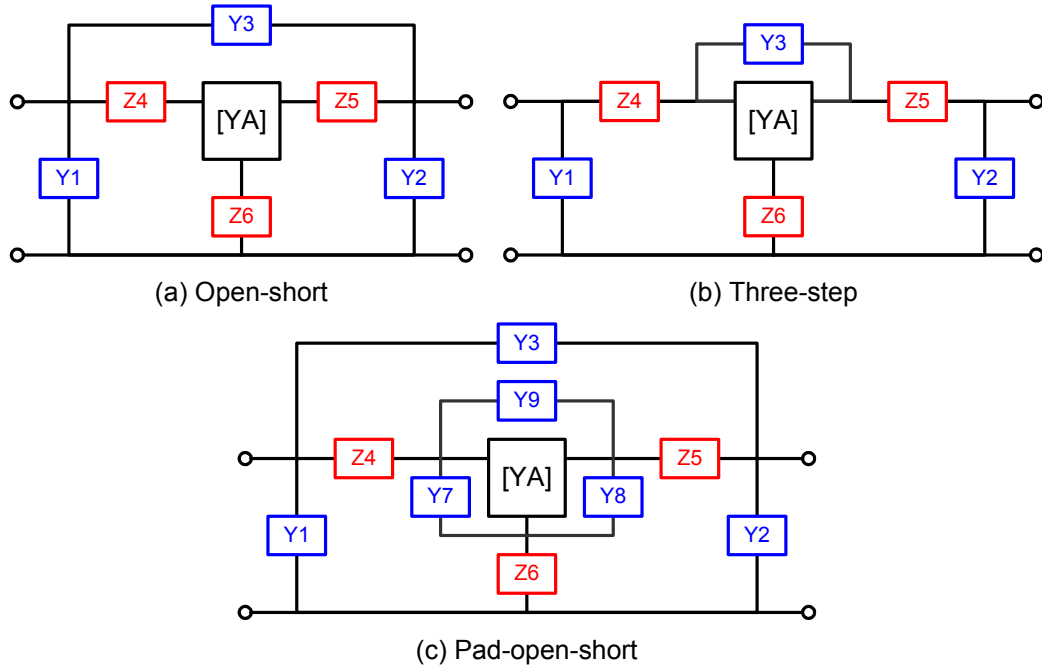


Fig. 1.4 The lumped equivalent circuits for (a) open-short, (b) pad-open-short, and (c) three-step de-embedding.

Fig. 1.5 shows the equivalent input resistance and capacitance, R_{in} and C_{in} , extracted from two-step calibration results [18] [19]. The system errors are calibrated using SOLT, while the on-wafer parasitics are removed using three different techniques,

open-short, pad-open-short, and the improved three-step. The de-embedding procedures are detailed in Appendix C. As compared in Fig. 1.5, the three methods give approximately the same R_{in} and C_{in} for the examined NMOS transistor, and all of them show an unphysical frequency dependence of C_{in} . This indicates that for transistor measurement, these three de-embedding methods all fail at frequencies above 50GHz, even though they are using different lumped equivalent circuit with different complexities. A four-port de-embedding technique, which describes the on-wafer parasitics as a four-port network, was developed in [20] with applications on SiGe HBTs. Advantages over open-short at frequencies above 30GHz were illustrated using simulated results. However the math is complex and no experimental results are presented. Furthermore, pad-open-short was shown to be more accurate than four-port for on-wafer inductor characterization in [17]. These issues need to be examined on CMOS technologies.

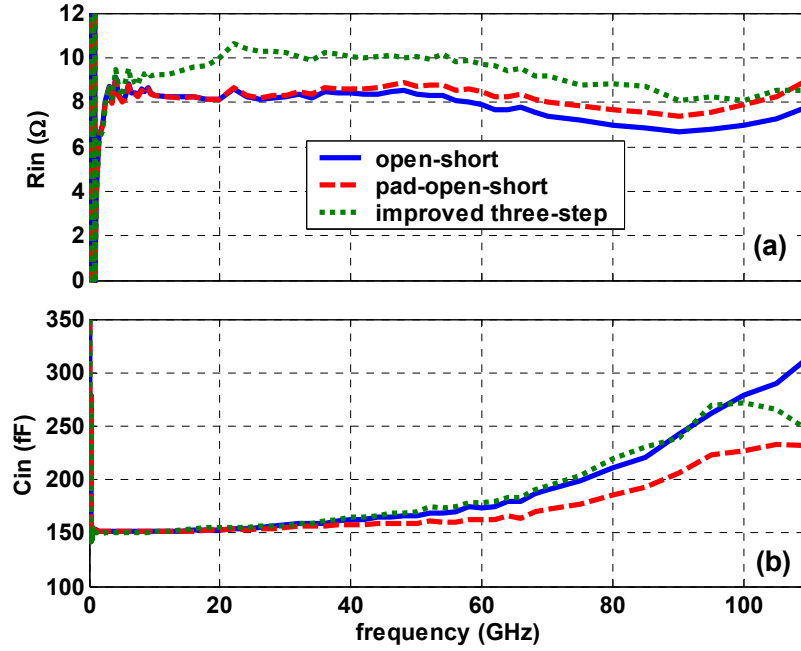


Fig. 1.5 The equivalent input resistance and capacitance extracted from open-short, pad-open-short, and improved three-step de-embedded results.

Two-step calibration can provide the most accurate system error information as long as the ISS standards are accurate. The disadvantage is that the system error calibration step is time consuming and need to be rechecked several times for hourly measurement. Also, two-step calibration involves a process to switch between the ISS substrate and the wafer. Another approach, the so called “single-step calibration”, defines only one reference plane, which is the reference plane at the device terminals. On-wafer standards are used to determining the error terms. The systematic errors and on-wafer parasitics are removed in a single step. The difficulty is that most IC processes cannot deposit a precision resistive load with good repeatability [21]. Due to the less accurate on-wafer standards, single-step calibration are expected to provide less accurate S-parameters when compared with two-step calibration, and thus not widely

used for on-wafer characterization. However, the same on-wafer standards are used for on-wafer de-embedding and these standards are assumed to be ideal for simplicity in two-step calibration. There is no occasion to have a huge difference between two-step calibration and single-step calibration using the same non-ideal n-wafer standards. With appropriate error calibration techniques, single-step calibration may be able to provide reasonably accurate results. This issue should be examined experimentally on advanced silicon technologies.

1.2 Intermodulation linearity measurement

The third order intercept point (IP3) is defined as the point where the 3rd order intermodulation (IM3) product equals the fundamental frequency product for a two-tone excited system. To extract IP3, the power levels of the fundamental and the IM3 products at the output have to be measured using a spectrum analyzer. Fig. 1.6 shows a two-tone intermodulation linearity measurement system with two identical Agilent performance signal generators (PSG) E8247 at the input and an Agilent 8563EC performance spectrum analyzer (PSA) at the output [22]. The signals generated by the two PSGs have the same power level, the same phase, but different frequencies. A power combiner with good isolation is required to combine the two signals. Otherwise, the power combiner itself may produce extra intermodulation products. The products will be amplified by the DUT, which leads to a much larger intermodulation product at the output, and thus introduce undesired errors when extract IP3 of the DUT. Proper attenuators maybe included before the power combiner to provide low enough input power level. DC bias circuits at the input and output are necessary for transistor

linearity characterization. The power spectrum is measured at the output by a PSA and output IP3 (OIP3) is calculated by an Agilent 85672A spurious response utility installed in the PSA. This utility can give not only the amplitude of the fundamental and IM3 products, but also the OIP3 value for the IM3 products.

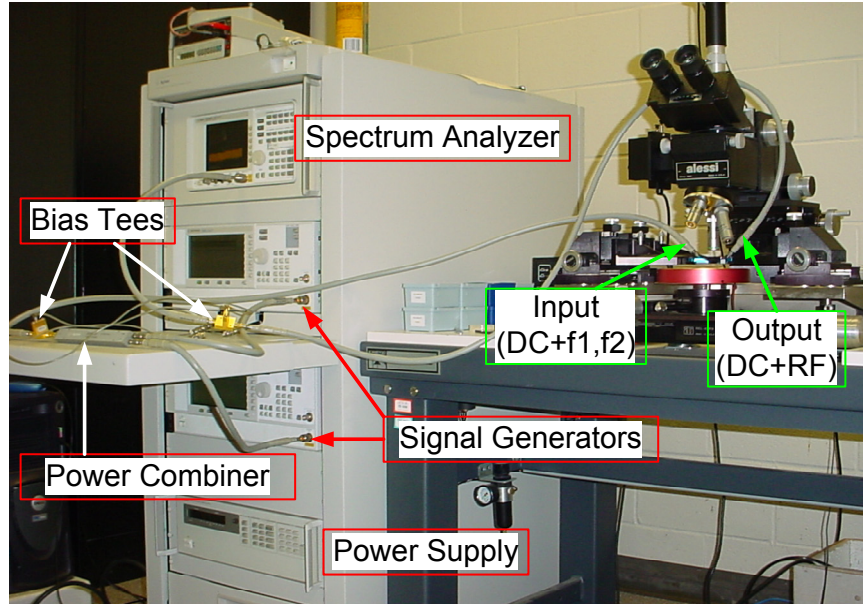


Fig. 1.6 An on-wafer intermodulation linearity measurement system.

Fig. 1.7 shows the fundamental and IM3 output products as a function of input power level P_{in} in dBm for a typical MOS transistor measurement. The solid lines are the measured power values in dBm for the fundamental output product and the IM3 output product, $P_{out,1st}$ and $P_{out,3rd}$. The dash straight lines are linear extrapolations of $P_{out,1st}$ and $P_{out,3rd}$ at a very low reference P_{in} . The reference P_{in} for extrapolation must be well below the 1dB compression point, which is -25dBm in Fig. 1.7. The 1 dB compression point is the input power level where the small signal gain drops by 1 dB, which sets the upper limit for small signal linearity analysis. The intercept point of the

two dash straight lines is the third order intercept point (IP3). The input power level at the IP3 point is IIP3, and the output power level at the IP3 point is OIP3. In Fig. 1.7, the 1dB compression point is -12dBm, IIP3=1.8dBm, OIP3=18dBm, and power gain=16.2 dB.

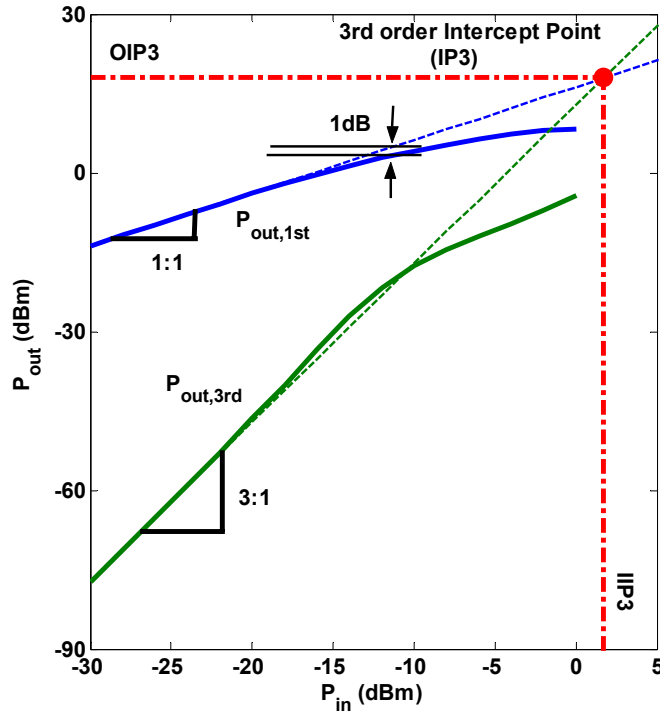


Fig. 1.7 The fundamental and IM3 output products versus input power for a two-tone excited system.

Before extracting IIP3, OIP3 and power gain, the power loss on the input and the output route, including RF cables and connectors, must be calibrated using a power meter. The power loss on the input and the output route must be calibrated using

$$P_{in}|_{dBm} = P'_{in}|_{dBm} - L_{in}|_{dB} \quad (1.1)$$

$$P_{out}|_{dBm} = P'_{out}|_{dBm} + L_{out}|_{dB} \quad (1.2)$$

P'_{in} is the power level generated by the signal generator. P'_{out} is the output power level monitored at the spectrum analyzer. P_{in} is the actual input power level at the gate of the NMOS transistor. P_{out} is the actual output power level at drain terminal of the transistor. L_{in} and L_{out} are the power losses on the input and output routes. L_{in} and L_{out} are frequency dependent, and need to be determined for each frequency before measurement. In practice, L_{in} is much larger than L_{out} , which can lead to a several dB shift on IIP3 and power gain. Relatively speaking, the value of OIP3 is much less sensitive to power calibration.

Instead of using two-tone measurement, IIP3 can also be determined using simulated or measured I-V data and small-signal parameters of the transistor, which just requires DC and S-parameters measurement. For both measurement and simulation, DC and S-parameters are much easier to obtained and much less time consuming. Fig. 1.8 compares first order IP3 with measured and simulated IP3. The derivation of first order IIP3 is detailed in Appendix I. K_{3g_m} is calculated using the 3rd order derivative of I_{DS} with respect to V_{GS} only. The first order IIP3 expression fails in modeling the position of the IP3 sweet spot and the gate voltage dependence of IIP3 in strong inversion region. Analytical IIP3 expressions containing more nonlinearities have been published [11] [22] [23] [24] [25] [26]. However the results are mainly for 0.13 μ m and older technologies, and the MOS model focused is BSIM3V3 [10] [22] [27]. Experimental results on 90nm technology and simulation results using BSIM4 model need to be examined as they become the main stream for RFIC designs.

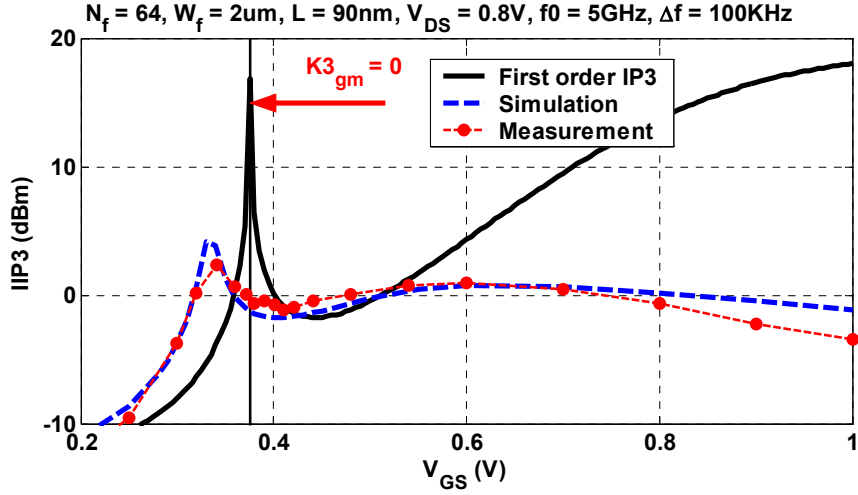


Fig. 1.8 IIP3 versus V_{GS} from first order IP3 theory, linearity simulation, and two-tone measurement.

1.3 Motivation and objectives

1.3.1 High-frequency RF CMOS characterization

Emerging gigabit wire-line and wireless communication applications require integrated circuits operating at frequencies above 60GHz [28] [29] [30] [31]. This demands accurate characterization and modeling of transistors at even higher frequencies. Essential to obtaining a good high-frequency model is the accuracy of the S-parameter measurement. VNA and RF probes capable of 110GHz S-parameter measurements are commercially available over 10 years [32]. However, very few results at such high frequencies are published. This is to a large extent due to the increased difficulty of error calibration for both system errors and on-wafer parasitics.

The industry practice is a two-step approach, which first correct the VNA system errors using well established calibration standards on an impedance standards substrate

(ISS), a process known as system error calibration, and then subtract the on-wafer pads and interconnect lines using on-wafer standards, a process known as on-wafer de-embedding. Short-open-load-thru (SOLT) calibration is one of the system error calibration methods embedded in all modern VNAs, e.g. VNA8510C, and is used in this dissertation where two-step calibration is involved. The de-facto standard technique of on-wafer de-embedding is open-short [14], which however fails for frequencies above 20-40 GHz, depending on layout design and process technology. Various alternatives to open-short have been proposed, including three-step [15], improved three-step [16], four-step [33], and pad-open-short [17]. These methods use more complicated, but still lumped equivalent circuits, and hence require more on-wafer standards. For instance, the three-step methods of [15] and [16] require four on-wafer standards. However, due to the lumped nature of the equivalent circuits used, these methods cannot capture the distributive nature of on-wafer parasitics, and fail above 50 GHz as already shown in Fig. 1.5. For transistor characterization at extremely high frequencies, on-wafer de-embedding methods that can accurately describe the distributive nature of on-wafer parasitics are urgently needed.

1.3.2 Four-port network for on-wafer parasitics

As discussed in Chapter 2, the accuracy of error calibration is determined by the error model, calibration standards, and calibration techniques. A unified 12-term model was developed in 1970s, and became a standard model for two-port VNAs. The SOLT calibration technique is implemented in all modern VNAs to solve the 12 error terms [34]. However, the 12-term error model was shown to be insufficient for high-frequency

measurement, since the leakage errors were modeled using only two error terms in the 12-term model [35] [36]. Same problem exists for error calibration techniques using 8(10)-term model. The most complete error model for two-port system is a 4x4 matrix, a 16-term error model, which is essentially a four-port error network relating four known waves and four unknown waves [35] [36]. Several advanced techniques solving the four-port network have been developed over the years [36] [37] [38] [39]. The 16-term model and the calibration techniques can in general be applied to remove both systematic errors and on-wafer parasitics.

This leads to an idea of describing everything between the probe tips and the device terminals as a four-port network instead of using lumped equivalent circuits [40] [20], an idea that is similar to the 16-term error adaptor in system error correction [36], at least mathematically. Fig. 1.9 (a) illustrates the four-port network for system errors, with two ports inside VNA and two ports at probe tips, which was described as 16-term or 15-term error model frequently [36] [38] [39] [41] [42] [43]. Fig. 1.9 (b) shows the four-port network for on-wafer parasitics with two ports at probe tips and two ports at device terminals, e.g. gate and drain for MOS transistors [20]. Note that all of the a waves are incident waves which entering the four-port network at each port, while all of the b waves are reflected waves which leaving the four-port network at each port. Therefore, the S-parameters of the four-port networks in Fig. 1.9 can be easily defined using the a and b waves. Analytical solutions of the four-port parasitic network were developed in [19] [17] [20] and [44], using three, four, and five on-wafer standards with varying degree of assumptions. For example, with reciprocal assumption, the number of unknowns is reduced to ten and only four on-wafer standards are necessary [45]. With

reciprocal and symmetric assumptions, the number of unknowns is reduced to six and only three on-wafer standards are necessary [17].

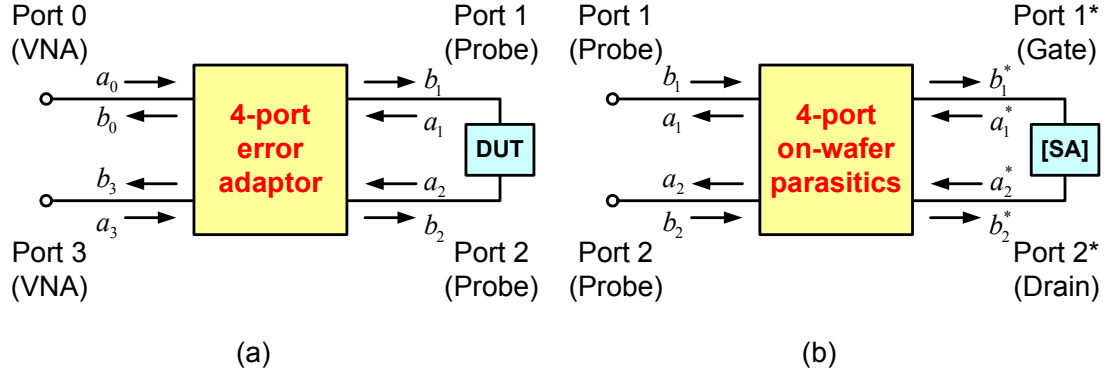


Fig. 1.9 (a) The four-port network for systematic errors. (b) The four-port network for on-wafer parasitics.

1.3.3 General four-port solution

Even though on-wafer parasitics is passive and the associated four-port network should be reciprocal, there are two practical reasons to seek for a solution for generic four-port network, which we will refer to as “general four-port solution.” First, in order to arrive at an analytical solution, a must for real-time fast measurement, on-wafer OPEN and SHORT standards are assumed to be ideal in all of the de-embedding algorithms, while the fabricated standards always have parasitics. In board measurements, inaccuracies of standards are known to lead to nonreciprocal S-parameters for physically passive structure [46]. A general four-port solution will allow us to examine the reciprocity of the four-port parasitics experimentally.

The second reason for seeking a general four-port solution is to directly obtain transistor S-parameters from the measured raw S-parameters without having to perform

system error calibration using ISS. This can result in significant saving in time and effort as ISS calibration is time consuming and needs to be repeated frequently, even during a day of measurement. Also, physical change of substrate is involved. Ideally, the same general four-port solution obtained for on-wafer parasitics de-embedding can be applied to raw S-parameters as is, to remove VNA system errors and on-wafer parasitics in a single step. Not all of the general four-port solutions can be used for single-step calibration. For instance, the solution of [20] can be used, while the solution of [19] cannot be used. The four-port de-embedding algorithms of [20] and [19] make no assumption of the nature of on-wafer parasitics, while the algorithm of [17] assumes that the four-port network for on-wafer parasitics is reciprocal and symmetric.

In this dissertation, two general four-port solutions that can be applied as single-step calibration are developed, 1) a Y-parameter based analytical solution and 2) a singular-value-decomposition (SVD) based numerical solution. With five on-wafer standards, both of them solve a generic four-port network and can be applied on the measured raw S-parameters without ISS calibration. The results were presented in *2007 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems* [19], *2007 IEEE Trans. On Electron Devices* [45] and *2007 IEEE/MTT-S International Microwave Symposium* [47]. The analytical four-port solution in [45] is much simpler than [20] and [19], and considers the parasitic capacitance of the non-ideal on-wafer load resistors. An added advantage of this solution is its intimate relation with open-short, which is then used to quantify the errors left after open-short de-embedding. However, the Y-parameter analytical solutions in [20], [19] and [45] are all limited by specified on-wafer standards and cannot take advantage of the redundancy available from the

measurements of five on-wafer standards, e.g. singularities [36]. These issues are ideally handled with the SVD based numerical solution in [47]. Although the SVD based four-port solution cannot give insight views of the parasitic network, it is easy to apply with multiple combinations of on-wafer standards and provides an indication of the validity of the solution. This dissertation presents detailed derivation of the analytical solution and the numerical solution, and demonstrates their utility on a 0.13 μm RF CMOS technology from 2 to 110GHz for both two-step calibration and single-step calibration.

1.3.4 Single-step calibration

With a general four-port solution, it is possible to solve the four-port network between the two ports inside VNA and the two ports at the device terminals. The known standards are fabricated on the same wafer as the desired device. This idea of utilizing on-wafer standards to remove systematic errors and on-wafer parasitics in a single step was not new. Actually it was introduced at the very beginning of VNA error correction. However, it is not widely used for transistor characterization for several reasons. First, error calibration using ISS standards are repeatable and traceable, which can be verified using stated references. For example, National Institute of Standards and Technology (NIST) in USA and the National Physical Laboratory (NPL) in UK provide and maintain reference standards. By comparing the calibrated VNA results with the reference S-parameters, the performance of VNAs can be verified. With on-wafer standards, the S-parameters of these standards are determined by the technology, which can vary a lot from process to process. It is hard to provide reference standards and

verification kits. Fortunately, the measurement comparison programs (MCP) provide another way to assure measurement accuracy. MCP compares the results of the same device that travel between the participating laboratories to avoid serious errors or provide verification on areas without reference standards. The MCP program illuminate us that the single-step calibrated results can be verified using two-step results for several on-wafer reference standards. Although, ISS calibration is still necessary for verification purpose, it still greatly reduces the measurement time since these reference results just need to be measured once for one wafer. It does not need to be repeated for every test structure.

Secondly, it is hard to accurately model the on-wafer standards. The standards on ISS substrate are modeled using non-ideal capacitance, inductance, and delay time based on physical analysis and verified using reference values. The accuracy of the on-wafer standards affects the accuracy of the error corrected S-parameters. The experimental results in Chapter 6 indicate that assuming ideal on-wafer standards leads to reasonably accurate results in the advance CMOS technology examined. The open capacitance, the short inductance, and the through delay are negligible because of the small dimension of the transistors. The non-ideality of on-wafer load resistor can be modeled using a parasitic capacitor in parallel with a perfect resistor. The experimental results in Chapter 6 indicate that single-step four-port calibrated results are practically identical to the two-step four-port calibrated results after switch error removal.

1.3.5 Validity of BSIM4 model for nonlinear RF modeling

The model parameters extracted from DC, CV, and S-parameters are based on a small-signal schematic, and accurate for small-signal modeling of transistors. For transistor modeling at signals higher than certain value, they do not represent the real transistor performance. In general, the linear model need to be verified using nonlinear simulation [8]. The intermodulation linearity simulation accuracy of the BSIM4 model, a widely used model for RF design, is examined against measurement, particularly in the moderate inversion region, where a linearity sweet spot exists and can be utilized for high linearity RF circuit design [48] [49]. In BSIM4, the moderate inversion region is modeled by mathematical smoothing functions interpolating between physics based approximations in the weak and strong inversion regions, instead of physics based surface potential approximation that can cover all levels of inversion. Its accuracy in linearity simulation, particularly in moderate inversion, therefore needs to be experimentally evaluated, as linearity simulation requires not only accurate modeling of the first order I-V relations, but also higher order derivatives. Note that we do not address simulation of harmonic or intermodulation distortion at $V_{DS}=0V$, a known problem for BSIM4 [50].

1.3.6 Third order intercept point modeling

The nonlinear performance of transistors is typically measured by the 1dB compression point and the third order intercept point (IP3). Using either measured or simulated I-V data, IP3 sweet spot biasing current can be determined from zero K_{3g_m} point based on first order IP3 theory [11] [51]. Circuits have been published to utilize

this zero K_{3g_m} point for high linearity LNA designs [48] [52]. However, experimental IP3 results indicate that the actual IP3 sweet spot V_{GS} is lower than the zero K_{3g_m} V_{GS} by a noticeable amount as already shown in Fig. 1.8 [53]. More accurate analytical IP3 expressions for CMOS devices involving more nonlinearities have been developed recently [22] [25] [26]. The complete IP3 expression developed in this dissertation considers not only transconductance nonlinearities, but also output conductance nonlinearities and cross terms. This expression is used to quantify the impact of these nonlinearities and explain the biasing, device size, and frequency dependence of IP3. Furthermore, guidelines for optimal biasing and sizing for high linearity are developed.

1.3.7 Third order intermodulation distortion characterization

Experimental IP3 results of CMOS devices have been examined using two-tone IP3 measurement [22]. However the results were primarily for 0.13 μ m and older technologies, and the model examined is BSIM3V3 [10] [22] [27]. This dissertation presents experimental characterization of IP3 in a 90nm RF CMOS process, as well as comparing measured IP3 with simulated IP3 using a BSIM4 model. For practical linearity characterization as well as optimal transistor sizing and biasing in circuit design, the linearity is examined as a function of biasing voltages and device sizes. An array of devices with different finger numbers are designed, fabricated and characterized as a function of V_{GS} and V_{DS} , at multiple frequencies. 2GHz, 5GHz, and 10GHz are selected because most current RFIC applications fall in this range. In particular, the sweet spot biasing current for practical large device sizes of interest to RFIC is investigated. The results were presented in *2008 IEEE Radio Frequency*

Integrated Circuits Symposium [53] and the extended paper was accepted by *2008 IEEE Trans. Microwave and Techniques* [54].

1.4 Outline of Contributions

Chapter 1 gives an overview of topics related to on-wafer transistor characterization including linear and nonlinear performance, and gives the motivation of this research. Chapter 2 presents layout details of on-wafer transistors and standards. Carefully designed GSG probing pads, metal ground plane, and shielding structures can help on-wafer parasitics de-embedding and transistor characteristics. Transistors with different gate connection topologies are compared.

The accuracy of S-parameters is determined by error models and correction techniques. Chapter 3 presents the four-port error adaptor concept, the classical 12-term model, and the most complete 16-term model for a two-port system. As a widely used system error calibration method, short-open-load-thru (SOLT) calibration is demonstrated in details. And, the idea of performing single-step calibration is introduced.

Starting from Chapter 4, the concept of four-port error adapter is extended to on-wafer parasitics de-embedding from systematic error calibration. A generic analytical four-port solution for on-wafer parasitics using Y-parameters is developed in Chapter 4. Five specified on-wafer standards, OPEN, SHORT, LEFT, RIGHT, and THRU, are necessary for solving the four-port network. A numerical way to evaluate the errors remaining after open-short de-embedding, and to examine the reciprocity and symmetry of on-wafer parasitics is given using experimental results. Chapter 5 presents a

numerical four-port solution for on-wafer parasitics using singular-value-decomposition (SVD). Although it does not give insight views of on-wafer parasitics, the SVD based solution is easy to apply and gives the most accurate de-embedded results. Although the set of standards can be any non-singular combination of five standards, the same OPEN, SHORT, LEFT, RIGHT, and THRU standards as used in analytical solution is used for comparison. Chapter 6 demonstrates the application of the two four-port solutions in Chapter 4 and Chapter 5 on single-step calibration.

Another topic that draws the attention of circuit designers is the linearity (nonlinearity) of the transistors, which determines upper limit of the spurious dynamic range of transistors or circuits. Chapter 7 evaluates the BSIM4 model for a 90nm RF CMOS technology, which is later used to generate the I-V and small-signal parameters needed to calculate IP3 analytically. Chapter 8 develops a valuable analytical IP3 expression for MOS transistor nonlinearity modeling. The expression is developed based on Volterra series theory using simulated I-V and S-parameters. Biasing, channel width, and frequency dependence of IP3 are well understood using this analytical expression. Chapter 9 compares the calculated IP3 with experimental IP3 for the 90nm RF CMOS technology. Guidelines for optimizing high-linearity applications are given based on experimental results and calculated IP3.

CHAPTER 2

ON-WAFER TEST STRUCTURE

Since a pair of Cascade infinity probes are used to contact the on-wafer test structure, there are several layout rules regarding probe pad placement and sizing that must be followed [55]. Typical contact size of Cascade infinity probes is $12\mu\text{m}\times 12\mu\text{m}$. To achieve reliable contact, it is recommended to further bring the probe down by 50-75 μm after the probe tip has made initial contact with the wafer surface, which leads to a 25-40 μm lateral skating. Thus, the minimum probing area recommended for general use is $50\mu\text{m}\times 50\mu\text{m}$ [55]. And, the minimum center-to-center space between pads is 100 μm . The sizing and spacing requirements for on-wafer probing make it impossible to place the probes directly on the terminals of a modern MOS transistor since the dimension of a typical MOS transistor is only several microns big. Probing pads and interconnect lines leading to the terminals of the transistor are necessary for on-wafer transistor characterization. The GSG probing pads designed for on-wafer characterization are illustrated in Section 2.2.

Ground shield was proved to be able to improve noise performance and on-wafer de-embedding [56] [57] [58]. The first metal layer is used to build the ground shield metal plane as detailed in Section 2.2. It was shown that different gate geometry can affect the DC and RF performance [56] [59] [60] [61]. An array of CMOS transistors with different gate pattern are carefully designed and fabricated. Several parameters that

are critical to RF and noise performance of CMOS transistors are extracted and compared for different gate pattern in Section 2.3.

2.1 Typical on-wafer transistor test structure

Fig. 2.1 (a) is the top view of an on-wafer test structure for a MOS transistor with probing pads and interconnections. GSG probing pads are designed for the GSG Cascade infinity probes, which can shield the signal path between two balanced ground paths and provide tight control on the fields around the signal probe. The dimension of the probing pads and interconnections are much larger than the transistor. Fig. 2.1 (b) gives a closer view of the MOS transistor under test. The four terminal MOS transistor is connected as a two-port system with source and substrate tied together to ground. The MOS transistor in general has multiple gate fingers to reduce gate resistance and a substrate ring around the whole active area to provide better shielding from adjacent structures. The channel width of MOS transistors can be modified by either changing the width of each finger or changing the number of fingers.

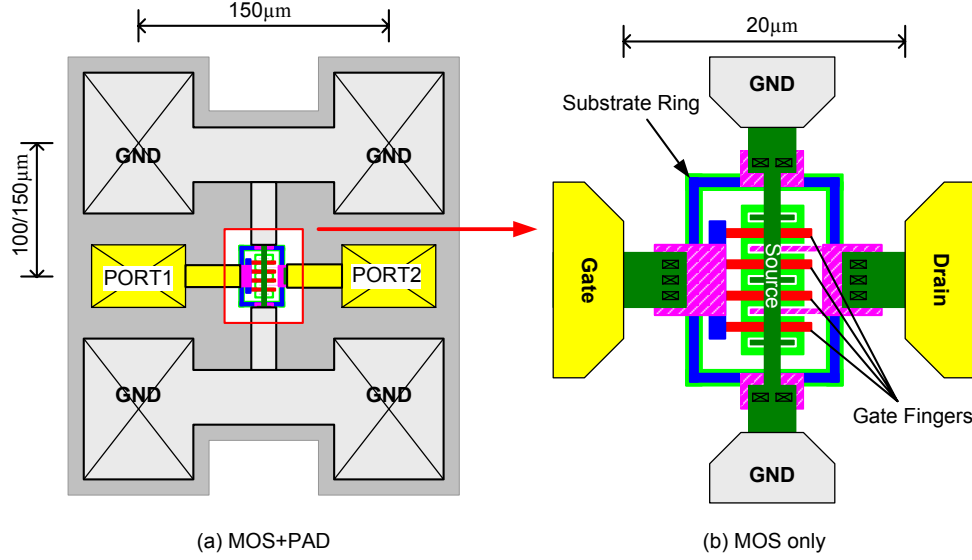
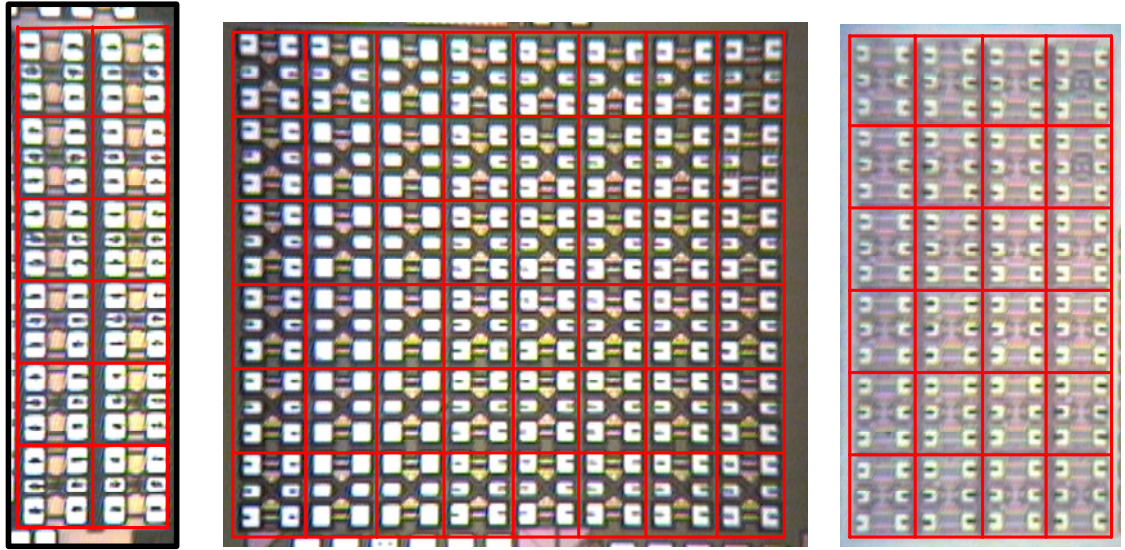


Fig. 2.1 The top view of an on-wafer test structure for transistors. (a) The whole test structure including probing pads. (b) The MOS transistor under test only. The dimension is not to scale.

Fig. 2.2 shows the pictures of the chips taken under the microscope. Fig. 2.2 (a) shows the chip fabricated on a $0.13\mu\text{m}$ RF CMOS technology for developing four-port calibration techniques. Each column contains five on-wafer standards and a $0.13\mu\text{m}$ NMOS transistor. An array of 90nm NMOS test structures with different gate connections and different layouts is fabricated on the 90nm chip in Fig. 2.2 (b). The measured S-parameters are used for characterizing the effects of different gate patterns on small-signal parameter extraction. The chip in Fig. 2.2 (c) contains an array of devices for intermodulation linearity characterization on 90nm CMOS technology. The necessary de-embedding standards are also included in Fig. 2.2 (b) and (c), which are laid close to the transistor structures to avoid space variation [62]. The small-signal parameters for the equivalent circuit used to calculate IP_3 can be extracted from the measured S-parameters.



(a) 0.13 μ m
CMOS

(b) 90nm CMOS

(c) 90nm CMOS

Fig. 2.2. Chip pictures of the fabricated transistor structures on three RF CMOS technologies. (b) and (c) are fabricated at different foundries.

2.2 Probing pad design considerations

Fig. 2.3 illustrates the cross section of a modern RF CMOS technology. It starts with a silicon substrate, which is normally lightly p-type doped. Active devices, including diodes, bipolar transistors, and CMOS transistors, and some of the passive devices, like poly resistors, are built on the very surface of the silicon substrate using doped materials. 1-4 thin metal layers (about 30nm thick) either aluminum or copper will be used for connections close to device terminals. These connections are thin and narrow, which can only handle low current, and has a higher resistance. Besides, the first metal layer is usually used as a metal ground plane under the probing pads and the interconnections to prevent the signal paths from coupling to the substrate [58] [63]. 2-4 thick metal layers (about 50nm thick) with low sheet resistance will be used for long

and high current connection. The RF layer, normally composing two thick aluminum metal layers (several micron thick), is used to build the probing pads. A passivation layer is used to protect the whole structure, and opening must be made on top of the probing pads.

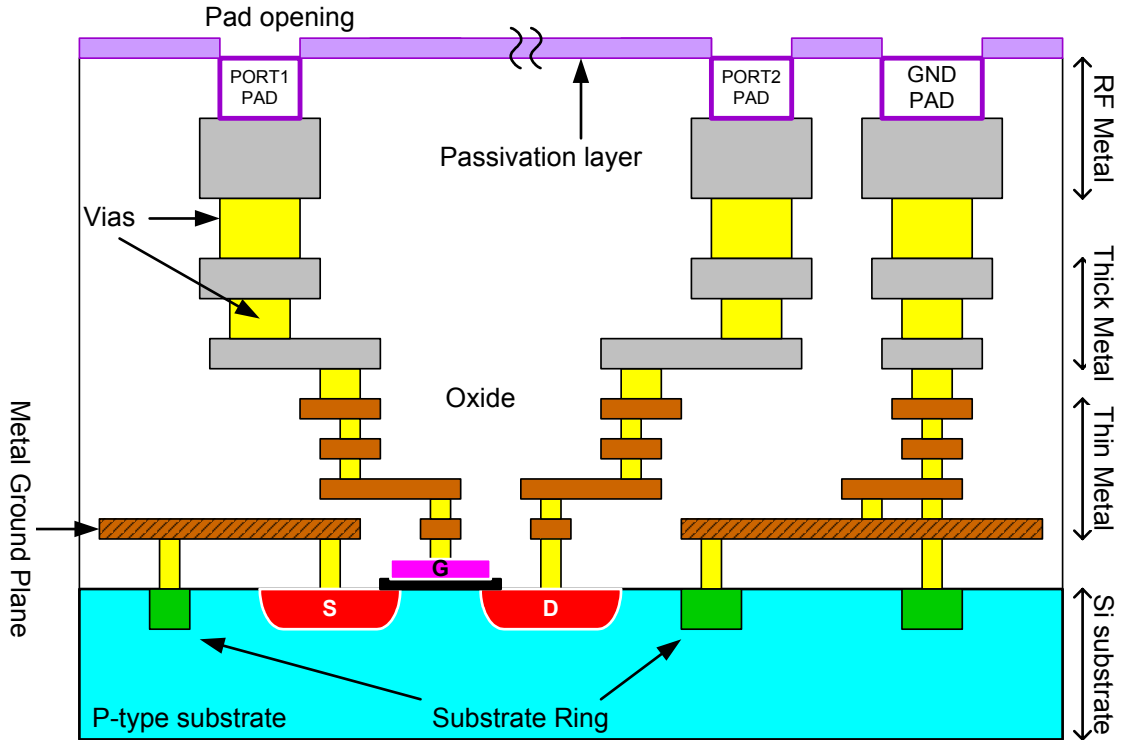


Fig. 2.3 Cross section view of an advanced RF CMOS technology. The dimension is to scale.

Fig. 2.4 shows the cross section view of three cuts along the test structure in Fig. 2.1 (a). The maximum pad height variation in a row of pads contacted by one GSG probe is $0.5\mu\text{m}$. To avoid pad height variation, the top metal layer for all ground pads and signal pads are the same. To support the overtravel of probe tips while probing, the pads are built using multiple metal layers, since even the thickest metal layer is less than

10 μ m thick. Fig. 2.4 (a) is just a copy of Fig. 2.1 (a) with three cut lines for the cross sections in Fig. 2.4 (b)-(d). The cross section of the GSG pads is shown in Fig. 2.4 (b). The ground pad is built using all metal layers, and the signal pad is built using the top two thick metal layers. The number of metal layers used for signal pads depends on the number of layers available and the metal layer thickness. The ground pads are all tied to the same metal ground plane built using the first metal layer to provide an as ideal as possible connection between the four ground pads. The metal ground plane exceeds the dimension of the signal pad by a size comparable to the total thickness of all metal layers to provide good electromagnetic isolation from the silicon substrate [58] [57]. A large number of substrate contacts are scattered over the wafer to provide good substrate connection and meet the requirement of doping and active area density.

Fig. 2.4 (c) shows the cross section of the cut along the middle between Port 1 and Port 2 pads, which shows that the connections between opposing ground pads (Port 1 to Port 2 side) are built using all available metal layers. This helps to provide an ideal and unified ground connection. The source is tied to the substrate ring locally, while the substrate ring is connected to the metal ground plane using short and wide metal lines. The grounded substrate ring can isolate the transistor from adjacent structures. Fig. 2.4 (d) is along a cut across the signal pads at Port 1 and Port 2. Not only the signal pads but also the interconnect lines to the transistor terminals are built using more than one metal layer. This will evidently reduce series parasitics of the leads, and increase the accuracy of the SHORT standard, but introduce coupling to the ground shield, thereby increase the loss. In general, it is safe to apply at least a few of the metal layers [57].

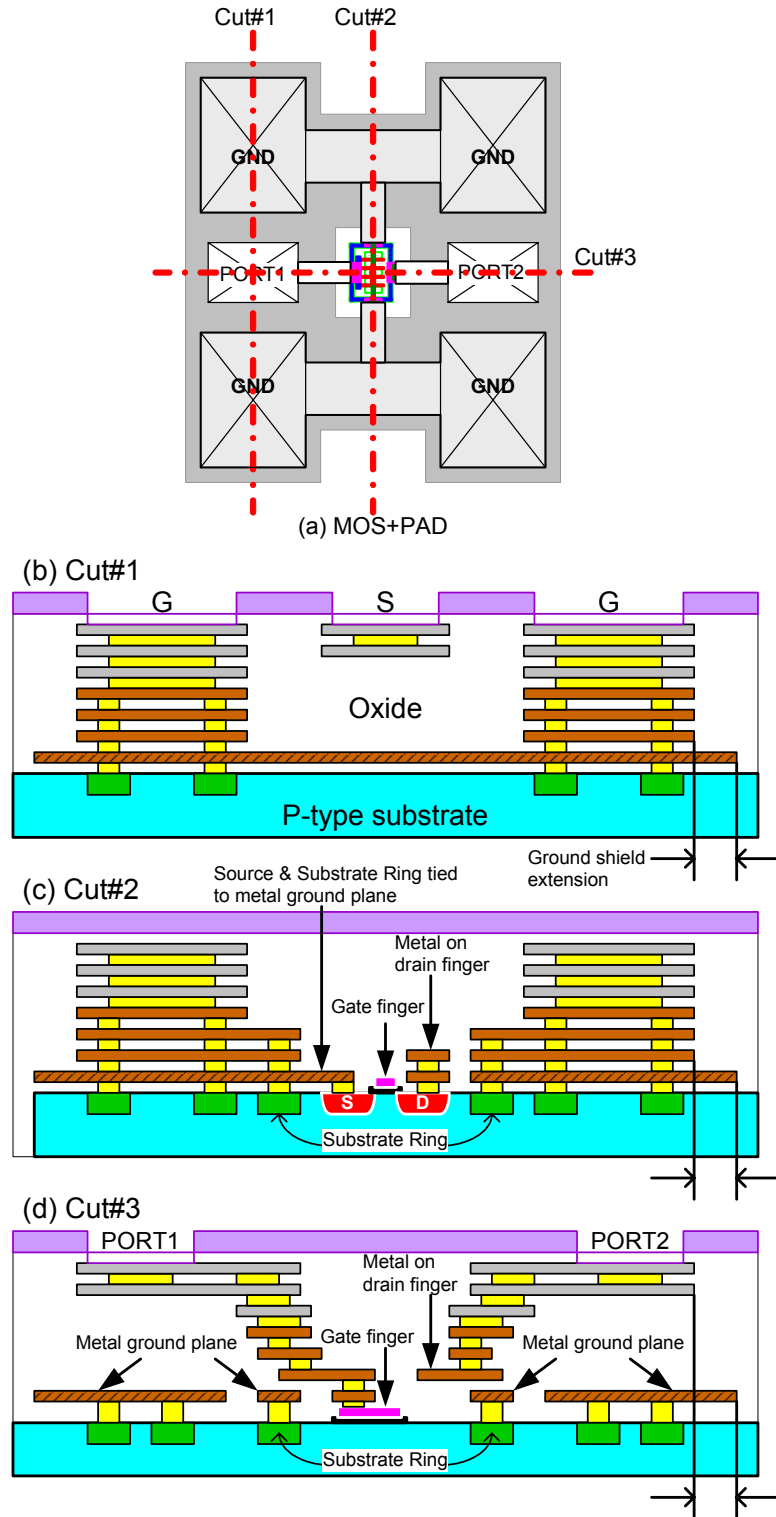


Fig. 2.4 The cross section view of GSG pads and MOS transistor along three cuts.

2.3 CMOS transistor design considerations

Fig. 2.5 gives the layout of a MOS transistor with 10 gate fingers. Each finger has a channel length of $0.13\mu\text{m}$, and a channel width of $5\mu\text{m}$ with double-sided gate contact. This leads to a “C-look” gate metal connection to Port 1. Port 2 is connected to the drain terminal of the transistor. The source terminal is tied to the substrate terminal and grounded. Multiple thick metal layers are used to connect the source and substrate to the ground pads to reduce substrate effect and nonidealities of on-wafer standards.

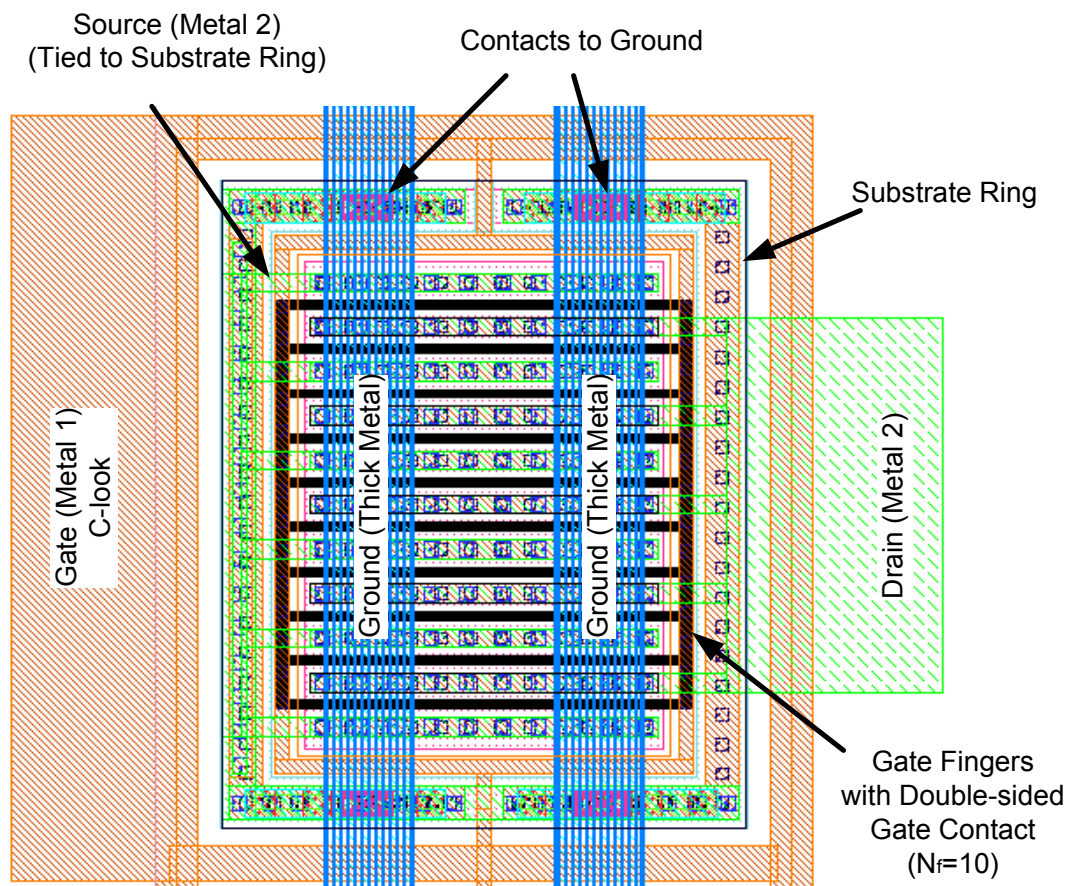


Fig. 2.5. Layout for one cell of the desired transistor.

Fig. 2.6 gives the layout of the desired NMOS transistor and the five on-wafer standards used to solve four-port error adaptors, OPEN, SHORT, LEFT, RIGHT, and THRU. Fig. 2.6 (a) is the desired NMOS transistor without pads and most of the interconnect lines. The total channel width of the transistor is $150\mu\text{m}$, and the channel length is $0.13\mu\text{m}$. The transistor contains three identical cells, i.e. multiplier factor=3. Each cell has the same layout as shown in Fig. 2.5. Without specification, the S-, Y-, and Z- parameters used to perform error correction from Chapter 4 to Chapter 6 are measured on this set of test structures fabricated on the $0.13\mu\text{m}$ chip in Fig. 2.2 (a). The reference plane is selected to be as close as possible to the gate and drain terminals of the transistor, which is marked out on the OPEN structure in Fig. 2.6 (b).

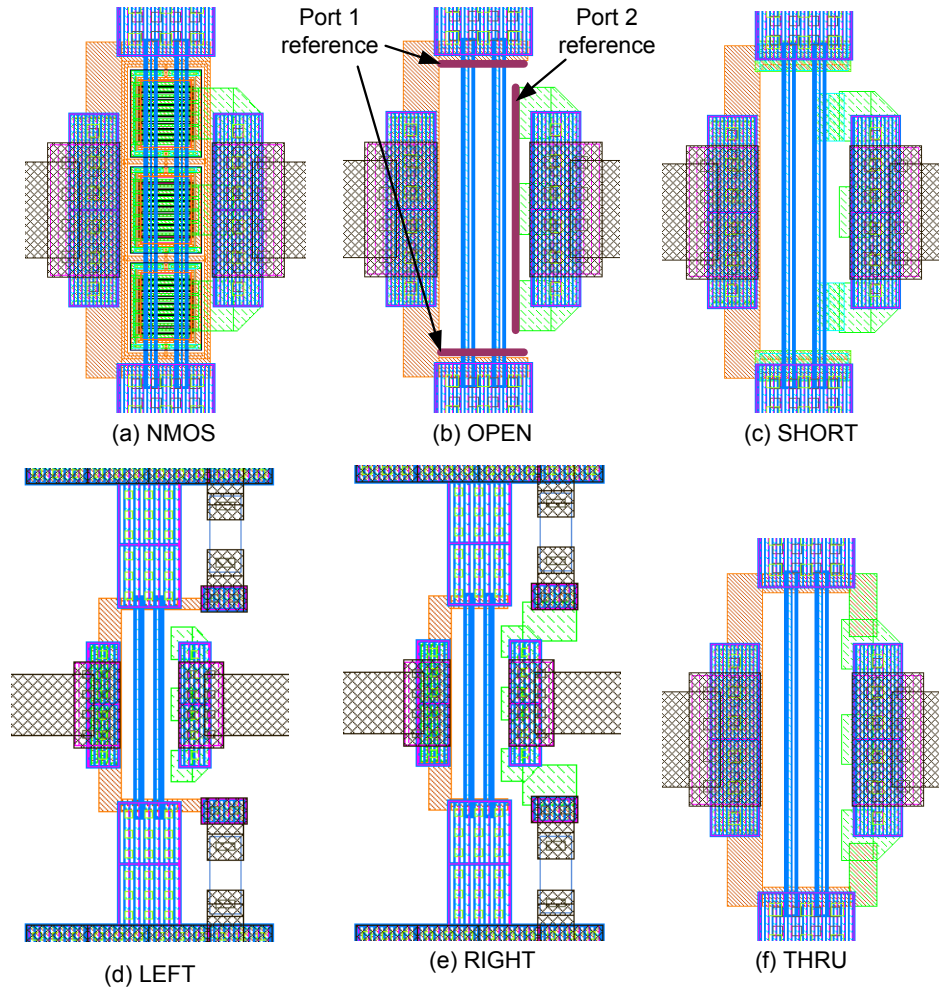


Fig. 2.6. Layout for the desired transistor, NMOS, and the on-wafer standards, OPEN, SHORT, LEFT, RIGHT and THRU.

The OPEN structure in Fig. 2.6 (b) just takes the transistor out together with the substrate ring and the necessary lowest layer metal connections. The SHORT structure in Fig. 2.6 (c) shorts the metal at the Port 1 and Port 2 reference plane to ground using short and wide metal lines. Multiple metal layers can be used if necessary. LEFT structure in Fig. 2.6 (d) has two 100Ω metal resistors connected to Port 1 in parallel to provide balanced signal flow at the GSG probe. In like manner, RIGHT structure in Fig. 2.6 (e) has the same two 100Ω resistors connected to Port 2 in parallel. One end of the

resistors is connected to the reference metal as close as possible to Port 1 or Port 2. The other end is terminates to ground. However, it is hard to connect this end to the same ground plane as the SHORT structure because of the size limitation of this back-end-of-line (BEOL) resistor. So, there is a reference plane variation between OPEN, SHORT and LEFT, RIGHT. Assuming the ground plane is very well connected throughout the whole structure, this variation is negligible. The THRU structure in Fig. 2.6 (f) simply shorts Port 1 reference to Port 2 reference in the shorted way. Since the metal line used to short Port 1 and Port 2 are wide and very short, comparable to the pattern of gate fingers, THRU standard can be considered as ideal THRU without delay and loss.

2.3.1 Gate pattern and multiplier factor

Fig. 2.7 shows the layout for three NMOS transistors with the same gate length and total gate width, but different gate patterns and multiplier factors (M). The three transistors have (a) double-sided gate contact with $M=1$, (b) single-sided gate contact with $M=1$, and (c) double-sided gate contact with $M=4$. Note that the “C-look” gate metal connection is used for double-sided gate contact to balance the current flow at the two-ends. The transistor with single-sided gate contact is directly connected to Port 1 using wide metal lines. This set of layout is used to investigate the impact of the “C-look” gate metal and the multiplier factor on the RF and noise performance.

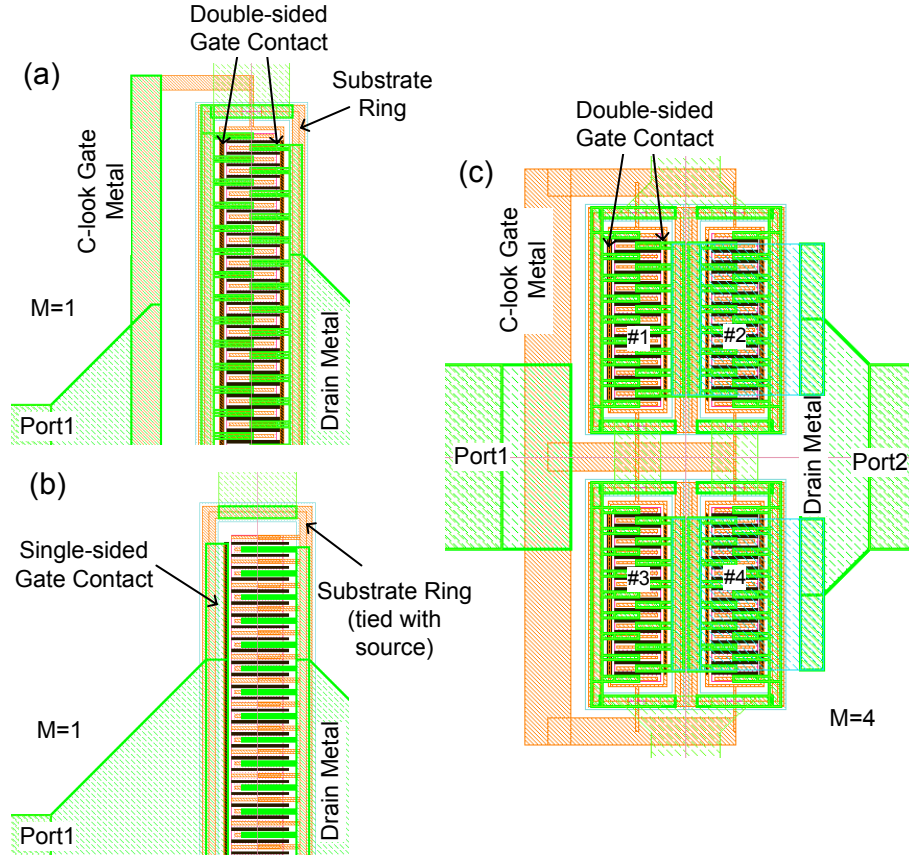


Fig. 2.7. Layout for NMOS transistors with different gate patterns and multiplier factors.

The gate resistance R_g , linear transconductance g_m , cut-off frequency f_T , and maximum oscillation frequency f_{\max} are the critical parameters for evaluating the RF and noise performances of a MOS transistor [56]. SOLT calibration is used for system error calibration. Since the parameters examined here are extracted at frequencies below 10GHz, open-short is valid in this frequency range [14] [45]. f_T is extracted using the -20dB/dec extrapolation method from the H_{21} versus frequency curve at each bias point. f_{\max} is extracted using the -20dB/dec extrapolation method from the Mason's unilateral

gain (MUG) versus frequency curve at each bias point [64]. An example of f_T and f_{\max} extraction is shown in Fig. 2.8 and Fig. 2.9. Maximum available gain (MAG) and maximum stable gain (MSG) do not follow the -20dB/dec slope, and thus are not used for f_{\max} extraction. R_g and g_m are extracted using [65]

$$R_g = \frac{\Re(Y_{21})}{\Im(Y_{11})^2} \Big|_{\text{low frequency}} \quad \text{and} \quad g_m = \Re(Y_{21}) \Big|_{\text{lowest frequency}} \quad (2.1)$$

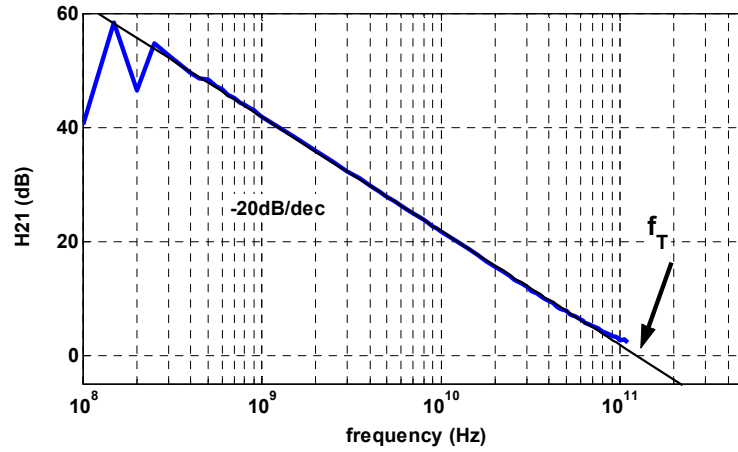


Fig. 2.8. An example for f_T extraction.

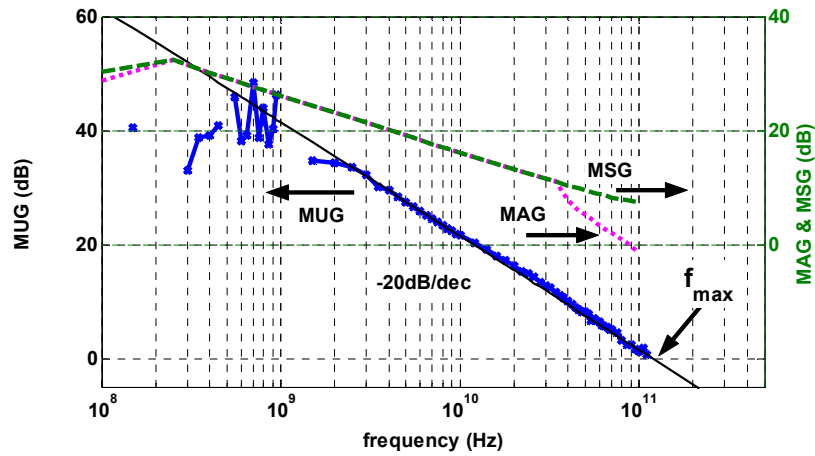


Fig. 2.9. An example for f_{\max} extraction.

Fig. 2.10 compares R_g , g_m , f_T and f_{\max} extracted for the three NMOS transistors. Fig. 2.10 (a) and (c) show that the transconductance and the cut-off frequency for the three transistors are approximately the same. Fig. 2.10 (b) shows that the transistor with double-sided gate contact and M=4 has the smallest gate resistance, and thus the best noise performance theoretically. Noise parameters are not measured due to lack of equipments. The transistor with single-sided contact and M=1 does not have the largest R_g as expected. Instead, the transistor with double-sided contact and M=1 gives the largest R_g . The reason may lie on the narrow metal connection from the reference plane to the double-sided gate contact, while a much wider metal connection is used in single-sided gate contact transistor in Fig. 2.7 (b). However, it is not possible to move the reference plane to the end of the narrow gate metal inside the substrate ring, as it is impossible to layout de-embedding standards in such small area. Fortunately, this problem can be solved by using different metal connections to the gate. For example, the transistor layout with M=4 greatly reduces the resistance on the narrow metal lines because it has four similar parallel connections. Fig. 2.10 (d) shows that f_{\max} is quite different for the three transistors. The device with single-sided gate contact and M=1 and the transistor with double-sided contact and M=4 gives the highest f_{\max} . This agrees with the lowest R_g of these two transistors.

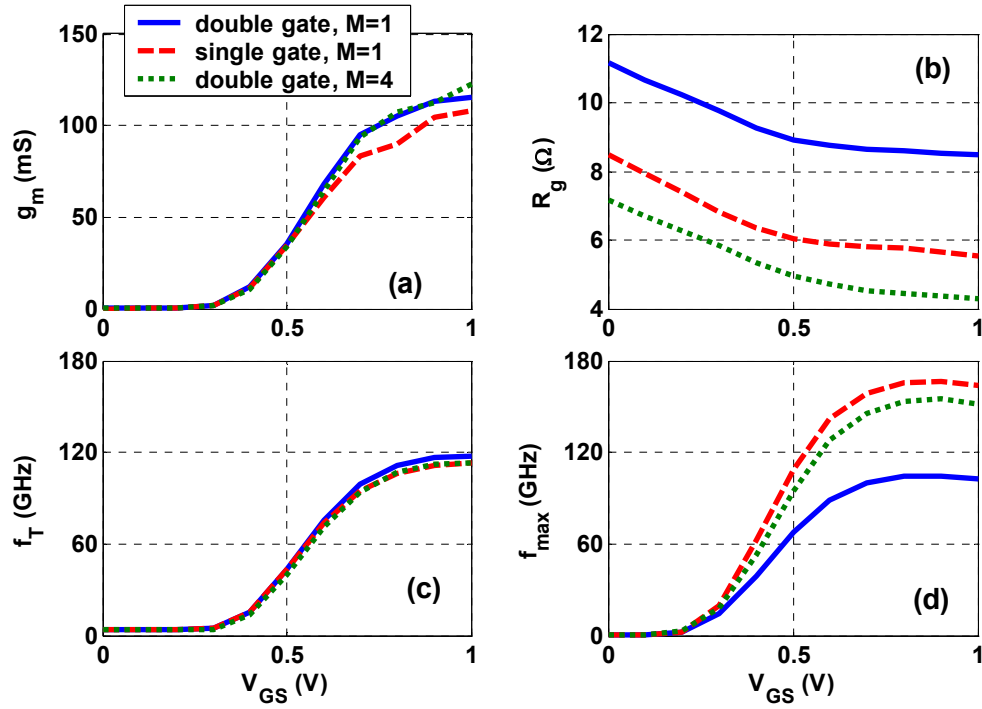


Fig. 2.10. Extracted parameters for three NMOS transistors with different gate patterns and multiplier factors.

2.3.2 Gate finger configuration

Fig. 2.11 shows the layout for three NMOS transistors with the same gate length and total gate width, but different number of fingers (N_f) and finger width (W_f). The number of fingers and the finger width of the three transistors are (a) $N_f=20$, $W_f=2\mu\text{m}$. (b) $N_f=10$, $W_f=4\mu\text{m}$. (c) $N_f=5$, $W_f=8\mu\text{m}$. All of the transistors are laid out using double-sided gate contact and “C-look” gate metal connection. Fig. 2.12 (a)-(d) compare R_g , g_m , f_T , and f_{\max} extracted from open-short de-embedded Y-parameters. Again, SOLT calibration and open-short are used for system error calibration and on-wafer parasitics de-embedding. The first two transistors with $N_f=20$, $W_f=2\mu\text{m}$ and $N_f=10$, $W_f=4\mu\text{m}$ are

practically the same for the four parameters extracted. The transistor with the longest W_f ($W_f=8\mu\text{m}$) has the largest R_g and thus the lowest f_{max} as expected. However, the R_g value difference does not follow ideal scaling rules of CMOS transistors. The reason may also lie on the narrow metal connection to the double-sided gate contact. The R_g value extracted is dominated by the resistance on the metal lines instead of the gate fingers.

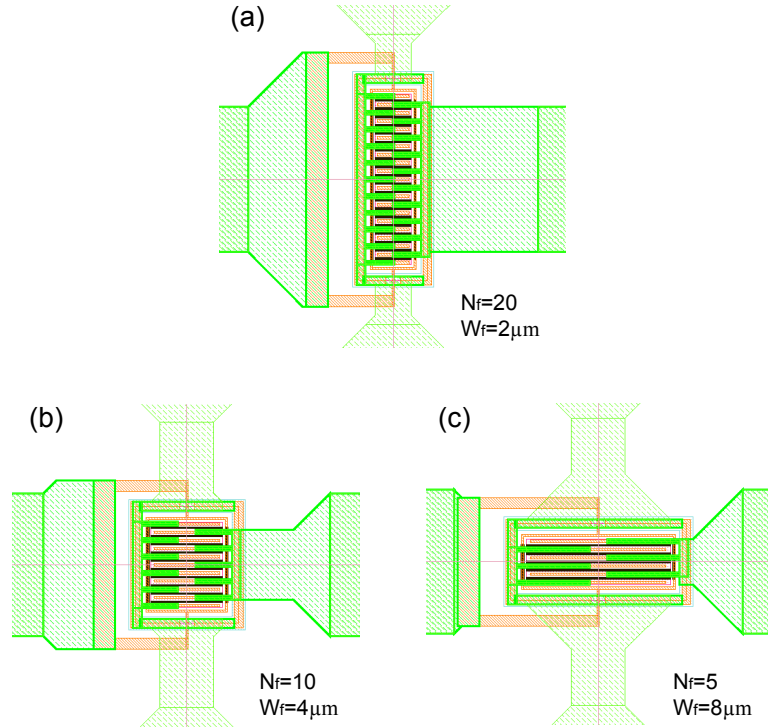


Fig. 2.11. Layout for three NMOS transistors with same total channel width but different finger width and finger number. $W_{\text{total}}=40\mu\text{m}$.

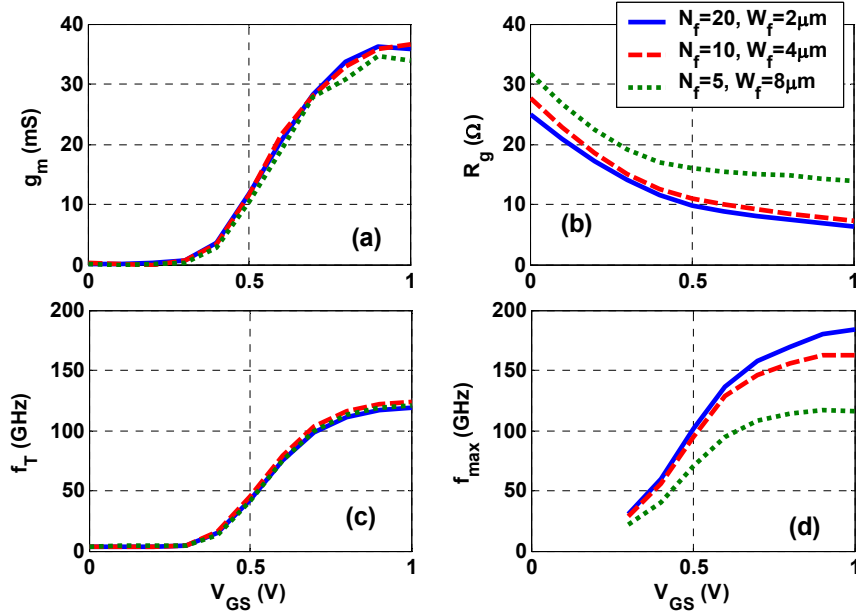


Fig. 2.12. Extracted parameters for three NMOS transistors with same total channel width but different finger width and finger number. $W_{\text{total}}=40\mu\text{m}$.

2.4 Summary

The layout rules concerning reliable on-wafer probing are detailed. It is recommended to use all metal layers for ground pad, and more than one top layer for signal pads. Ground shield need to be carefully designed. The transistor characteristic fluctuation caused by layout variation is examined. Double-sided gate contact does not necessarily provide lower gate resistance. The gate pattern needs to be optimized. Otherwise, the metal lines connecting out can have considerable impact on gate resistance. On the other hand, the selection of reference plane is of great important for transistor characterization.

CHAPTER 3

ERROR MODELS FOR TWO-PORT S-PARAMETER MEASUREMENT

Of paramount importance in on-wafer transistor characterization at RF frequencies is to properly correct the errors introduced by the VNA system and on-wafer parasitics [66] [13] [21] [12]. The demand for increased measurement accuracy in on-wafer S-parameter measurement can be achieved by improving the hardware, the models used for characterizing measurement errors, the calibration methods used for calculating these errors, and the definitions of calibration standards [34]. The type of the error model depends on the hardware topology of the VNA. There are three-receiver VNA and four-receiver VNA for two-port measurement. The three-receiver VNA has one reference receiver for detecting the incident signal, and two measurement receivers, one at each port. The corresponding error model is a 12-term error model, 6 for forward direction, 6 for reverse direction [67] [68]. For double-reflectometer VNA with four receivers, a 8-term error model was introduced and solved in S-parameters and T-parameters [69] [70] [71]. The leakage terms can be added to the 8-term error model, one for each measurement direction, increasing the number of error coefficients to 10 [72]. Both the 8(10)-term model and the 12-term models are used for four-receiver VNA, which even have their calibration procedures embedded in modern VNAs. The error models and their corresponding calibration techniques are compared in [73] [74] [68] [75]. If required, several techniques with different conversion

equations can be used to convert the 12-term model into a 8(10)-term model [75] [76]. These equations are slightly different but are based on the same physical principle. One may also apply the 8(10)-term model for the three-receiver VNA, with an assumption that the source match equals the load match of the test set, which holds only in the case of an ideal switch. For a real system, this may lead to intolerable measurement inaccuracy. Only the 12-term model guarantees the entire description of three-receiver VNA [34]. The reasons will be detailed in Section 3.4.

However, both the 8(10)-term model and the 12-term model make an arbitrary assumption that the leakage terms bypassing the unknown two-port are negligible. Further measurement experiments and practical experiences reveal that the leakage terms can have a very complicated nature. A much more general concept of error model was introduced by Speciale and Franzen in 1977 [37] [35]. The systematic errors of a n -port VNA are represented by a $2n$ -port virtual error adapter, with its n -port connected to the n -port unknown network, and its other n -port connected to the ideal, error-free VNA. The error adapter consists of $2n \times 2n$ coefficients and describes all possible paths between the $2n$ receivers. For two-port measurement, the error adapter is a four-port network, which involves 4×4 error coefficients, i.e. a 16-term model. The 16-term model is only solvable for four-receiver ($2n$ -receiver) VNA. However, it is also possible to define a full error model for three-receiver ($n+1$ receiver) VNA. This includes significantly more error coefficients, for example, the 22-term model for a three-receiver two-port VNA, compared with the 16-term model for four-receiver two-port VNA [42]. The four-port error adapter can not only be applied on systematic error removal, but also be used to remove on-wafer parasitics as it does not make any

assumptions of the error network. The four-port error network is described in Section 3.5. There are also techniques published to solve the 16-term model in S- or T-parameters using five standards [35] [37] [36] [41] [43] [39]. Two general approaches to solve the four-port error network using five standards are developed in this dissertation, an analytical solution based on Y-parameters in Chapter 4, and a numerical solution using SVD and T-parameters in Chapter 5.

3.1 Two-port S-parameter measurement

Fig. 3.1 is the block diagram for the two-port S-parameter measurement system in Fig. 1.2, which includes VNA8510C system, DC power supply, and a control computer. The measurement is controlled by a MATLAB program on the computer through a USB to GPIB controller. For each measurement, the program first biases the DUT by sending GPIB commands to the DC power supply, then starts one single frequency sweep by sending GPIB commands to the VNA's processor. The DC voltage is added to the DUT through two bias tees inside the VNA test set. Two DC cables connect the outputs of the DC power supply to the test set from the backside of the two equipments, which are illustrated using dash lines in Fig. 3.1.

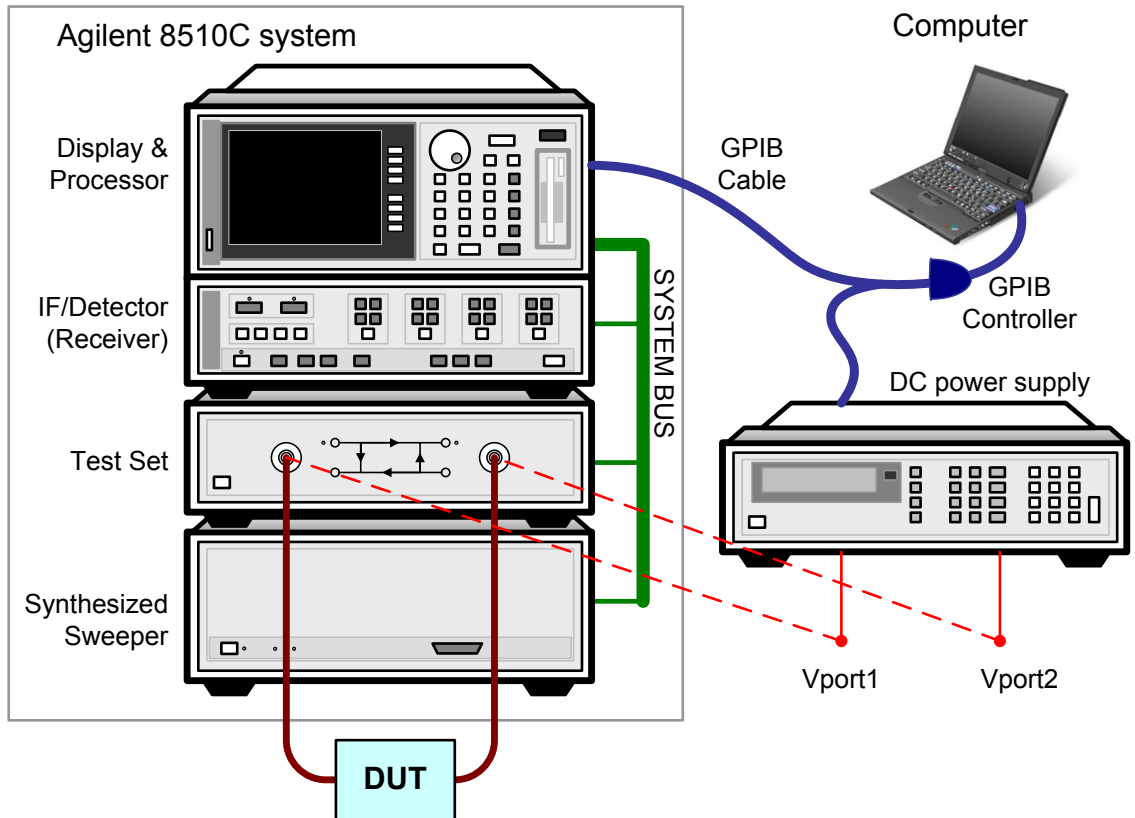


Fig. 3.1 Block diagram for two-port S-parameter measurement using Agilent 8510C system.

Fig. 3.2 shows the simplified block diagram of a two-port system involving a four-receiver VNA and an unknown two-port. The two bias tees are assumed to be ideal for AC signals and thus not included. In Section 3.2, it will be shown that the errors introduced by the bias tees are actually included in the four-port error adapter. A dual reflectometer is attached to the input of the unknown two-port DUT, and another one is attached to the output. Thus, the VNA has four receivers, two at each port, to capture the incident and reflected waves at each port. A switch changes the direction of the incident power to the unknown DUT for forward and reverse measurements, and terminates the unknown DUT at an impedance Z_0 . The four S-parameters exported by

the VNA are actually the ratios of the incident and reflected waves monitored by the two dual-reflectometer. $S_{11} = b_0 / a_0$ and $S_{21} = b_3 / a_0$ are calculated when the switch is at forward position as a_0 is the incident signal and b_0 and b_3 are the reflected waves. $S_{12} = b_0 / a_3$ and $S_{22} = b_3 / a_3$ are calculated when the switch is at reverse position as a_3 is now the incident signal.

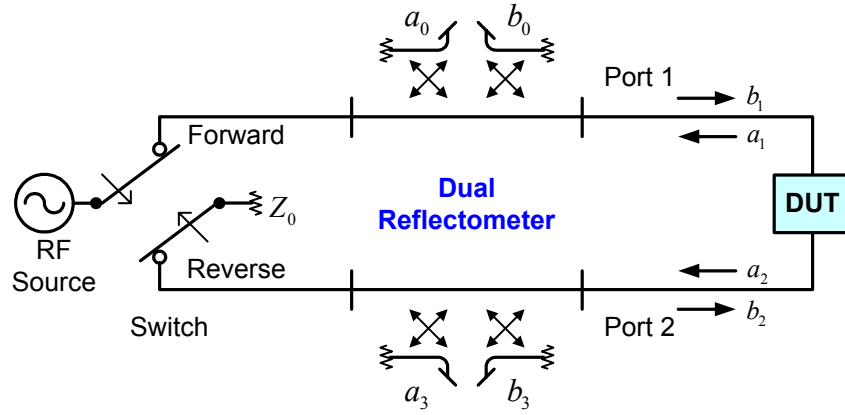


Fig. 3.2 A two-port VNA system with four receivers.

If Z_0 is a perfect matched load and the switch is ideal, the waves and the S-parameters can be related through

$$\begin{aligned} b_0 &= S_{11}a_0 + S_{12}a_3 \\ b_3 &= S_{21}a_0 + S_{22}a_3 \end{aligned} \quad (3.1)$$

Under forward mode, $a_3 = 0$, the equation reduces to $b_0 = S_{11}a_0$ and $b_3 = S_{21}a_0$. Under reverse mode, $a_0 = 0$, (3.1) becomes $b_0 = S_{12}a_3$ and $b_3 = S_{22}a_3$. Therefore, the measured S-parameters, S^M , are the wave ratios calculated,

$$S^M = \begin{bmatrix} S_{11}^M & S_{12}^M \\ S_{21}^M & S_{22}^M \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}. \quad (3.2)$$

If Z_0 is not a perfect matched load or the switch is nonideal, i.e. $a_3 \neq 0$ in forward mode, and $a_0 \neq 0$ in reverse mode. The waves measured under forward mode and reverse mode can be combined as

$$\begin{bmatrix} b_0 & b_0' \\ b_3 & b_3' \end{bmatrix} = \begin{bmatrix} S_{11}^M & S_{12}^M \\ S_{21}^M & S_{22}^M \end{bmatrix} \begin{bmatrix} a_1 & a_1' \\ a_3 & a_3' \end{bmatrix}. \quad (3.3)$$

The superscript “'” differs the waves measured in reverse mode from the waves measured in forward mode. S^M can be calculated from the wave ratios as [39]

$$S^M = \begin{bmatrix} \frac{S_{11} - S_{12}S_{21}\Gamma_1}{D} & \frac{S_{12} - S_{11}S_{12}\Gamma_2}{D} \\ \frac{S_{21} - S_{22}S_{21}\Gamma_1}{D} & \frac{S_{22} - S_{21}S_{12}\Gamma_2}{D} \end{bmatrix}, \quad \Gamma_1 = \left. \frac{a_3}{b_3} \right|_{forward}, \quad \Gamma_2 = \left. \frac{a_0'}{b_0'} \right|_{reverse}. \quad (3.4)$$

$D = 1 - S_{21}S_{12}\Gamma_1\Gamma_2$. Γ_1 and Γ_2 are the two additional wave ratios measured under forward and reverse mode while probing a THRU standard, which can only be measured by four-receiver VNAs. The process to remove the switch errors caused by the non-ideal switch and imperfect Z_0 load is called “switch error removal”, which can only be performed on four-receiver VNAs. Fortunately, most of the modern VNAs are four-receiver VNA. The derivation of the equations and a step-by-step guide to measure Γ_1 and Γ_2 are detailed in Appendix D.

Denote S^{DUT} as the S-parameters of the unknown two-port. The directions of the waves in Fig. 3.2 are defined in a manner that simplifies the error adapter description in Section 3.2. Thus, the directions of a_1 , b_1 , a_2 , and b_2 give

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = S^{DUT} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \quad (3.5)$$

Since the real world measurement system is not perfect, there are random errors and systematic errors contributing to the measurement of the unknown two-port S^{DUT} , i.e. $S^M \neq S^{DUT}$. For example, Fig. 3.3 (a) shows the magnitude of the measured $S_{11}^{M,load}$ of an ideal resistive termination with $S_{11}^{DUT,load} = 0$. $|S_{11}^{M,load}|$ has 0.01 peak-to-peak variations with respect to frequency. Fig. 3.3 (b) shows the measured $|S_{11}^{M,short}|$ for an ideal short with $S_{11}^{DUT,short} = -1$. $|S_{11}^{M,short}|$ has an obvious frequency dependence, and the values are far away from one. These ideal devices are fabricated on Alumina substrate, modeled based on physical parameters, and verified by National Institute of Standards and Technology (NIST) [77] [78] [12]. So, the variations are not in the ideal load or short. Instead, these errors are introduced by the measurement system. The random errors, e.g. thermal drift, can only be described statistically, which cannot be systematically corrected. The systematic errors are reproducible and can be corrected using computational techniques. However full correction is impossible, due to superimposed random fluctuations in the measured results [12]. The linear systematic errors introduced by the imperfect reflectometer can be modeled by a fictitious two-port error adapter between the reflectometer and the unknown one-port. This results in a perfect reflectometer with no loss, no mismatch, and no frequency response errors.

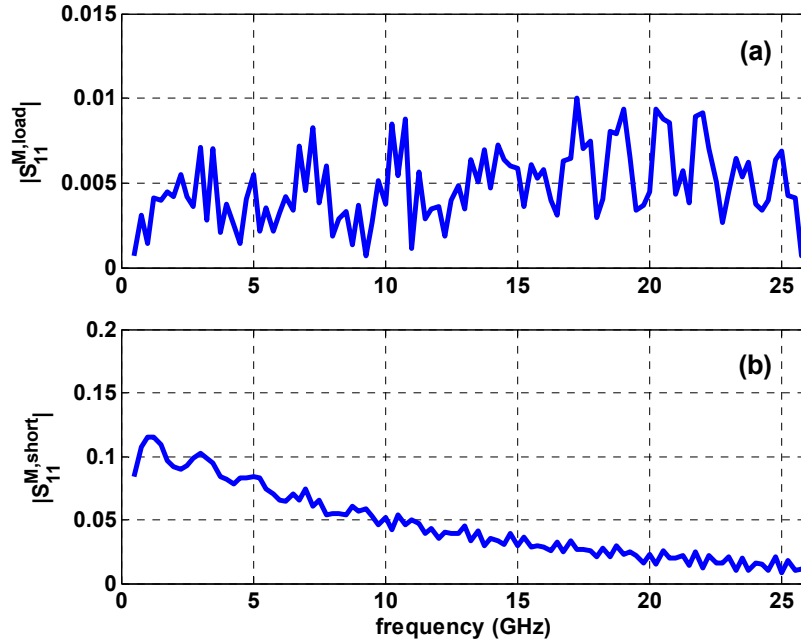


Fig. 3.3 The magnitude of the measured S_{11} of an ideal (a) LOAD and (b) SHORT.

3.2 Error adaptor concept

In general, all of the linear errors of the imperfect reflectometers, including directivity errors, frequency response errors, and port match errors, can be lumped into an error adaptor. This fictitious error adaptor is a four-port network, containing 16 error terms since four-port network is presented as a 4×4 matrix mathematically. Fig. 3.4 shows the two-port system with a four-port error adaptor inserted between the perfect reflectometer and the unknown DUT. Port 0 and Port 3 are the two perfect measurement ports inside the VNA, while Port 1 and Port 2 are the two terminals of the unknown two-port. a_k is the incident wave to the four-port error adaptor, while b_k is the reflected wave to the error adaptor. Without specification, the directions of the waves in error models and calibration techniques are all defined in the same manner. The subscript is

the port number where the wave is monitored. $k=0,1,2,3$. Note that the two bias tees are three-port components. The return losses and insertion losses of the bias tee are included in the four-port error adapter, but the leakage errors to the DC power supply are not. However, the leakages to the DC power supply do not affect the main signal path, and it is safe to ignore these leakages without any loss in accuracy [68] [79].

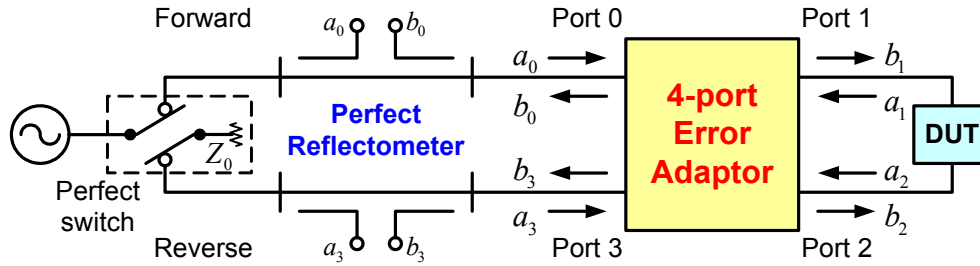


Fig. 3.4 The four-port system error adaptor for two-port S-parameter measurement.

3.3 The simplest 8-term error model

The 8-term model simply doubles the 4-term model for a one-port system at the two ports [70] [80]. The signal flow graph for the whole error adapter and the DUT is illustrated in Fig. 3.5. The two error adapters at the two ports are named as X-adaptor and Y-adaptor. The error terms are represented using S-parameters. Two additional leakage terms are added to the 8-term model which turn it to a 10-term model as shown in Fig. 3.6 [67]. The first explicit solution for 8-term model was introduced in 1971 by Kruppa and Sodomsy. Three reflection standards, open, short, matched load, and one through standard with the two ports connected together are used to calculate the error terms in S-parameters [70]. The error terms can be either solved using S-parameters or T-parameters, and modified approaches for different test structures are developed in [67]

[68] [69] [71] [72] [79] [81] [82]. The solution is not shown here as it is not used during transistor characterization in this dissertation.

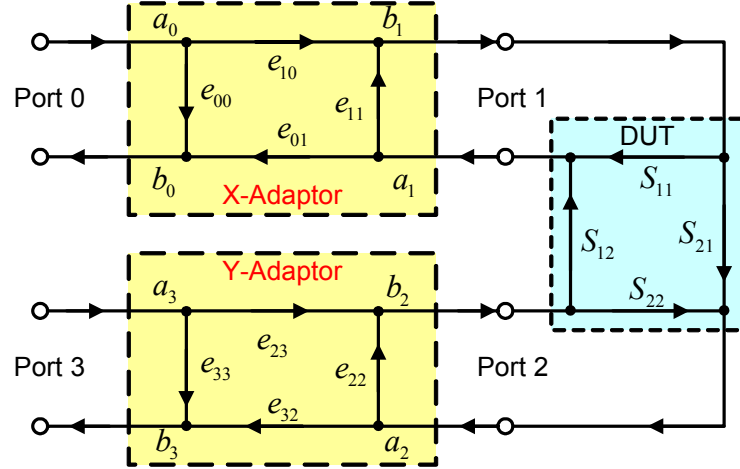


Fig. 3.5 Signal flow graph of 8-term error model for a two-port system.

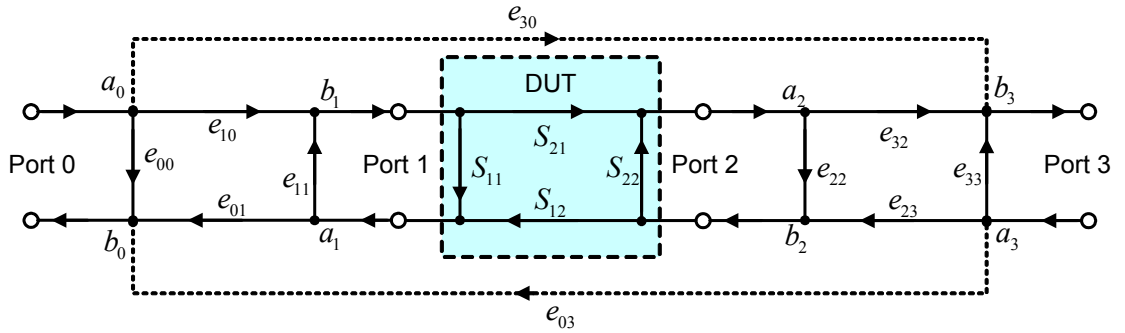


Fig. 3.6 The modified 10-term error model with two leakage errors added.

3.4 The classical 12-term error model

The classical 12-term model handles the switch error problem by using two separate error models for forward and reverse mode. This error model can be applied for both four-receiver VNA, and three-receiver VNA. The switch errors no longer need

to be removed using (3.4). This error model is still widely used in error correction techniques, e.g. short-open-load-thru (SOLT) calibration.

3.4.1 Forward mode

Under forward mode, the incident wave a_0 , the reflected wave b_0 , and the transmitted wave b_3 can be measured by both three-receiver VNA and four-receiver VNA. Fig. 3.7 shows the block diagram of a two-port VNA configured for forward measurement. Fig. 3.8 illustrates the possible signal paths using a signal flow graph for forward mode operation, based on 8(10)-term model. e_{30} represents the leakage path between the incident signal receiver, a_0 , and the transmission receiver, b_3 . Γ_3 lumps the impact of non-ideal switch or non-ideal Z_0 termination. Using signal flow graph analysis, the a_3 node can be removed, and the signal flow graph in Fig. 3.9 is equivalent to the signal flow graph in Fig. 3.8, with

$$(e_{22})^* = e_{22} + \frac{e_{32}e_{23}\Gamma_3}{1 - e_{33}\Gamma_3}, (e_{32})^* = \frac{e_{32}}{1 - e_{33}\Gamma_3}, \Gamma_3 = \frac{a_3}{b_3}. \quad (3.6)$$

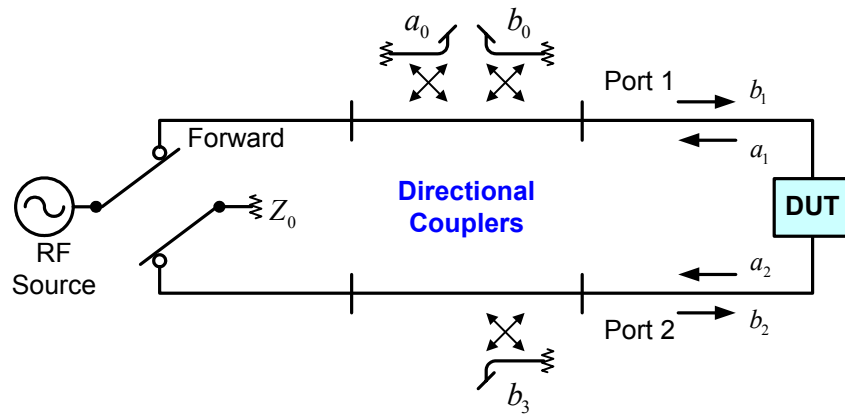


Fig. 3.7 A two-port S-parameter measurement system configured for forward mode.

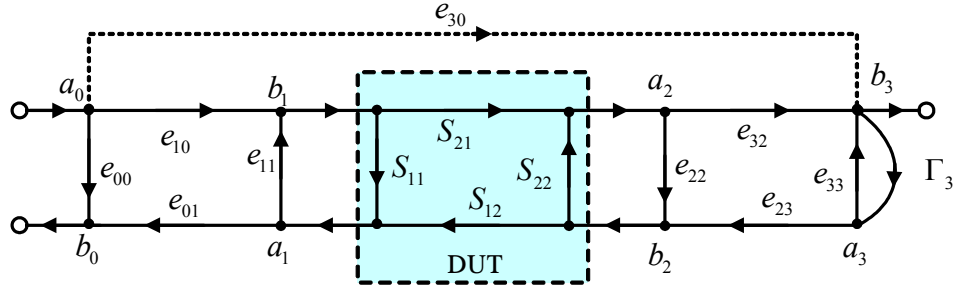


Fig. 3.8 Forward mode signal flow graph for two-port system including non-ideal Z_0 termination.

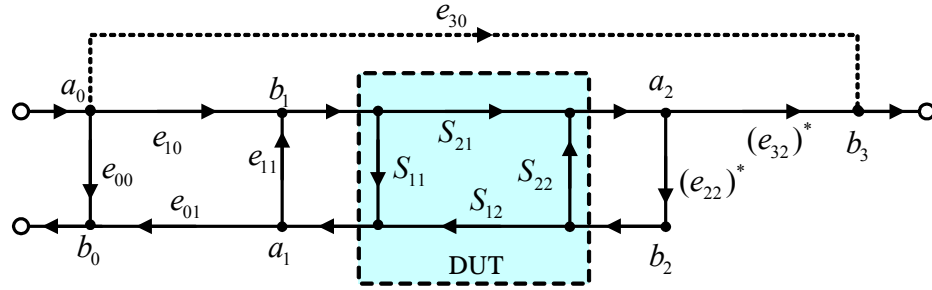


Fig. 3.9 Simplified forward mode signal flow graph.

3.4.2 Reverse mode

Fig. 3.10 shows the block diagram for reverse configuration. Under reverse mode, the incident wave a_3 , the reflected wave b_3 , and the transmitted wave b_0 are measured by a three-receiver VNA or a four-receiver VNA. Fig. 3.11 illustrates the signal flow graph for reverse mode operation using S-parameters. e_{03} represents the leakage path between the incident signal receiver, a_3 , and the transmission receiver, b_0 . Γ_0 lumps the impact of non-ideal switch or non-ideal Z_0 termination. Similarly, the a_0 node can be removed using signal flow graph analysis, and the signal flow graph in Fig. 3.12 is equivalent to the signal flow graph in Fig. 3.11, with

$$(e_{11})^* = e_{11} + \frac{e_{10}e_{01}\Gamma_0}{1-e_{00}\Gamma_0}, (e_{01})^* = \frac{e_{01}}{1-e_{00}\Gamma_0}, \Gamma_0 = \frac{a_0}{b_0} \quad (3.7)$$

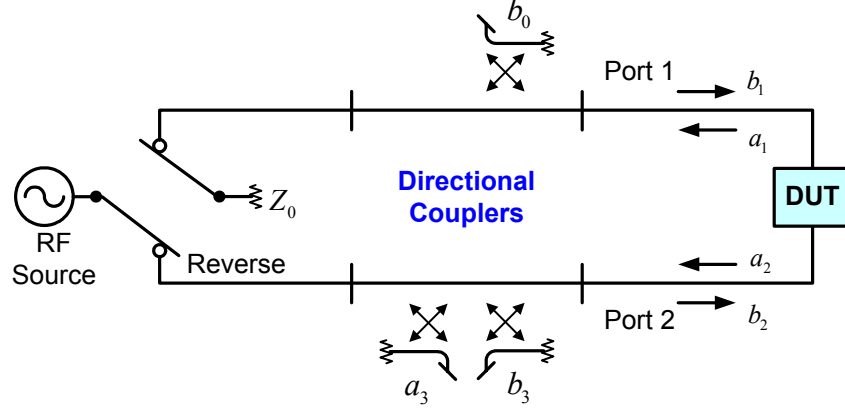


Fig. 3.10 A two-port S-parameter measurement system configured for reverse mode.

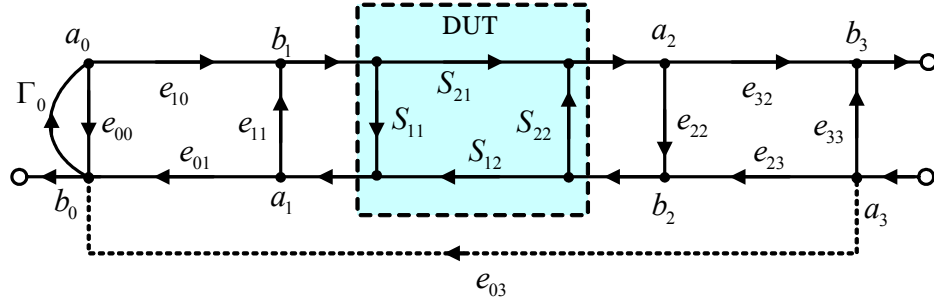


Fig. 3.11 Reverse mode signal flow graph for two-port system including non-ideal Z_0 termination.

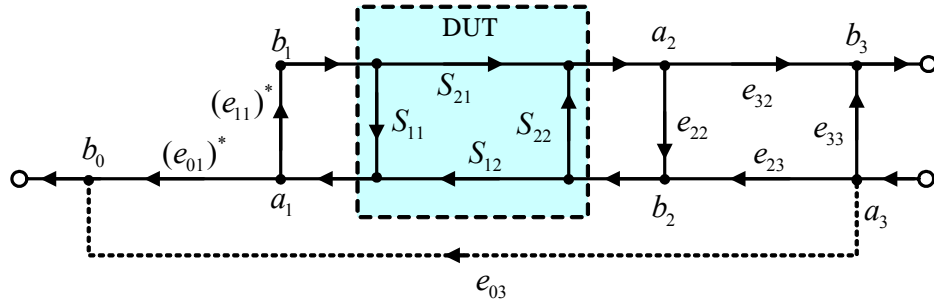


Fig. 3.12 Simplified reverse mode signal flow graph for two-port system.

3.4.3 12-term model

Fig. 3.13 redraws the signal flow graph for forward mode in Fig. 3.9. Note that e_{32} and e_{22} in Fig. 3.13 are not the same e_{32} and e_{22} that defined in 8-term model, instead they are the $(e_{32})^*$ and $(e_{22})^*$ calculated in (3.6), which involve the impact of switch errors since separate error adaptors are used for forward and reverse mode. This does not affect the error calibration procedures at all. Based on signal flow graph analysis, the measured wave ratios S_{11} and S_{21} are functions of the unknown S^{DUT} as [80]

$$S_{11} = e_{00} + \frac{(e_{10}e_{01})(S_{11}^{DUT} - e_{22}\Delta_{S^{DUT}})}{1 - e_{11}S_{11}^{DUT} - e_{22}S_{22}^{DUT} + e_{11}e_{22}\Delta_{S^{DUT}}} . \quad (3.8)$$

$$S_{21} = e_{30} + \frac{(e_{10}e_{32})(S_{21}^{DUT})}{1 - e_{11}S_{11}^{DUT} - e_{22}S_{22}^{DUT} + e_{11}e_{22}\Delta_{S^{DUT}}} . \quad (3.9)$$

where $\Delta_{S^{DUT}} = S_{11}^{DUT}S_{22}^{DUT} - S_{21}^{DUT}S_{12}^{DUT}$. The 6 (5 after normalization) error terms for forward mode are directivity error e_{00} , port match error e_{11} and e_{22} , frequency response error $e_{10}e_{01}$ and $e_{10}e_{32}$. The leakage errors e_{10} , e_{01} , and e_{32} cannot be completely determined because they can only be measured as products as shown in (3.8) and (3.9). Thus, only $e_{10}e_{01}$ and $e_{10}e_{32}$ can be solved, which is sufficient for calibration. This is equivalent to normalizing the error terms by e_{10} , as illustrated in Fig. 3.14 with the normalized values on the branches. The 6 error terms for reverse mode are directivity error e'_{33} , port match error e'_{11} and e'_{22} , frequency response error $e'_{23}e'_{01}$ and $e'_{23}e'_{32}$.

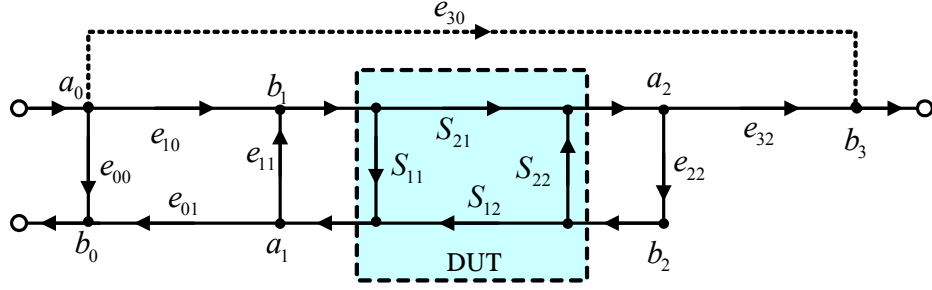


Fig. 3.13 Forward mode signal flow graph for two-port system.

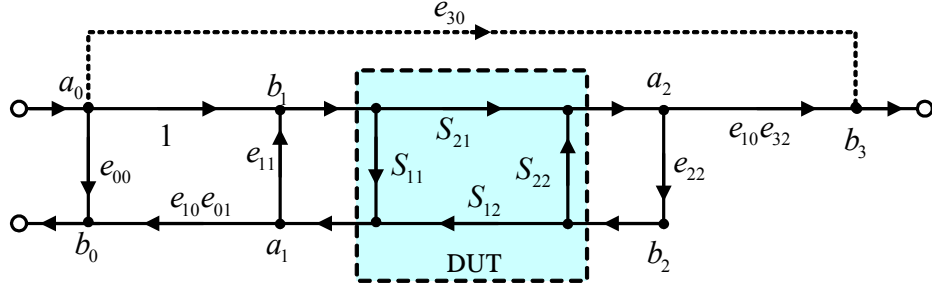


Fig. 3.14 Normalized 6-term error model for forward mode.

Since the 6-term model in Fig. 3.14 involves lumped error terms, these error terms no longer represent signal paths, instead they are just mathematical coefficients. To separate the error terms in forward mode and reverse mode, a superscript “'” is used to identify the waves and error terms in reverse mode. The normalized 6-term model for reverse mode is illustrated using the signal flow graph in Fig. 3.15. The measured wave ratios S_{22} and S_{12} are related to S^{DUT} as [80]

$$S_{22} = e'_{33} + \frac{(e'_{23}e'_{32})(S_{22}^{DUT} - e'_{11}\Delta_{S^{DUT}})}{1 - e'_{11}S_{11}^{DUT} - e'_{22}S_{22}^{DUT} + e'_{11}e'_{22}\Delta_{S^{DUT}}} . \quad (3.10)$$

$$S_{12} = e'_{03} + \frac{(e'_{23}e'_{01})(S_{12}^{DUT})}{1 - e'_{11}S_{11}^{DUT} - e'_{22}S_{22}^{DUT} + e'_{11}e'_{22}\Delta_{S^{DUT}}} . \quad (3.11)$$

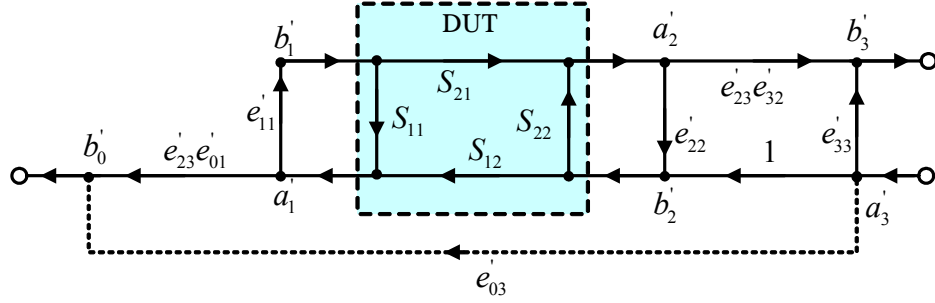


Fig. 3.15 Normalized 6-term error model for reverse mode.

Fig. 3.14 and Fig. 3.15 give the complete 12-term model. With 12 forward and reverse measurements, (3.8)-(3.11) give 12 equations. The 12 unknowns can be determined by solving the 12 equations simultaneously. Once the 12 error terms are determined, the S-parameters of the unknown two-port can be calculated as [80] [83]

$$S_{11}^{DUT} = \frac{\left(\frac{S_{11} - e_{00}}{e_{10}e_{01}} \right) \left(1 + \frac{S_{22} - e'_{33}}{e'_{23}e'_{32}} e'_{22} \right) - e_{22} \left(\frac{S_{21} - e_{30}}{e_{10}e_{32}} \right) \left(\frac{S_{12} - e'_{03}}{e'_{23}e'_{01}} \right)}{D}, \quad (3.12)$$

$$S_{21}^{DUT} = \frac{\left(\frac{S_{21} - e_{30}}{e_{10}e_{32}} \right) \left[1 + \left(\frac{S_{22} - e'_{33}}{e'_{23}e'_{32}} \right) (e'_{22} - e_{22}) \right]}{D}, \quad (3.13)$$

$$S_{12}^{DUT} = \frac{\left(\frac{S_{12} - e'_{03}}{e'_{23}e'_{01}} \right) \left[1 + \left(\frac{S_{11} - e_{00}}{e_{10}e_{01}} \right) (e_{11} - e'_{11}) \right]}{D}, \quad (3.14)$$

$$S_{22}^{DUT} = \frac{\left(\frac{S_{22} - e'_{33}}{e'_{23}e'_{32}} \right) \left(1 + \frac{S_{11} - e_{00}}{e_{10}e_{01}} e_{11} \right) - e'_{11} \left(\frac{S_{21} - e_{30}}{e_{10}e_{32}} \right) \left(\frac{S_{12} - e'_{03}}{e'_{23}e'_{01}} \right)}{D}, \quad (3.15)$$

$$D = \left(1 + \frac{S_{11} - e_{00}}{e_{10}e_{01}} e_{11} \right) \left(1 + \frac{S_{22} - e'_{33}}{e'_{23}e'_{32}} e'_{22} \right) - \left(\frac{S_{21} - e_{30}}{e_{10}e_{32}} \right) \left(\frac{S_{12} - e'_{03}}{e'_{23}e'_{01}} \right) e_{22}e'_{11}. \quad (3.16)$$

Note that all four measured S-parameters are used to calculate any one S-parameter in S^{DUT} , and each of the equations in (3.12)-(3.15) contains error terms calculated under forward and reverse mode. Thus, both the forward 6-term and the reverse 6-term affect the results of S^{DUT} , since essentially the forward error terms and the reverse error terms describe the same VNA system.

3.4.4 SOLT calibration

The classical 12-term model has been widely used for over 10 years. One of its well-established technique is the so called short-open-load-through (SOLT) calibration, or thru-open-short-match (TOSM) calibration, which is implemented on all modern VNAs [34]. During SOLT calibration, 12 measurements on four standards are done to solve the 12 error terms, 6 from forward mode, and 6 from reverse mode. The 6 forward measurements are three forward reflection measurements on OPEN, SHORT, and LOAD standards (S_{11}), one forward isolation measurement on two-port LOAD (S_{21}), one forward match and one forward transmission measurements on two-port THRU (S_{11} and S_{21}). Similarly, the 6 reverse measurements are S_{22} on OPEN, SHORT, and LOAD, S_{12} on two-port LOAD, S_{22} and S_{12} on two-port THRU. The accuracy of SOLT calibration depends critically on the fabrication and modeling tolerance of the standards. Additional procedures, such as improving the calibration standard models, or the use of standards initially characterized with respect to the reference calibration, can enhance the accuracy of the SOLT calibration [84] [85]. Fig. 3.16 shows the SHORT, LOAD, THRU standards on a Cascade impedance standard substrate (ISS) 101-190.

OPEN is defined as an open in air with a minimum distance of 250 μ m above the chuck surface. LOAD is built using two thin-film 100 Ω resistors in parallel [86] [13] [78]. The four standards are characterized using physical measurements and verified by National Institute of Standards and Technology (NIST) LRM/LRRM calibration [77] [78] [12].

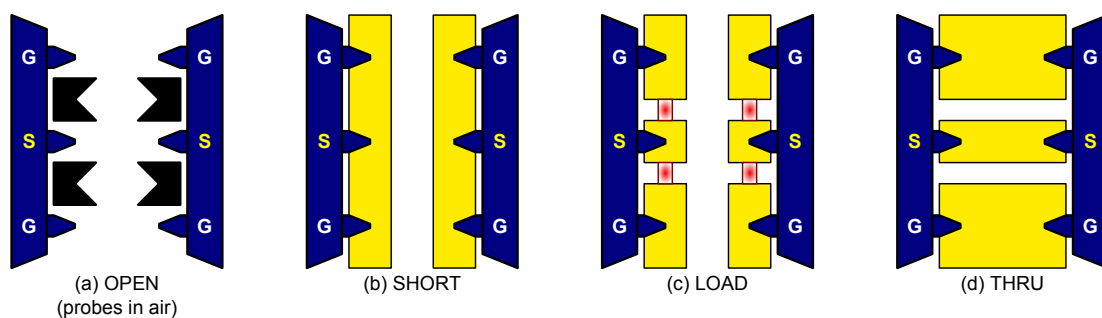


Fig. 3.16 (a) OPEN, (b) SHORT, (c) LOAD, and (d) THRU standards for SOLT calibration on Cascade ISS 101-190.

A significant assumption of SOLT calibration is that the calibration standards must be well known. In practice, the internal routine of VNAs uses simple models defined by several coefficients for each standard [83]. The coefficients of the four standards and the RF probes must be well defined in the VNA calibration kit for SOLT calibration. The accuracy of calibration significantly depends on the accuracy of these coefficients. Appendix E provides a table of the calibration coefficients for Cascade RF infinity probe with 100 μ m pitch size and Cascade ISS 101-190 with 1 pico-second delay. VNA8510C can store two CalKits in the system. The Calibration coefficients can be loaded into VNA system from a floppy disk that came with calibration standards, or manually entered into VNA following the steps in Appendix E. The 12 error terms

determined from SOLT calibration can be saved as a CalSet. Below are the headlines of a CalSet file,

```
CITIFILE A.01.01
#NA VERSION HP8510C.07.16
NAME CAL_SET
#NA REGISTER 5
VAR FREQ MAG 93
DATA E[1] RI
DATA E[2] RI
DATA E[3] RI
DATA E[4] RI
DATA E[5] RI
DATA E[6] RI
DATA E[7] RI
DATA E[8] RI
DATA E[9] RI
DATA E[10] RI
DATA E[11] RI
DATA E[12] RI
#NA SWEEP_TIME 1.839999E-1
#NA POWER1 -2.5E1
#NA POWER2 -2.5E1
#NA PARAMS 30
#NA CAL_TYPE 5
#NA DOMAIN_TYPE 0
#NA POWER_SLOPE 0.0E0
#NA POWER_SLOPE2 0.0E0
... ..
```

Fig. 3.17 show S_{11} and S_{21} of a 0.13 μm NMOS transistor for 2-110 GHz. Raw data is the measured S-parameters without any error calibration. Corrected data is the data with system error calibrated using SOLT calibration. For parameter extraction and device modeling, both real part and imaginary part of the S-parameters are important. Error correction is necessary at all frequencies.

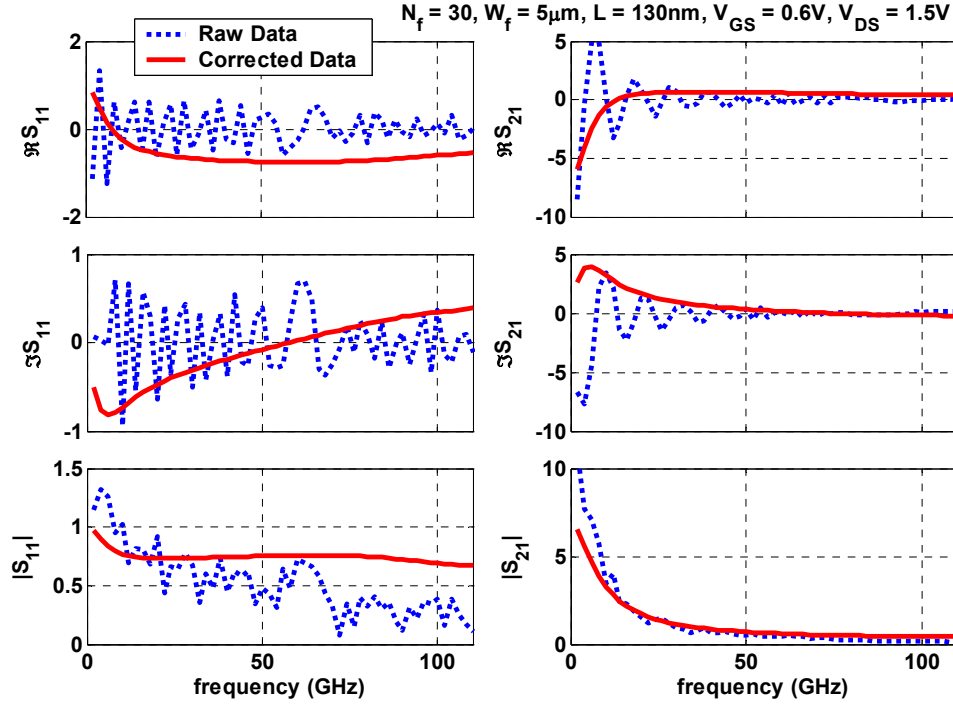


Fig. 3.17 Raw and corrected data for S_{11} and S_{21} of a $0.13\mu\text{m}$ NMOS transistor.

3.5 The most complete 16-term error model

The most complete mathematic model for a four-port network is 16-term model, since four-port network is essentially a 4×4 matrix. Fig. 3.18 shows the signal flow graph of the four-port error adaptor, containing 16 error terms. The reflection at each port contributes four error terms including two directivity errors (e_{00} and e_{33}) and two port match errors (e_{11} and e_{22}). The transmission from measurement ports to DUT terminals introduces four frequency response error terms; e_{10} , e_{01} , e_{32} and e_{23} . The coupling between the four ports adds eight leakage error terms marked with dash lines in Fig. 3.18. When the couplings are negligible, it will be reduced to the 8-term model in Section 3.3.

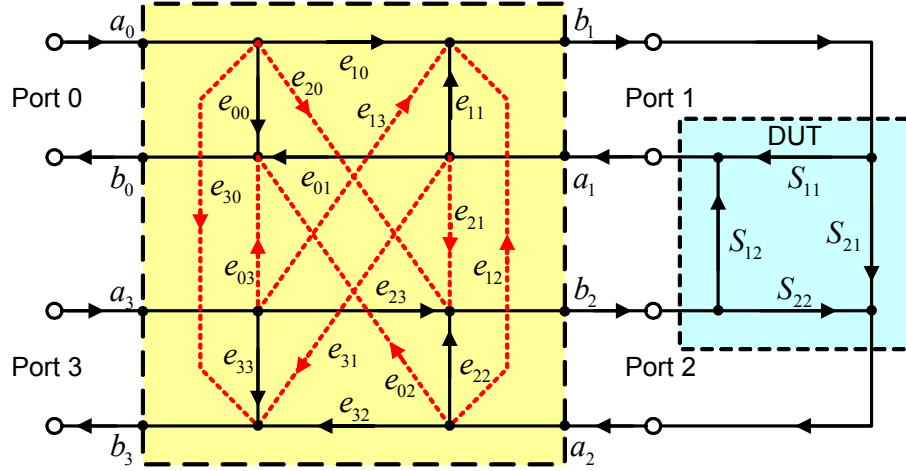


Fig. 3.18 Signal flow graph of the 16-term model for a two-port system.

The 16 error terms are actually the S-parameters of the four-port network, which can be defined using the incident and reflected waves at each port as

$$\begin{bmatrix} b_0 \\ b_3 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} e_{00} & e_{03} & | & e_{01} & e_{02} \\ e_{30} & e_{33} & | & e_{31} & e_{32} \\ e_{10} & e_{13} & | & e_{11} & e_{12} \\ e_{20} & e_{23} & | & e_{21} & e_{22} \end{bmatrix} \begin{bmatrix} a_0 \\ a_3 \\ a_1 \\ a_2 \end{bmatrix}, \quad (3.17)$$

For simplicity, the above expression is rewritten using 2×2 matrices as

$$\begin{bmatrix} b^M \\ b^{DUT} \end{bmatrix} = \begin{bmatrix} E_1 & E_3 \\ E_2 & E_4 \end{bmatrix} \begin{bmatrix} a^M \\ a^{DUT} \end{bmatrix}. \quad (3.18)$$

E_1 , E_2 , E_3 , and E_4 are 2×2 matrices defined as

$$E_1 = \begin{bmatrix} e_{00} & e_{03} \\ e_{30} & e_{33} \end{bmatrix}, E_2 = \begin{bmatrix} e_{10} & e_{13} \\ e_{20} & e_{23} \end{bmatrix}, E_3 = \begin{bmatrix} e_{01} & e_{02} \\ e_{31} & e_{32} \end{bmatrix}, E_4 = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix}. \quad (3.19)$$

The vectors b^M , a^M , b^{DUT} and a^{DUT} are 2×1 wave vectors defined at the perfect VNA side (Port 0 and Port 3) and the DUT side (Port 1 and Port 2).

$$b^M = \begin{bmatrix} b_0 \\ b_3 \end{bmatrix}, b^{DUT} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, a^M = \begin{bmatrix} a_0 \\ a_3 \end{bmatrix}, a^{DUT} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \quad (3.20)$$

Based on the directions of the waves in Fig. 3.18, the S-parameters measured by the VNA, S^M , and the S-parameters of the unknown DUT, S^{DUT} , are defined as

$$\begin{bmatrix} b_0 \\ b_3 \end{bmatrix} = S^M \begin{bmatrix} a_0 \\ a_3 \end{bmatrix}, \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = S^{DUT} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad (3.21)$$

i.e. $b^M = S^M a^M$ and $a^{DUT} = S^{DUT} b^{DUT}$. Thus, S^M and S^{DUT} can be related through a nonlinear equation in terms of E as

$$S^M = E_1 + E_3 \left[\left(S^{DUT} \right)^{-1} - E_4 \right]^{-1} E_2, \quad (3.22)$$

or

$$S^{DUT} = \left[E_2 \left(S^M - E_1 \right)^{-1} E_3 + E_4 \right]^{-1}. \quad (3.23)$$

It can also be written as

$$\left(E_3 - E_1 E_2^{-1} E_4 \right) S^{DUT} + S^M \left(E_2^{-1} E_4 \right) S^{DUT} + E_1 E_2^{-1} - S^M E_2^{-1} = [0]_{2 \times 2}. \quad (3.24)$$

It is difficult to solve E from the nonlinear relationship in (3.22) and (3.23). However, using transmission parameters (T-parameters), S^M and S^{DUT} can be related through a linear relation in terms of error matrix [35]. In that case, error terms can be solved using linear algebra algorithms [36]. (3.18) can be rewritten using T-parameters as

$$\begin{bmatrix} b^M \\ a^M \end{bmatrix} = \begin{bmatrix} T_1 & T_3 \\ T_2 & T_4 \end{bmatrix} \begin{bmatrix} a^{DUT} \\ b^{DUT} \end{bmatrix}. \quad (3.25)$$

where

$$T_1 = \begin{bmatrix} t_1 & t_5 \\ t_2 & t_6 \end{bmatrix}, T_2 = \begin{bmatrix} t_3 & t_7 \\ t_4 & t_8 \end{bmatrix}, T_3 = \begin{bmatrix} t_9 & t_{13} \\ t_{10} & t_{14} \end{bmatrix}, T_4 = \begin{bmatrix} t_{11} & t_{15} \\ t_{12} & t_{16} \end{bmatrix}. \quad (3.26)$$

Recall that $b^M = S^M a^M$ and $a^{DUT} = S^{DUT} b^{DUT}$, (3.25) can be rewritten as [36]

$$T_1 S^{DUT} - S^M T_2 S^{DUT} + T_3 - S^M T_4 = [0]_{2 \times 2}. \quad (3.27)$$

This is equivalent to [35]

$$S^M = (T_1 S^{DUT} + T_3) (T_2 S^{DUT} + T_4)^{-1}, \quad (3.28)$$

or

$$S^{DUT} = (T_1 - S^M T_2)^{-1} (S^M T_4 - T_3). \quad (3.29)$$

Comparing (3.27) and (3.24), the elements in E and T can be related through

$$\begin{aligned} T_1 &= E_3 - E_1 E_2^{-1} E_4 & E_1 &= T_3 T_4^{-1} \\ T_2 &= -E_2^{-1} E_4 & E_2 &= T_4^{-1} \\ T_3 &= E_1 E_2^{-1} & E_3 &= T_1 - T_3 T_4^{-1} T_2 \\ T_4 &= E_2^{-1} & E_4 &= -T_4^{-1} T_2 \end{aligned} \quad \text{and} \quad (3.30)$$

Since the matrices in (3.27) are 2×2 matrices, each two-port measurement will give four linear equations in terms of T . Four calibration standard measurements seem to give enough linear equations to solve the 16 elements in T , but in fact this is not true. Only 14 parameters can be solved by making four measurements for two reasons [41]. First of all, the set of equations is homogeneous, and the maximum number of nonzero unknowns can be solved is 15, because the only possible solution will be an all zero solution if the coefficient matrix is full rank. Therefore, 15 error terms can be solved as a function of the 16th no matter how many standards are measured. Secondly, because of the singularity conditions, besides the freely chosen 16th parameter, one error term

remains unknown, and it can be solved using the fifth measurement. Numerical examples in Appendix G show that the set of equations is ill-conditioned for any four passive standards.

The previous 8-term model and 12-term model can also be represented using the four-port network as they are actually describing the same set of systematic errors. The 8-term model is just a special case of 16-term model with negligible leakage terms. Only directivity, port match, and frequency response terms are considered in 8-term model. The leakage terms e_{30} and e_{03} in Fig. 3.13 and Fig. 3.15, can be added to the signal flow graph, which increases the number of error terms to 10, and can be determined individually using LOAD standard. Thus the 4×4 error matrix for 8(10)-term model is defined as

$$\begin{bmatrix} b_0 \\ b_3 \\ b_1 \\ b_2 \end{bmatrix} = \left[\begin{array}{cc|cc} e_{00} & 0(e_{03}) & e_{01} & 0 \\ 0(e_{30}) & e_{33} & 0 & e_{32} \\ e_{10} & 0 & e_{11} & 0 \\ 0 & e_{23} & 0 & e_{22} \end{array} \right] \begin{bmatrix} a_0 \\ a_3 \\ a_1 \\ a_2 \end{bmatrix}. \quad (3.31)$$

The 12-term error model in Section 3.4 is equivalent to two error matrices, one for forward mode, one for reverse mode [87]. For forward mode, a_3 is not available, so the matrix becomes

$$\begin{bmatrix} b_0 \\ b_3 \\ b_1 \\ b_2 \end{bmatrix} = \left[\begin{array}{cc|cc} e_{00} & & e_{01} & e_{02} \\ e_{30} & & e_{31} & e_{32} \\ e_{10} & & e_{11} & e_{12} \\ e_{20} & & e_{21} & e_{22} \end{array} \right] \begin{bmatrix} a_0 \\ \text{---} \\ a_1 \\ a_2 \end{bmatrix}. \quad (3.32)$$

For reverse mode, it is

$$\begin{bmatrix} b'_0 \\ b'_3 \\ b'_1 \\ b'_2 \end{bmatrix} = \left[\begin{array}{cc|cc} e'_{03} & & e'_{01} & e'_{02} \\ & e'_{33} & e'_{31} & e'_{32} \\ e'_{13} & & e'_{11} & e'_{12} \\ e'_{23} & & e'_{21} & e'_{22} \end{array} \right] \begin{bmatrix} a'_3 \\ a'_1 \\ a'_2 \end{bmatrix}. \quad (3.33)$$

There are 12 error terms in forward mode, and 12 error terms in reverse mode if all leakages are considered. It is published in 1997 as a 22-term model, because only 11 of the 12 error terms can be solved for either forward or reverse mode [42]. Six standards will be wanted to solve this 22-term model [42]. As long as two separate error matrices are used for forward and reverse mode, switch error is naturally removed as discussed in Section 3.4.

3.6 Error adaptor for single-step calibration

Not only systematic errors, but also on-wafer parasitics can be described as a four-port network. On-wafer parasitics are the probing pads, and interconnect lines leading to the device terminals, which actually connects the two ports at the two signal pads, Port 1 and Port 2, to the two ports at the gate and drain of the desired CMOS transistor. That essentially defines a four-port network between the two probes and the two transistor terminals. Fig. 1.9 (a) and (b) show the two four-port networks for systematic errors and on-wafer parasitics. The four-port relations derived in Section 3.5 do not make any assumption about the properties of the four-port network. So, the same equations can be applied on system error four-port or on-wafer parasitics four-port. Since the two ports at the probe tips are shared by the two four-port networks, it is possible to combine the two four-port networks into one. Fig. 3.19 shows the combined four-port network. The technique that solves the combined four-port network between the perfect VNA and the

transistor terminals using on-wafer standards is called “single-step calibration”. The error models and calibration techniques discussed above can be applied without modification. However, single-step calibration is not widely used in the past because of traceability issue and less accurate on-wafer standards compared with ISS standards [34] [86].

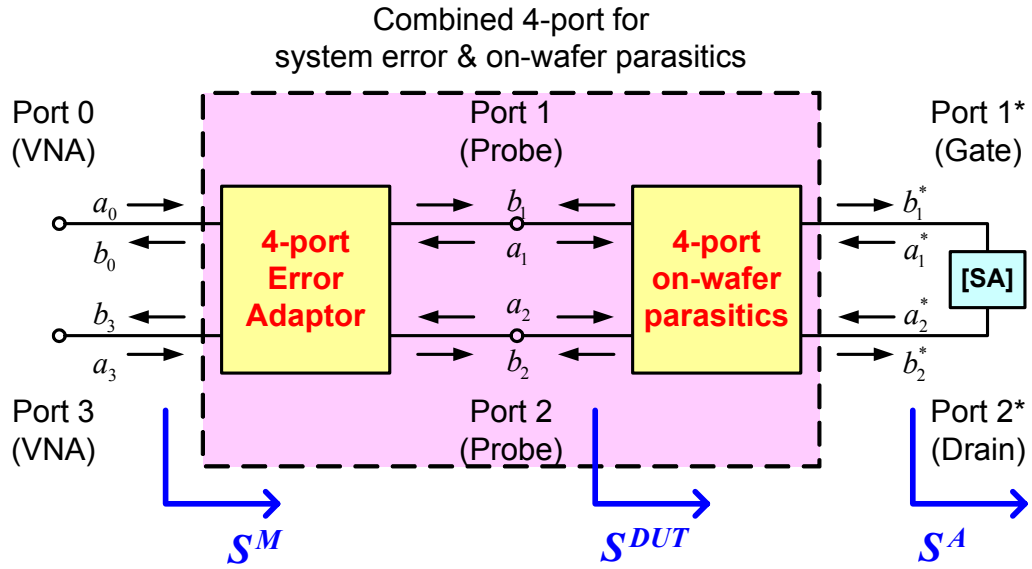


Fig. 3.19 The combined four-port network including system errors and on-wafer parasitics.

For transistor characterization purpose, S-parameters are usually measured on a large number of transistors, which may take hours or days. For two-step calibration, the accuracy of ISS calibration need to be rechecked frequently as systematic errors may drift during hourly measurements, e.g. temperature changes. This is time consuming and requires a manual switch of the test wafer and the ISS substrate. This problem is naturally solved with single-step calibration. However, as mentioned in Section 1.3.4,

there are two problems need to be solved. The first is how to verify the accuracy of the results. The second is how to model the non-ideal parasitics of on-wafer standards.

This dissertation uses two-step calibration results as a reference to evaluate the accuracy of single-step calibration. Since the verification only need to be done for several reference test structures before large amount of measurements, it can still reduce the time for measurement and help automation of large volume measurements. The non-ideal parasitics of on-wafer standards is also examined in this dissertation. First of all, OPEN, SHORT standards can be assumed to be ideal from the experimental results in Section 4.2. The same assumption is applied in on-wafer de-embedding step for two-step calibration. Secondly, the length of on-wafer THRU is much shorter than the THRU on ISS substrate, because the dimension of the transistor is usually much less than the distance between the two signal pads. Thirdly, on-wafer resistor standard can be modeled using a similar mathematical model as ISS calibration does. The parasitics of on-wafer resistor can be lumped as a parallel capacitance whose value is determined from low frequency measurement. Since the parasitic capacitance will not drift a lot for a fixed process, the value just need to be checked once for one process.

Fig. 3.19 shows the two four-port error adapters for systematic errors, on-wafer parasitics, and the four-port network combining systematic errors and on-wafer parasitics. Note that the direction of the a and b waves at the probes are defined differently for the four-port error adapter and the four-port on-wafer parasitics, to keep the rules that all a waves are incident waves entering the four-port, and all of the b waves are the reflected waves leaving the four-port. For simplicity, the following S-parameters are defined.

1. S^M is the measured S-parameter of the unknown two-port without switch error.
2. S^{DUT} is the measured S-parameter of the unknown two-port after ISS calibration. The on-wafer parasitics, probing pads and interconnects is still involved in S^{DUT} .
3. S^A is the actual S-parameter of the unknown two-port without system errors and on-wafer parasitics, which means the S-parameter after two-step calibration or single-step calibration.

S^M , and S^A can be easily defined using waves with directions shown in Fig. 3.19 as

$$\begin{bmatrix} b_0 \\ b_3 \end{bmatrix} = S^M \begin{bmatrix} a_0 \\ a_3 \end{bmatrix}, \begin{bmatrix} a_1^* \\ a_2^* \end{bmatrix} = S^A \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} \quad (3.34)$$

It is a little complicated to defining S^{DUT} . When applied for systematic error calibration,

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = S^{DUT} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad (3.35)$$

since b_1, b_2 are the incident waves to DUT, and a_1, a_2 are the reflected waves to DUT.

When applied for on-wafer parasitics de-embedding

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = S^{DUT} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad (3.36)$$

because, now, b_1, b_2 leave DUT, and a_1, a_2 enter DUT. When applied for single-step calibration, it does not matter because S^{DUT} do not show up in the calibration procedures.

3.7 Summary

Error adaptor concept for two-port S-parameter measurement is introduced. The error adaptor is a fictitious linear network that is inserted between the measurement ports and the unknown two-port. For two-port measurement, the error adaptor is a four-port network, which is described using a 4×4 matrix or 16 error terms, since there are four waves at the two measurements ports and four waves at the DUT terminals. The 16 term error model is the most complete error model for two-port S-parameter measurement. 8-term and 12-term error model can be viewed as special cases of 16-term. The advantage of 12-term model is that switch error is naturally removed because the two error adaptors for forward and reverse mode are completely separated. Thus, 12-term model can be applied on three-receiver VNA and four-receiver VNA. SOLT calibration is based on the 12-term error model, and implanted in all modern VNAs. For high frequency applications, especially when the leakage errors are not negligible when compared with other error terms, 16-term error model is needed. The complete 16 error model can be used to describe both system errors and on-wafer parasitics since both of them are four-port networks. Single-step calibration combines the two four-port networks into one, with two ports inside VNA and two ports at the transistor terminals. When on-wafer standards are available, systematic errors and on-wafer parasitics can be removed in a single step.

CHAPTER 4

GENERIC ANALYTICAL FOUR-PORT SOLUTION

As the operating frequency increases, the distributive nature of on-wafer parasitics becomes significant. The de-embedding techniques based on lumped equivalent circuit for probing pads and interconnections fail, including open-short, pad-open-short, and three-step in Appendix C. The distributive nature of on-wafer parasitics is naturally accounted for by describing the on-wafer parasitics as a four-port network, i.e. a 4×4 matrix. The four-port network is located between the two external ports at the two probe tips and the two internal ports at the two-port device terminals [20]. This four-port parasitics network was shown to be solvable using five on-wafer standards [20] [19]. These solutions, however, are complicated and involve taking square roots, and thus choice of positive and negative signs. Furthermore, the solution in [20] does not give insight into the relationship between open-short and four-port solutions, while the solution in [19] cannot be applied for single-step calibration. The solution developed below retains the open-short relation of [17], is much simpler mathematically than both [20] and [19], does not involve taking square root, and is applicable to both two-step and single-step calibration. All of these improvements are achieved without loss of accuracy.

One of the standards used is an on-wafer load resistor, which was assumed to be ideal in [20] and [19], but always has parasitics in reality. The reciprocal and symmetric

four-port solution in [17] showed that the parasitic capacitance associated with this load resistor can affect the de-embedding results for on-wafer inductor measurements. In this solution, we first determine the parasitic capacitance of the load resistor using low frequency open-short de-embedding, e.g. below 30 GHz, and then include its effect in four-port de-embedding procedures.

The relationship between open-short de-embedding and four-port de-embedding derived in [17] is further examined using two matrices of the general four-port solution, which reduce to identity matrices at low frequencies where open-short is valid. New criteria for examining reciprocity and symmetry of the solved four-port network are developed. Using a reciprocal and symmetric solution, four-port de-embedding and pad-open-short de-embedding were previously shown to be close for inductors, and pad-open-short was concluded to be superior to four-port due to better tolerance to parasitic capacitance in [17]. We examine this issue for transistor measurements and show that these conclusions cannot be generalized, at least to this experiment. Instead, pad-open-short gives inaccurate results at high frequencies that are close to open-short.

This chapter details the derivation of an analytical four-port solution for on-wafer parasitics using Y-parameters. With five on-wafer standards, OPEN, SHORT, LEFT, RIGHT, and THRU, the 16 error terms in Y-format can be determined. Experimental results are presented and compared with de-embedding methods using lumped equivalent circuits on 0.13 μ m RF CMOS technology. An indicator to quantify the validity of open-short de-embedding is given.

4.1 Four-port network in Y-parameters

Fig. 4.1 illustrates the four-port description of on-wafer parasitics using port currents and voltages. Port 1 and port 2 are formed by the two probe tips, i.e. the two GSG pads of the whole DUT. Port 1* and 2* are terminated at the two terminals of the two-port device, e.g. the gate and drain of the examined NMOS transistor. We define current and voltage vectors, I_e , V_e , I_i , and V_i as follows:

$$V_e = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, I_e = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, V_i = \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix}, I_i = \begin{bmatrix} I_1^* \\ I_2^* \end{bmatrix}. \quad (4.1)$$

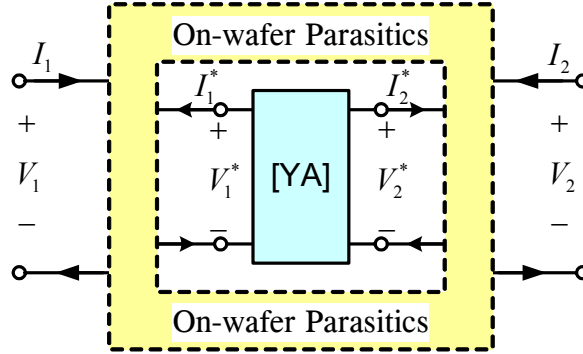


Fig. 4.1. Block diagram of the on-wafer parasitics four-port network using I-V representation.

The subscript e means external, while the subscript i means internal. These voltage and current vectors can be related through four 2×2 admittance matrices, Y_{ee} , Y_{ei} , Y_{ie} , and Y_{ii} as:

$$\begin{bmatrix} I_e \\ I_i \end{bmatrix} = \begin{bmatrix} Y_{ee} & Y_{ei} \\ Y_{ie} & Y_{ii} \end{bmatrix} \begin{bmatrix} V_e \\ V_i \end{bmatrix}. \quad (4.2)$$

Denoting the two-port Y-parameters of the whole DUT as Y^{DUT} and the actual two-port Y-parameters of the intrinsic transistor as Y^A , we have $I_e = Y^{DUT} V_e$ and $I_i = -Y^A V_i$. Y^A can then be related to Y^{DUT} as [20]:

$$Y^{DUT} = Y_{ee} - Y_{ei} \left(Y^A + Y_{ii} \right)^{-1} Y_{ie}, \quad (4.3)$$

or

$$Y^A = -Y_{ii} - Y_{ie} \left(Y^{DUT} - Y_{ee} \right)^{-1} Y_{ei}. \quad (4.4)$$

The 16 unknowns in Y_{ee} , Y_{ei} , Y_{ie} , and Y_{ii} can be determined by measuring at least four on-wafer standards with known Y^A since each measurement gives four equations. Actually five on-wafer standards are necessary when the standards are combinations of open, short, matched load, in addition to a through line. Once Y_{ee} , Y_{ei} , Y_{ie} , and Y_{ii} are known, the actual Y-parameters Y^A of any transistor, can be easily retrieved from the measured Y^{DUT} .

4.2 General four-port Solution

4.2.1 Relationship between open-short and four-port

Substituting the Y-parameters of an ideal OPEN and an ideal SHORT into (4.3), i.e. $Y^{A,open} = [0]_{2 \times 2}$ and $\left(Y^{A,short} \right)^{-1} = [0]_{2 \times 2}$, the measured Y-parameters of OPEN and SHORT can be obtained as [20]:

$$Y^{DUT,open} = Y_{ee} - Y_{ei} \left(Y_{ii} \right)^{-1} Y_{ie}, \quad (4.5)$$

$$Y^{DUT,short} = Y_{ee}. \quad (4.6)$$

Note that ideal OPEN and SHORT are used in all analytical de-embedding methods to achieve an analytical solution. The equivalent two-port networks of ideal OPEN and SHORT standards are shown in Fig. 4.2 (b) and (c). The SHORT measurement directly yields Y_{ee} . However, solving Y_{ei} , Y_{ie} and Y_{ii} proves difficult, because of the nonlinear relationship between Y^{DUT} and Y^A due to matrix inversion and multiplication.

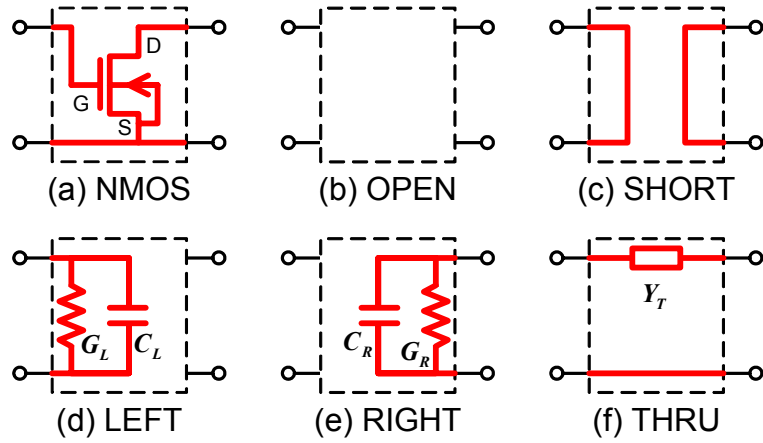


Fig. 4.2. The equivalent two-port network of the intrinsic NMOS transistor and the five on-wafer standards OPEN, SHORT, LEFT, RIGHT and THRU.

Recall that the open-short de-embedded Y-parameters Y^{OS} is given by [14]:

$$Y^{OS} = \left[\left(Y^{DUT} - Y^{DUT,open} \right)^{-1} - \left(Y^{DUT,short} - Y^{DUT,open} \right)^{-1} \right]^{-1}. \quad (4.7)$$

Substituting (4.3), (4.5) and (4.6) into (4.7) leads to a simple relationship between Y^{OS} and Y^A [17]:

$$Y^{OS} = Y_{ei} \left(Y_{ii} \right)^{-1} Y^A \left(Y_{ii} \right)^{-1} Y_{ie}. \quad (4.8)$$

Derivation details can be find in Appendix F. Denoting $A = Y_{ei} (Y_{ii})^{-1}$ and $B = (Y_{ii})^{-1} Y_{ie}$,

(4.8) can be rewritten as

$$Y^{OS} = AY^AB, \quad (4.9)$$

or

$$Y^A = A^{-1}Y^{OS}B^{-1}. \quad (4.10)$$

A and B are 2×2 matrices, which relate to the Y-parameters of the four-port error adaptor through

$$A = Y_{ei} (Y_{ii})^{-1} = \frac{1}{\det(Y_{ii})} \begin{bmatrix} y_{11}^{ei}y_{22}^{ii} - y_{12}^{ei}y_{21}^{ii} & y_{12}^{ei}y_{11}^{ii} - y_{11}^{ei}y_{12}^{ii} \\ y_{21}^{ei}y_{22}^{ii} - y_{22}^{ei}y_{21}^{ii} & y_{22}^{ei}y_{11}^{ii} - y_{21}^{ei}y_{12}^{ii} \end{bmatrix}, \quad (4.11)$$

$$B = (Y_{ii})^{-1} Y_{ie} = \frac{1}{\det(Y_{ii})} \begin{bmatrix} y_{11}^{ie}y_{22}^{ii} - y_{21}^{ie}y_{12}^{ii} & y_{12}^{ie}y_{22}^{ii} - y_{22}^{ie}y_{12}^{ii} \\ y_{21}^{ie}y_{11}^{ii} - y_{11}^{ie}y_{21}^{ii} & y_{22}^{ie}y_{11}^{ii} - y_{12}^{ie}y_{21}^{ii} \end{bmatrix}. \quad (4.12)$$

y_{mn}^{ei} , y_{mn}^{ie} , and y_{mn}^{ii} are (m, n) elements of the 2×2 matrices Y_{ei} , Y_{ie} , and Y_{ii} , $m, n = 1, 2$.

Expanding the matrices in (4.9), the elements of open-short de-embedded Y-parameters are

$$Y^{OS} = \begin{bmatrix} a_{11}Y_{11}^Ab_{11} + a_{12}Y_{21}^Ab_{11} + a_{11}Y_{12}^Ab_{21} + a_{12}Y_{22}^Ab_{21} & a_{11}Y_{11}^Ab_{12} + a_{12}Y_{21}^Ab_{12} + a_{11}Y_{12}^Ab_{22} + a_{12}Y_{22}^Ab_{22} \\ a_{21}Y_{11}^Ab_{11} + a_{22}Y_{21}^Ab_{11} + a_{21}Y_{12}^Ab_{21} + a_{22}Y_{22}^Ab_{21} & a_{21}Y_{11}^Ab_{12} + a_{22}Y_{21}^Ab_{12} + a_{21}Y_{12}^Ab_{22} + a_{22}Y_{22}^Ab_{22} \end{bmatrix}. \quad (4.13)$$

a_{ij} and b_{ij} are (i, j) elements of A and B , $i, j = 1, 2$.

Instead of directly solving the 16 unknowns in Y_{ee} , Y_{ei} , Y_{ie} , and Y_{ii} , as was done in [20], only the 8 elements in A and B need to be solved after performing open-short de-embedding [17]. Strictly speaking, only 15 of the 16 unknowns can be solved, due to the ratio nature of S-parameter measurements, similar to the situation in 16-term error calibration [41] [80]. For the same reason, only 7 of the 8 unknowns in A and B can be

fully solved, which is sufficient for de-embedding purpose [19]. Three additional on-wafer standards, LEFT, RIGHT, and THRU, are used in this dissertation to find out the 8 (7 solvable) unknowns left after open-short de-embedding.

4.2.2 Open-short de-embedded LEFT, RIGHT, and THRU

We now examine the three additional standards, LEFT, RIGHT, and THRU, as illustrated in Fig. 4.2 (d)-(f). The Y-parameters for actual LEFT, RIGHT, and THRU standards are modeled by:

$$Y^{A,left} = \begin{bmatrix} Y_L & 0 \\ 0 & 0 \end{bmatrix}, Y^{A,right} = \begin{bmatrix} 0 & 0 \\ 0 & Y_R \end{bmatrix}, Y^{A,thru} = \begin{bmatrix} Y_T & -Y_T \\ -Y_T & Y_T \end{bmatrix}. \quad (4.14)$$

Note that the on-wafer load resistor in LEFT and RIGHT, which are assumed to be purely resistive in [20] and [19], are represented as Y_L and Y_R to account for non-idealities of on-wafer resistors. The primary non-ideality is a parallel capacitance, as shown by their open-short de-embedded Y-parameters at relatively low frequencies where open-short is accurate. Thus Y_L and Y_R are modeled as $Y_L = G_L + j\omega C_L$ and $Y_R = G_R + j\omega C_R$ as shown in Fig. 4.2 (d) and (e). The admittance and parasitic capacitance, G_L , G_R , C_L , and C_R , are extracted from open-short de-embedded LEFT and RIGHT below 30 GHz. If high precision low parasitics resistors are used, which are increasingly available in RF SiGe BiCMOS and RF CMOS processes, one may determine G_L and G_R from DC measurements and neglect C_L and C_R . The Y_T in $Y^{A,thru}$ becomes infinity for an ideal through line with zero length. A small length THRU is typically used in transistor measurement to allow signal propagation from

input to output. As a result, the s and t terms used to represent the non-ideal THRU in [19] are close, thus only a single Y_T term is used here, which helps to considerably simplify the general four-port solution and make the new solution applicable to single-step calibration. Y_T does not need to be known as it will be cancelled out during de-embedding.

4.2.3 Analytical solution of A and B

The open-short de-embedded Y-parameters of LEFT and RIGHT, $Y^{OS, left}$ and $Y^{OS, right}$, can be related to elements of A and B by substituting $Y^{A, left}$ and $Y^{A, right}$ in (4.14) into (4.9). Both $Y^{A, left}$ and $Y^{A, right}$ have 3 zero elements, thus the final product of AY^AB only contains simple product of the elements in A and B . For convenience, we use M and N defined below instead of $Y^{OS, left}$ and $Y^{OS, right}$:

$$M = Y^{OS, left} / Y_L = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} \\ a_{21}b_{11} & a_{21}b_{12} \end{bmatrix}, \quad (4.15)$$

$$N = Y^{OS, right} / Y_R = \begin{bmatrix} a_{12}b_{21} & a_{12}b_{22} \\ a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}, \quad (4.16)$$

where a_{ij} , b_{ij} , M_{ij} , and N_{ij} are the (i, j) elements of A , B , M , and N , $i, j = 1, 2$. Note that M and N are known matrices for the following procedures.

At first glance, one may attempt to solve the 8 elements of A and B from the 8 equations provided by LEFT and RIGHT (4 each in (4.15) and (4.16)). This, however, is not the case, as only three of the 4 equations provided by each measurement are independent. For example, the ratios of M_{21}/M_{11} and M_{22}/M_{12} both give a_{21}/a_{11} .

Thus, only three unknowns can be solved as a function of the fourth unknown for a LEFT or a RIGHT measurement.

Using the relationship between the elements of M and N and the unknowns in A and B , some of the unknowns can be solved first. To make the solution easier and clearer, A and B are normalized to A' and B' using $A' = kA$, $B' = k^{-1}B$ by a constant k . As we still have $Y^{OS} = A'Y^AB'$ and $Y^A = (A')^{-1}Y^{OS}(B')^{-1}$, we can replace A and B by A' and B' respectively for de-embedding purpose.

The normalization factor k is chosen based on multiple considerations. First, it must not affect the accuracy of the de-embedded results. Second, the errors remaining after open-short can be easily examined from the elements of the normalized matrices. Third, it will reduce to unity if the four-port network is reciprocal. A choice satisfying these requirements is $k = \sqrt{b_{11}/a_{11}}$:

$$A' = kA = \sqrt{b_{11}/a_{11}}A = \sqrt{a_{11}b_{11}} \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} \\ \frac{a_{21}}{a_{11}} & \frac{a_{22}}{a_{11}} \end{bmatrix} \text{ and} \quad (4.17)$$

$$B' = k^{-1}B = \sqrt{a_{11}/b_{11}}B = \left(\sqrt{a_{11}b_{11}}\right)^{-1} \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} \end{bmatrix}. \quad (4.18)$$

After normalization, there are only 7 elements that need to be solved in A' and B' . We first solve as many terms of A' and B' as possible from M and N , using (4.15) and (4.16):

$$a_{11}b_{11} = M_{11}, \quad a_{11}b_{12} = M_{12}, \quad \text{and} \quad \frac{a_{21}}{a_{11}} = \frac{M_{21}}{M_{11}}, \quad (4.19)$$

$$a_{11}b_{21} = N_{21} \frac{a_{11}}{a_{22}}, \quad a_{11}b_{22} = N_{22} \frac{a_{11}}{a_{22}}, \quad \text{and} \quad \frac{a_{12}}{a_{11}} = \frac{N_{12}}{N_{22}} \frac{a_{22}}{a_{11}}. \quad (4.20)$$

6 of the 7 elements are now solved as functions of the 7th, a_{22}/a_{11} , which we define as

$$\lambda \triangleq a_{22}/a_{11}.$$

For a given set of measured data, a_{21}/a_{11} can be calculated in two ways, either as M_{21}/M_{11} or as M_{22}/M_{12} from (4.15). The analysis below will show that a_{21}/a_{11} calculated from M_{21}/M_{11} gives better error tolerance. Assuming the actual LEFT is not ideal, there will be small error term ε added to $Y^{A, \text{left}}$ as

$$Y^{A, \text{left}} = \begin{bmatrix} Y_L & \varepsilon \\ \varepsilon & \varepsilon \end{bmatrix}. \quad (4.21)$$

The M matrix, which involves the measurement errors in the LEFT measurement and the calculation errors during open-short de-embedding, can be written as

$$M = Y^{OS, \text{left}} / Y_L = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} \\ a_{21}b_{11} & a_{21}b_{12} \end{bmatrix} + \begin{bmatrix} \Delta & \Delta \\ \Delta & \Delta \end{bmatrix}. \quad (4.22)$$

Δ combines the non-ideality factor ε , the measurement errors and the calculation errors.

The physical nature of the LEFT and RIGHT standards dictates that M_{11} and N_{22} are the largest elements in M and N , and close to one, respectively, as confirmed by measurements. Hence, $\Delta \ll a_{11}b_{11}$, and M_{21}/M_{11} can be calculated as

$$\frac{M_{21}}{M_{11}} = \frac{a_{21}b_{11} + \Delta}{a_{11}b_{11} + \Delta} \approx \frac{a_{21}b_{11}}{a_{11}b_{11}} + \frac{\Delta}{a_{11}b_{11}} = \frac{a_{21}}{a_{11}} + O(\Delta). \quad (4.23)$$

As $O(\Delta)$ is a very small number, M_{21}/M_{11} is relatively accurate even with non-ideal

LEFT structure. However, M_{22}/M_{12} is calculated as

$$\frac{M_{22}}{M_{12}} = \frac{a_{21}b_{12} + \Delta}{a_{11}b_{12} + \Delta} \neq \frac{a_{21}}{a_{11}}. \quad (4.24)$$

Since $a_{11}b_{12}$ and $a_{21}b_{12}$ are close to zero from the physics nature of LEFT, and can be comparable to Δ , M_{22}/M_{12} may give inaccurate a_{21}/a_{11} . A similar situation exists for a_{12}/a_{11} . Hence solutions with M_{11} and N_{22} as denominators should be used to obtain better error tolerance.

Fig. 4.3 plots the real part of de-embedded y_{21} as a function of frequency, from which g_m is extracted. The g_m extracted from the results with $a_{21}/a_{11} = M_{21}/M_{11}$ is much smoother and more accurate than the one extracted from the results with $a_{21}/a_{11} = M_{22}/M_{12}$. The other unknowns are all determined based on the same principle as shown in (4.19) and (4.20). A' and B' can be rewritten using (4.19), (4.20), and $\lambda \triangleq a_{22}/a_{11}$ as:

$$A' = \sqrt{M_{11}} \begin{bmatrix} 1 & \lambda \frac{N_{12}}{N_{22}} \\ \frac{M_{21}}{M_{11}} & \lambda \end{bmatrix}, \quad B' = \frac{1}{\sqrt{M_{11}}} \begin{bmatrix} M_{11} & M_{12} \\ \frac{N_{21}}{\lambda} & \frac{N_{22}}{\lambda} \end{bmatrix}. \quad (4.25)$$

λ is the only unknown left which relates the unknowns solved from the open-short de-embedded LEFT and RIGHT, and can be solved using the THRU standard.

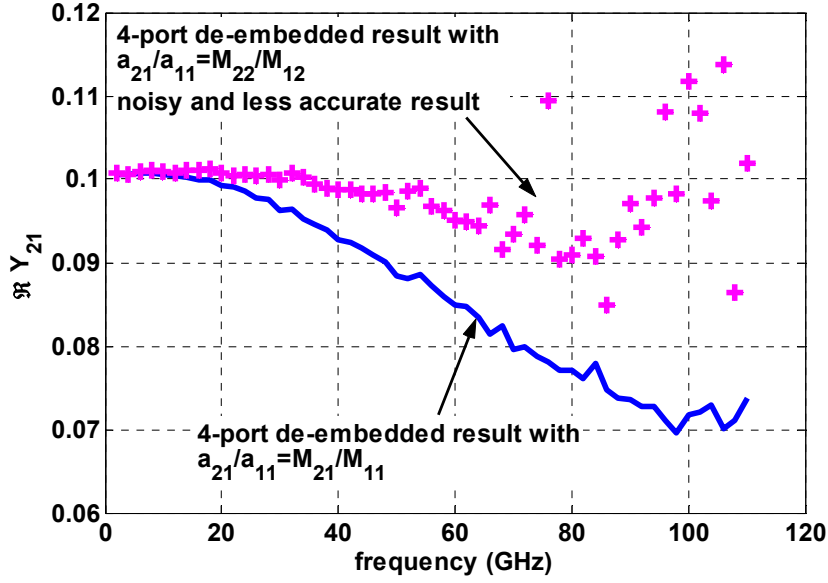


Fig. 4.3. The real part of four-port de-embedded y_{21} using different a_{21}/a_{11} choices. The a_{21}/a_{11} defined from M_{22}/M_{12} is clearly noisier, and should not be used.

The open-short de-embedded Y-parameters of THRU, $Y^{OS,thru}$, can be calculated by substituting $Y^{A,thru}$ in (4.14) into $Y^{OS,thru} = AY^{A,thru}B$ as:

$$Y^{OS,thru} = Y_T \begin{bmatrix} (a_{11} + a_{12})(b_{11} + b_{21}) & -(a_{11} + a_{12})(b_{12} + b_{22}) \\ -(a_{21} + a_{22})(b_{11} + b_{21}) & (a_{21} + a_{22})(b_{12} + b_{22}) \end{bmatrix} \quad (4.26)$$

By taking ratios of the elements of $Y^{OS,thru}$, the equations including λ can be constructed as:

$$\alpha = \frac{y_{21}^{OS,thru}}{y_{11}^{OS,thru}} = \frac{y_{22}^{OS,thru}}{y_{12}^{OS,thru}} = -\frac{a_{21} + a_{22}}{a_{11} + a_{12}} = -\frac{M_{21}/M_{11} + \lambda}{1 + \lambda N_{12}/N_{22}}, \quad (4.27)$$

$$\beta = \frac{y_{12}^{OS,thru}}{y_{11}^{OS,thru}} = \frac{y_{22}^{OS,thru}}{y_{21}^{OS,thru}} = -\frac{b_{12} + b_{22}}{b_{11} + b_{21}} = -\frac{M_{12} + N_{22}/\lambda}{M_{11} + N_{21}/\lambda}. \quad (4.28)$$

Therefore we have four options to solve λ , $y_{21}^{OS,thru} / y_{11}^{OS,thru}$, $y_{22}^{OS,thru} / y_{12}^{OS,thru}$, $y_{12}^{OS,thru} / y_{11}^{OS,thru}$, and $y_{22}^{OS,thru} / y_{21}^{OS,thru}$. The de-embedded transistor Y-parameters are practically the same for all four choices in our experiment. Below, λ is obtained from α as:

$$\lambda = -\frac{\alpha + M_{21} / M_{11}}{1 + \alpha N_{12} / N_{22}}, \quad \alpha = \frac{y_{21}^{OS,thru}}{y_{11}^{OS,thru}}. \quad (4.29)$$

This general four-port solution here is much simpler than that of [19].

4.2.4 Summary of general four-port de-embedding

To summarize, for two-step four-port parasitics de-embedding, the main procedures are:

1. Perform VNA system error calibration using Impedance Standard Substrate (ISS).
2. Measure S-parameters of on-wafer standards and the desired transistor or any two-port DUT. The S-parameters are transformed to Y- and Z-parameters using equations in Appendix B [88].
3. Perform open-short de-embedding on measured LEFT, RIGHT, THRU, and the DUT to obtain $Y^{OS,left}$, $Y^{OS,right}$, $Y^{OS,thru}$, and $Y^{OS,dut}$.
4. Extract G_L , G_R , C_L , and C_R from $Y_{11}^{OS,left}$ and $Y_{22}^{OS,right}$ at low frequencies, e.g. below 30 GHz.
5. Calculate M and N using (4.15) and (4.16).
6. Solve λ from open-short de-embedded THRU, $Y^{OS,thru}$, using (4.29).
7. Find out the elements of A' and B' from M , N , and λ using (4.25).

8. Calculate $Y^{A,dut}$ for the examined transistor using $Y^{A,dut} = (A')^{-1} Y^{OS,dut} (B')^{-1}$.

4.2.5 Impact of non-ideal load in LEFT and RIGHT

In [20] and [19], on-wafer load resistor was assumed to be purely resistive. However, open-short de-embedded Y-parameters of LEFT and RIGHT show that there is parasitic capacitance in parallel with the resistance. The parasitic capacitance C_L and C_R can be extracted from open-short de-embedded LEFT and RIGHT, $y_{11}^{OS,left}$ and $y_{22}^{OS,right}$. The impact of these capacitances is examined by setting $C_L = C_R = 0$ during de-embedding procedures. Fig. 4.4 plots the general four-port de-embedded Y-parameters with and without including C_L and C_R . The transistor Y-parameters are noticeably different, especially for the input admittance $y_{11}^{A,dut}$ and the effective transconductance $\Re\{y_{21}^{A,dut}\}$. This difference indicates that the extracted small signal parameters can be affected, for example, the effective gate resistance extracted using $R_{in} = \Re\{1/Y_{11}^{A,dut}\}$ and the effective gate capacitance extracted using $C_{in} = -1/[2\pi f \Im\{1/Y_{11}^{A,dut}\}]$ [18].

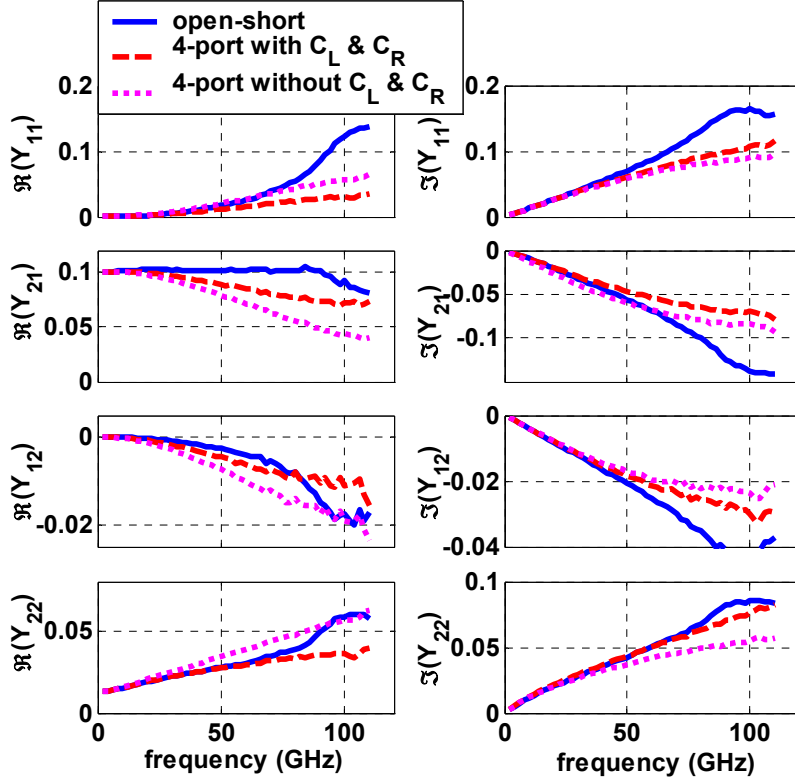


Fig. 4.4. The four-port de-embedded transistor Y-parameters with and without including parasitic capacitance in Y_L and Y_R . For comparison, open-short de-embedded results are also plotted. No reciprocal assumptions are made for four-port parasitics de-embedding.

Fig. 4.5 compares R_{in} and C_{in} extracted from Y_{11} in Fig. 4.4. The R_{in} extracted without C_L and C_R is 2Ω larger than the R_{in} extracted with C_L and C_R . The general four-port R_{in} with C_L and C_R is close to the open-short R_{in} but shows improved frequency dependence. The closeness is expected as open-short is valid below 30 GHz. The C_{in} extracted from open-short gives a very strong and unphysical frequency dependence, while the C_{in} extracted from general four-port is almost frequency independent, no matter C_L and C_R are included or not. In strong inversion, for an oxide

thickness of only a few nanometers, the effective gate capacitance is expected to be approximately constant even at 100 GHz for a MOSFET of such short channel length.

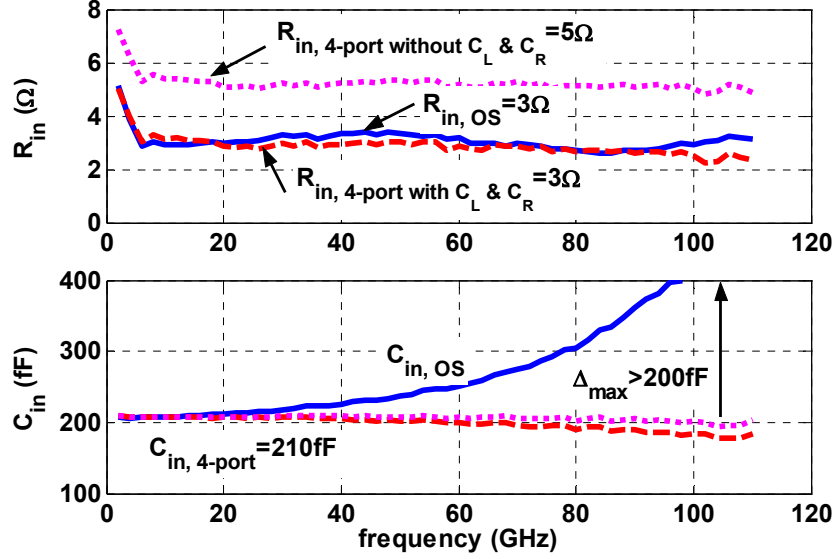


Fig. 4.5. Effective gate resistance and capacitance extracted from four-port de-embedded results with and without parasitic capacitance included in LEFT and RIGHT. Open-short de-embedded results are also shown for comparison. No reciprocal assumptions are made.

4.2.6 Quantifying errors of open-short

By examining the elements in A' and B' , the errors remaining after open-short can be quantified, because $Y^{OS} = A'Y^AB'$. Clearly, only when A' and B' are both identity matrices, open-short will be the same as four-port, i.e. $Y^{OS} = Y^A$. The deviation of A' and B' from identity matrix is thus an indicator of the (in)validity of open-short. Fig. 4.6 plots the real and imaginary parts of the 8 elements in A' and B' . At low frequencies, a'_{11} , b'_{11} , a'_{22} , and b'_{22} are close to unity with zero imaginary part, and all of the other

elements are close to zero in both real and imaginary part. This indicates that A' and B' are both identity matrices and open-short is valid within this frequency range. As frequency goes above 50 GHz, the deviation of A' and B' from identity matrix becomes noticeable and open-short loses its accuracy.

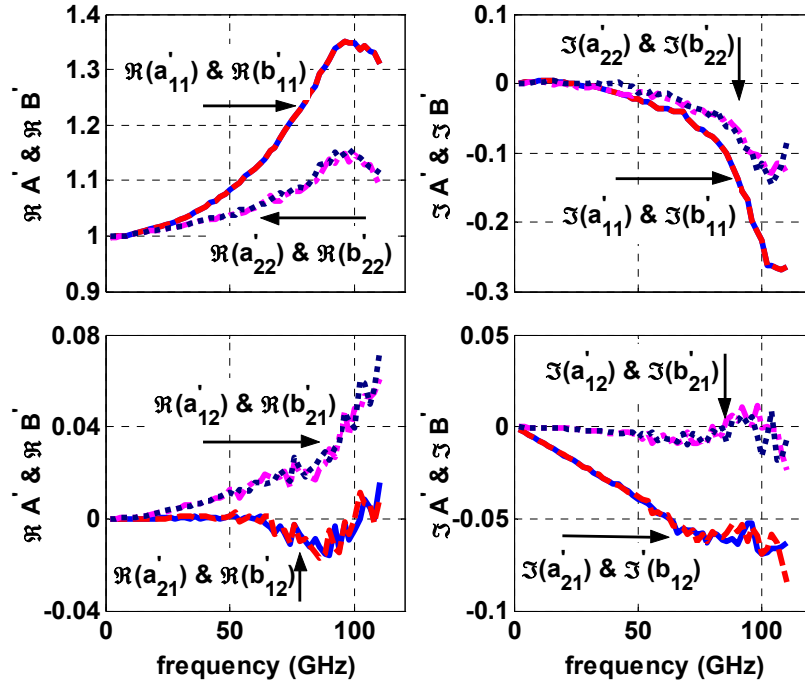


Fig. 4.6. The elements of A' and B' versus frequency.

4.2.7 Reciprocity and symmetry of the four-port parasitics

It was observed in board measurement that non-idealities in the OPEN and SHORT standards can lead to non-reciprocal parameters for passive structures [46]. Ideal OPEN and SHORT, however, are necessary in all de-embedding methods to achieve analytical solution. It is therefore necessary to check if the solved four-port parasitics is still reciprocal or not, and significant deviation from reciprocity would indicate significant

error in the analytical solution. Here we propose a simple criterion to examine reciprocity. From Fig. 4.6, we notice that $a'_{11} = b'_{11}$, $a'_{22} = b'_{22}$, $a'_{12} = b'_{21}$ and $a'_{21} = b'_{12}$, i.e. $A' = (B')^T$. Although $a'_{11} = b'_{11}$ is always true as a result of our choice of normalization, the agreement of the other 3 (6) elements suggests that the solved four-port parasitics is reciprocal and the on-wafer OPEN and SHORT standards may indeed be viewed as ideal. Accordingly, the de-embedded Y-parameters using reciprocal assumption are almost identical to the general four-port results. However, both A' and B' are clearly not symmetric matrices, which is a direct result of our asymmetric layout design necessitated by our choice of the desired reference planes.

4.3 Reciprocal four-port solution and pad-open-short

Reciprocal four-port network means $Y_{ee} = Y_{ee}^T$, $Y_{ei} = Y_{ie}^T$, and $Y_{ii} = Y_{ii}^T$ [20]. Thus, we will have $A = B^T$ and $k = \sqrt{b_{11}/a_{11}} = 1$ from (4.11) and (4.12). Therefore, the number of unknowns can be reduced to 4. All of them can be directly solved from open-short de-embedded LEFT and RIGHT as:

$$A = A' = \begin{bmatrix} \sqrt{M_{11}} & \sqrt{\frac{N_{12}N_{21}}{N_{22}}} \\ \sqrt{\frac{M_{12}M_{21}}{M_{11}}} & \sqrt{N_{22}} \end{bmatrix}, \quad B = B' = A^T, \quad (4.30)$$

because we have $a_{11} = b_{11}$, $a_{22} = b_{22}$, $a_{12} = b_{21}$, and $a_{21} = b_{12}$ in (4.15) and (4.16) and $k = \sqrt{b_{11}/a_{11}} = 1$ in (4.17) and (4.18). The de-embedded results using general four-port solution and reciprocal four-port solution for on-wafer parasitics are very close and can

be viewed as identical. Given that only one THRU structure is saved, we suggest that the general four-port solution to be used, as consistency between reciprocity and ideal OPEN and SHORT can be checked, and single-step calibration can be made.

Note that with reciprocity, there are only 10 independent terms left in the original 4×4 matrix describing the four-port on-wafer parasitics. On the other hand, the pad-open-short of [17] uses a 9-element equivalent circuit. It was then suggested and concluded in [17] with inductor data that pad-open-short is better than four-port, as it gives comparable results, but does not require using on-wafer load resistors. We reexamine this issue for active RF CMOS transistors in Fig. 4.7, where open-short, pad-open-short, and reciprocal four-port results are compared. The Y-parameters of PAD is estimated from layout using extraction tools, as was done in [17]. Above 50 GHz, open-short is much less accurate, as the lumped equivalent circuit with 6 elements fails. Although pad-open-short includes 9 elements in the lumped equivalent circuit, the improvement over open-short is very limited. The reciprocal four-port with 10 error terms does a much better job particularly above 50 GHz. The main reason for the success of the 10 term reciprocal four-port method, we believe, is that it does not rely on lumped equivalent circuit, and has little to do with the use of one more term than pad-open-short. One may use an equivalent circuit with more than 10 elements and still obtain less accurate results, as lumped circuits fail at higher frequencies. Our results strongly suggest that for higher frequency transistor measurements, four-port is necessary and superior to pad-open-short, despite the need for on-wafer load resistors.

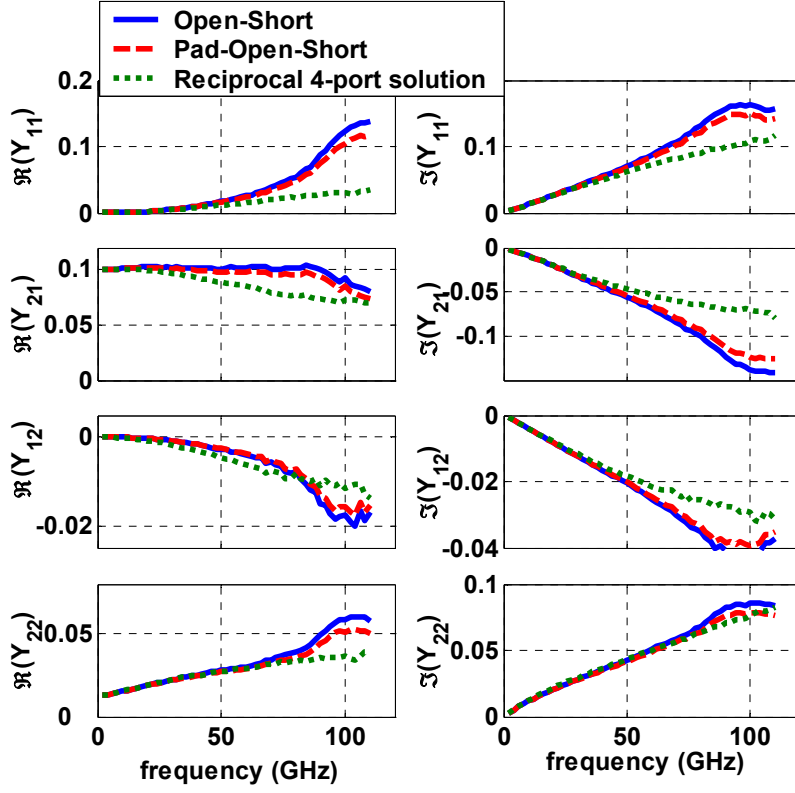


Fig. 4.7. Reciprocal four-port de-embedded transistor Y-parameters versus the results using open-short and pad-open-short de-embedding.

4.4 Summary

A new general four-port solution for on-wafer transistor measurements is developed and its utility is demonstrated on a $0.13\mu\text{m}$ RF CMOS process. The impact of non-ideal on-wafer load resistor is examined, and can be accounted for by including the parallel parasitic capacitances. Through proper normalization, easy to use new criteria are developed for quantifying the difference between open-short and four-port, as well as for examining reciprocity and symmetry of the four-port parasitics. Despite the assumption of ideal OPEN and SHORT, as was done in all de-embedding methods for achieving analytical solution, the solved four-port network for on-wafer parasitics is

shown to be reciprocal. Comparison with pad-open-short shows that for transistor measurements, pad-open-short does not provide significant improvement over open-short, and four-port is necessary despite the need to use on-wafer load resistors.

CHAPTER 5

NUMERICAL FOUR-PORT SOLUTION

On-wafer transistor S-parameter measurement is fundamentally important in both laboratory and production testing. The most complete system error model is the 16-term model [36], which accounts for all of the possible signal paths between the four waves measured inside the VNA and the four waves at the two terminals of the DUT, as illustrated in Section 3.5. The idea of describing everything between the probe tips and the device terminals as a four-port network [17] [20], is essentially the same as the 16 term error adaptor concept in system error calibration, at least mathematically, as illustrated in Fig. 5.1. Analytical equations for determining the 4×4 Y-parameters of the four-port network are developed in Chapter 4 and [19] [20], using five on-wafer standards. However, these analytical solutions can only be applied if the specified five standards, OPEN, SHORT, LEFT, RIGHT, and THRU, are available. Once other standards are used, new equations need to be derived. Also, due to its analytical nature, the solutions cannot take advantage of the redundancy available from the measurements of five on-wafer standards, and does not provide information on the relevant importance of the 16 terms of the parasitic four-port. Furthermore, analytical solutions do not provide information on systematic measurement errors. These issues are ideally handled with a singular value decomposition (SVD) based solution which solves the 4×4 S- and T-parameters of the parasitics four-port. Experimental results are demonstrated on a

0.13 μm RF CMOS process. Note that SVD was first used to solve for the T-parameters of the 16 term error model in [36].

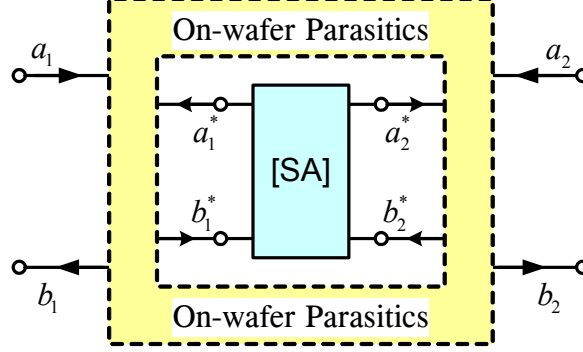


Fig. 5.1 The four-port error adaptor for on-wafer parasitics in wave representation.

5.1 Four-port parasitic network in T-parameters

Fig. 5.1 illustrates the four-port error adaptor for on-wafer parasitics. a and b are the incident and reflected waves at each port. Port 1 and Port 2 are the two probe tips, while Port 1* and Port 2* are the two device terminals. The linear equation relationship in (3.27) can also be applied for on-wafer parasitics. Rewrite the equation as below

$$T_1 S^A - S^{DUT} T_2 S^A + T_3 - S^{DUT} T_4 = [0]_{2 \times 2}. \quad (5.1)$$

$$T = \begin{bmatrix} T_1 & T_3 \\ T_2 & T_4 \end{bmatrix} = \begin{bmatrix} t_1 & t_5 & t_9 & t_{13} \\ t_2 & t_6 & t_{10} & t_{14} \\ t_3 & t_7 & t_{11} & t_{15} \\ t_4 & t_8 & t_{12} & t_{16} \end{bmatrix}. \quad (5.2)$$

S^A is the S-parameter of the internal transistor itself, and S^{DUT} is the S-parameter of the transistor plus the probing pads and interconnects as defined in Section 3.6. Each

measurement on a known two-port standard gives a pair of known S^A and S^{DUT} and four linear equations in terms of T as expanding (5.1) gives

$$\begin{bmatrix} S_{11}^A & 0 & -S_{11}^A S_{11}^{DUT} & -S_{11}^A S_{12}^{DUT} & S_{21}^A & 0 & -S_{21}^A S_{11}^{DUT} & -S_{21}^A S_{12}^{DUT} & 1 & 0 & -S_{11}^{DUT} & -S_{12}^{DUT} & 0 & 0 & 0 & 0 \\ 0 & S_{11}^A & -S_{11}^A S_{21}^{DUT} & -S_{11}^A S_{22}^{DUT} & 0 & S_{21}^A & -S_{21}^A S_{21}^{DUT} & -S_{21}^A S_{22}^{DUT} & 0 & 1 & -S_{21}^{DUT} & -S_{22}^{DUT} & 0 & 0 & 0 & 0 \\ S_{12}^A & 0 & -S_{12}^A S_{11}^{DUT} & -S_{12}^A S_{12}^{DUT} & S_{22}^A & 0 & -S_{22}^A S_{11}^{DUT} & -S_{22}^A S_{12}^{DUT} & 0 & 0 & 0 & 0 & 1 & 0 & -S_{11}^{DUT} & -S_{12}^{DUT} \\ 0 & S_{12}^A & -S_{12}^A S_{21}^{DUT} & -S_{12}^A S_{22}^{DUT} & 0 & S_{22}^A & -S_{22}^A S_{21}^{DUT} & -S_{22}^A S_{22}^{DUT} & 0 & 0 & 0 & 0 & 0 & 1 & -S_{21}^{DUT} & -S_{22}^{DUT} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ \vdots \\ t_{13} \\ t_{14} \\ t_{15} \\ t_{16} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (5.3)$$

S_{mn}^A and S_{mn}^{DUT} are the elements of S^A and S^{DUT} , $m, n=1,2$. t_k is the elements of T , $k=1-16$. Written in matrix, the above four linear equations are $C_{4 \times 16} T_{16 \times 1} = [0]_{4 \times 1}$. $C_{4 \times 16}$ is the coefficient matrix for each two-port measurement. For two two-port standards, 2×4 equations will be obtained, and the coefficient matrix will be $C_{8 \times 16}$. For n two-port standards, $n \times 4$ equations will be obtained, and the coefficient matrix will be $C_{(n \times 4) \times 16}$. In principle, four two-port standards are sufficient to solve the 16 unknowns. However, in practice, five on-wafer standards are required, and only 15 unknowns can be fully determined.

5.2 SVD based four-port Solution

Given the four linear equations in (5.3), four on-wafer measurements give 16 linear equations which can be rewritten in matrix as $C_{16 \times 16} T_{16 \times 1} = [0]_{16 \times 1}$. It seems like that the 16 unknowns can all be fully determined using the 16 equations. However, the only possible 16-term solution is an all zero solution because the set of equations is homogenous. In linear algebra, the rank of the coefficient matrix determines the number

of the unknowns can be solved. If $C_{16 \times 16}$ is full rank, the only possible solution is $T_{16 \times 1} = [0]_{16 \times 1}$. Since it is impossible that the error terms are all zero, the rank of the coefficient matrix is less than 16, that is to say that the maximum number of unknowns can be determined is 15, no matter how many on-wafer standards are measured.

The equations may be solved by normalizing the unknowns to one of the unknown terms, preferable one whose magnitude is close to unity. In Appendix H, the t_4 term was used as normalization factor for one-port error correction, essentially because the frequency response $e_{10}e_{01}$ can only be solved as product. Since $T_4 = E_2^{-1}$ and T_1, T_3, T_4 are functions of E_2^{-1} in (3.30), and the diagonal elements of E_2, e_{10} and e_{23} , are related to frequency response, the diagonal elements of T_4, t_{11} and t_{16} are good choices for normalization. t_{16} is used as the normalization factor in this dissertation. The normalized T matrix is

$$T' = \begin{bmatrix} T'_1 & T'_3 \\ T'_2 & T'_4 \end{bmatrix} = \begin{bmatrix} t'_1 & t'_5 & t'_9 & t'_{13} \\ t'_2 & t'_6 & t'_{10} & t'_{14} \\ t'_3 & t'_7 & t'_{11} & t'_{15} \\ t'_4 & t'_8 & t'_{12} & 1 \end{bmatrix}. \quad (5.4)$$

$t'_k = t_k / t_{16}, k=1-15$. After normalization, (5.1) can be rewritten as

$$T'_1 S^A - S^{DUT} T'_2 S^A + T'_3 - S^{DUT} T'_4 = [0]_{2 \times 2}. \quad (5.5)$$

The four linear equations for each measurement is then rewritten as

$$\begin{bmatrix}
S_{11}^A & 0 & -S_{11}^A S_{12}^{DUT} & -S_{11}^A S_{12}^{DUT} & S_{21}^A & 0 & -S_{21}^A S_{11}^{DUT} & -S_{21}^A S_{12}^{DUT} & 1 & 0 & -S_{11}^{DUT} & -S_{12}^{DUT} & 0 & 0 & 0 \\
0 & S_{11}^A & -S_{11}^A S_{21}^{DUT} & -S_{11}^A S_{22}^{DUT} & 0 & S_{21}^A & -S_{21}^A S_{21}^{DUT} & -S_{21}^A S_{22}^{DUT} & 0 & 1 & -S_{21}^{DUT} & -S_{22}^{DUT} & 0 & 0 & 0 \\
S_{12}^A & 0 & -S_{12}^A S_{11}^{DUT} & -S_{12}^A S_{12}^{DUT} & S_{22}^A & 0 & -S_{22}^A S_{11}^{DUT} & -S_{22}^A S_{12}^{DUT} & 0 & 0 & 0 & 0 & 1 & 0 & -S_{11}^{DUT} \\
0 & S_{12}^A & -S_{12}^A S_{21}^{DUT} & -S_{12}^A S_{22}^{DUT} & 0 & S_{22}^A & -S_{22}^A S_{21}^{DUT} & -S_{22}^A S_{22}^{DUT} & 0 & 0 & 0 & 0 & 0 & 1 & -S_{21}^{DUT}
\end{bmatrix}
\begin{bmatrix}
t_1' \\
t_2' \\
t_3' \\
t_4' \\
\vdots \\
t_{13}' \\
t_{14}' \\
t_{15}'
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
S_{12}^{DUT} \\
S_{22}^{DUT}
\end{bmatrix}. \quad (5.6)$$

Denoting the 15 normalized unknowns as $T'_{15 \times 1}$, $C_{16 \times 16} T_{16 \times 1} = [0]_{16 \times 1}$ can be rewritten as

$$A_{16 \times 15} T'_{15 \times 1} = B_{15 \times 1}, \text{ where } C_{16 \times 16} = \begin{bmatrix} A_{16 \times 15} & -B_{15 \times 1} \end{bmatrix}.$$

As discussed in Section 3.4, the e_{10} and e_{23} error terms cannot be measured independently because of the ratio nature of S-parameters [35] [36] [41] [42]. Since T-parameters represent the same four-port network as the E matrix, the same singularities exist in T and E matrices, although they are not that obvious in T matrix. By numerical simulation it was shown in [39] that the equations are singular for any four standards. The condition numbers of several sets of four standards are shown in Appendix G. There must be additional assumptions of the four-port network if 15 unknowns are solved using four standards [39] [41]. Five on-wafer standards are strictly needed for a general four-port solution. One of the standards should be a two-port standard or a through connection, e.g. THRU in Fig. 2.6. There is no upper limit for the number of standards. However, if the five standards chosen are nonsingular, adding more standards will not greatly improve the de-embedded results as shown in Appendix G.

Given pairs of S^A and S^{DUT} of five known standards, the elements of T can be determined from twenty (5×4) linear equations using (5.6). The set of linear equations can be written as $A_{20 \times 15} T'_{15 \times 1} = B_{20 \times 1}$. The set of equations $A_{20 \times 15} T'_{15 \times 1} = B_{20 \times 1}$ is over

determined, and can be solved using SVD. The solution is $T'_{15 \times 1} = A_{20 \times 15}^\dagger B_{20 \times 1}$, where $A_{20 \times 15}^\dagger$ is the pseudo-inverse of $A_{20 \times 15}$. The 15-term solution $T'_{15 \times 1} = A_{20 \times 15}^\dagger B_{20 \times 1}$ is sufficient for calculating actual S^A from measured S^{DUT} for any unknown DUT using an alternative expression of (5.1) as

$$S^A = (T'_1 - S^{DUT} T'_2)^{-1} (S^{DUT} T'_4 - T'_3). \quad (5.7)$$

To make the comparison between analytical four-port solution and numerical four-port solution easier, the same NMOS transistor and OPEN, SHORT, LEFT, RIGHT and THRU standards used for the analytical four-port solution in Section 4.2 are used here as a demonstration. Back-end-of-line resistors are used as on-wafer load in LEFT and RIGHT. Imperfect on-wafer load resistors are still modeled as $Y_L = G_L + j\omega C_L$ and $Y_R = G_R + j\omega C_R$ as shown in Fig. 4.2 (d) and (e). The value of G_L , C_L , G_R , and C_R are determined from low-frequency open-short de-embedding. According to analytical four-port solution, neglecting the capacitance does not introduce significant errors in most parameters.

The above process can be directly applied to solve for the combined four-port network including both system errors and on-wafer parasitics. The S^{DUT} will be replaced by the measured raw S-parameters, S^M . The results will be show in Section 6.2.

5.3 Experimental results for on-wafer parasitics de-embedding

Fig. 5.2 shows typical de-embedding results on a 32 finger N-type MOS transistor at one typical bias, $V_{GS}=1.5V$, $V_{DS}=1.5V$. Each finger has a gate width of $5\mu m$ and a length of $0.13\mu m$. S-parameters are measured using a HP 8510XF system, from 2GHz

to 110GHz. ISS calibration using SOLT is first performed. The ISS calibrated S-parameters of the five on-wafer standards shown in Fig. 2.6 are then used to determine the four-port T matrix (15 independent terms). The transistor S-parameters S^A are obtained using (5.7) and converted to Y-parameters. The de-embedded results using the popular open-short [14], pad-open-short [17], both of which are based on lumped equivalent circuits, are compared with the analytical four-port solution in Section 4.2.

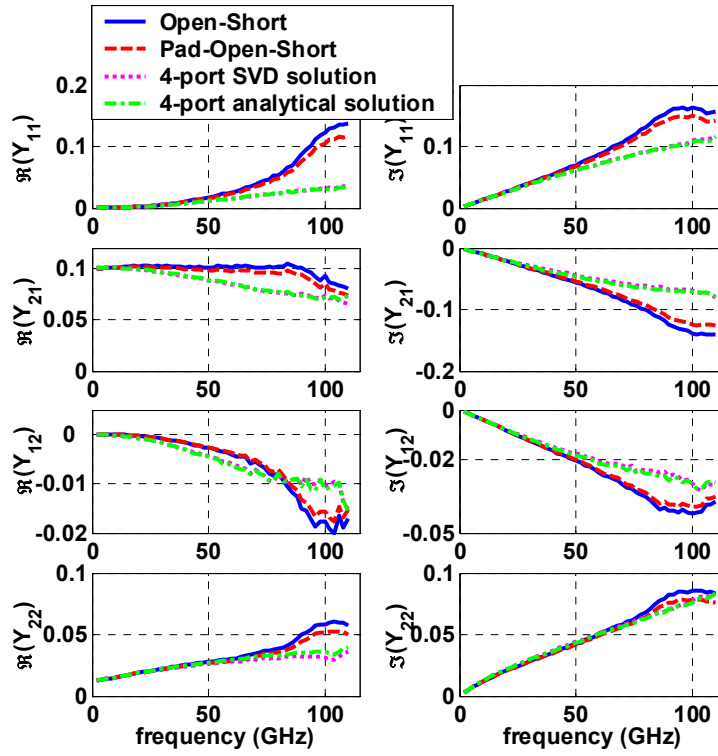


Fig. 5.2 Comparison of Y-parameters between open-short, pad-open-short, SVD based numerical four-port solution, and analytical four-port solution.

The SVD results are practically identical to the analytical four-port results in Section 4.2, which are carefully chosen among several possible solutions based on

singularity considerations. With SVD, singularity is naturally handled [36], and no special measurements need to be taken. Redundancy is handled by SVD as well [81].

The SVD and analytical four-port results are significantly different from the open-short and pad-open-short results. The frequency dependence of Y_{11} from four-port results is more physical. To observe this better, we plot out the effective gate resistance $R_{in} = \Re\{1/Y_{11}\}$ and effective gate capacitance $C_{in} = -1/\left[2\pi f \Im\{1/Y_{11}\}\right]$ [27] in Fig. 5.3. While open-short and pad-open-short give the same R_{in} as four-port solutions, they give a very strong and unphysical frequency dependence of the effective gate capacitance. In strong inversion, for an oxide thickness of only a few nanometers, the value of the effective gate capacitance is expected to be approximately constant even at 100GHz for a MOSFET of such short channel length.

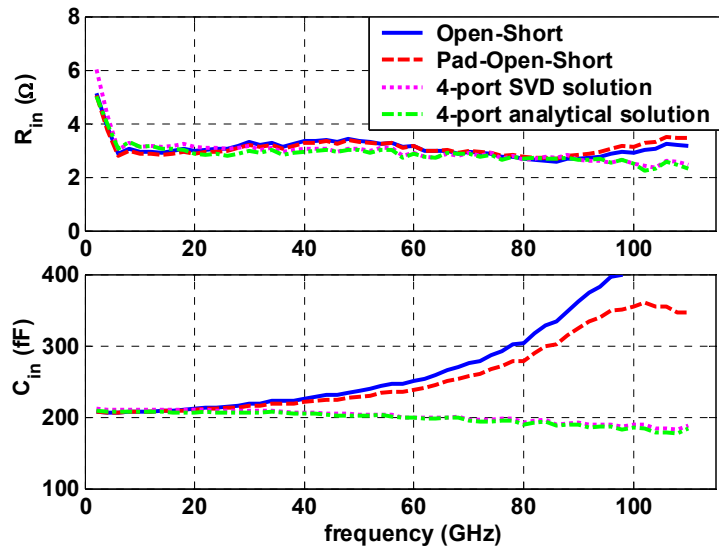


Fig. 5.3 Comparison of effective gate resistance and capacitance between open-short, pad-open-short, SVD based four-port solution, and analytical four-port solution.

5.4 Reduction of Error Terms and Number of Standards

Using SVD, the relevant importance of different error terms can be examined efficiently. For both the parasitics four-port and the combined four-port used in single-step calibration, only 12 terms of the full 16 terms are important as shown below. This reduces the number of standards from 5 to 4. SHORT, LEFT, RIGHT, and THRU are a good combination needed for the final T solution. One may then omit the OPEN structure. The saving in chip area is not significant, and it does not allow the use of open-short de-embedding at lower frequencies for extraction of capacitive parasitics of the left and right loads. Having open-short can be useful as this serves as the reference from traditional on-wafer de-embedding and open-short is known to be accurate at lower frequencies when used with ISS calibration. Comparison with open-short can prove useful at the algorithm development stage as consistency of four-port with open-short at low frequencies indicates a correct four-port solution.

5.4.1 Quantify error terms for four-port on-wafer parasitics

To quantify the impact of the 16 error terms, the SVD solved T -parameters are transferred to S -parameters since S -parameters give straightforward physical meanings of the signal paths. Because the solved T terms are normalized, the relationship between the normalized T terms and the S -parameters of the four-port network need to be developed first.

Using the relationship between the E and T elements in (3.30), the E elements calculated from the normalized T elements are

$$\begin{aligned}
E'_1 &= T'_3 (T'_4)^{-1} = E_1 \\
E'_2 &= (T'_4)^{-1} = t_{16} E_2 \\
E'_3 &= T'_1 - T'_3 (T'_4)^{-1} T'_2 = \frac{1}{t_{16}} E_3 \\
E'_4 &= -(T'_4)^{-1} T'_2 = E_4
\end{aligned} \tag{5.8}$$

E'_1, E'_2, E'_3, E'_4 represent the four-port network after normalization. The S-parameters of the four-port network without normalization is

$$E = \begin{bmatrix} E'_1 & t_{16} E'_3 \\ \frac{1}{t_{16}} E'_2 & E'_4 \end{bmatrix} = \left[\begin{array}{cc|cc} e_{00} & e_{03} & e_{01} & e_{02} \\ e_{30} & e_{33} & e_{31} & e_{32} \\ \hline e_{10} & e_{13} & e_{11} & e_{12} \\ e_{20} & e_{23} & e_{21} & e_{22} \end{array} \right]. \tag{5.9}$$

Although, E'_1, E'_2, E'_3 , and E'_4 are no longer the original S-parameters of the four-port network, it does not affect the relative importance of the error terms. Fig. 5.4 shows the magnitude of the error terms in E'_1, E'_2, E'_3 , and E'_4 . The normalization factor, t_{16} , does not affect the comparison between the diagonal elements and the non-diagonal elements in each 2×2 matrix. In Section 3.3, the 8-term model assumes that the leakage terms, $(e_{30}, e_{03}), (e_{31}, e_{02}), (e_{20}, e_{13})$, and (e_{21}, e_{12}) , are negligible. From Fig. 5.4, it is clear that this assumption works well for the whole frequency range as the magnitude of the non-diagonal elements of each 2×2 matrix is at least 20dB lower than the magnitude of the diagonal elements. Note that although the magnitude of the diagonal elements of E_2 and E_3 are much larger than the diagonal elements of E_1 and E_4 , it cannot be concluded that the diagonal elements of E_2 and E_3 are dominant elements, because the elements in E_2 and E_3 are normalized S-parameters.

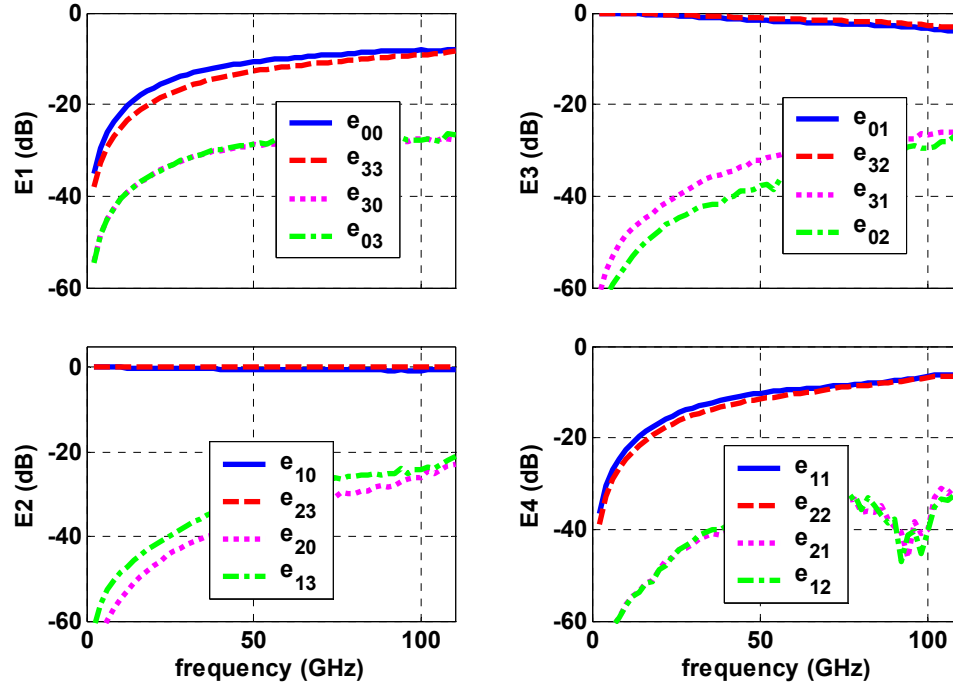


Fig. 5.4 The magnitude of the S-parameters for the four-port on-wafer parasitics.

5.4.2 8-term solution using three on-wafer standards

Since the non-diagonal elements of E'_1 , E'_2 , E'_3 , and E'_4 are much less than the diagonal elements, it is possible that 8 error terms is sufficient for on-wafer de-embedding. Note that here the 8 error terms are e_{00} , e_{11} , e_{10} , e_{01} , e_{22} , e_{33} , e_{23} , and e_{32} as shown in Section 3.3. Because SVD solves the T-parameters of the four-port on-wafer parasitics, the 8-term E matrix is transformed to T matrix using

$$\begin{aligned}
T_1' &= \frac{1}{t_{16}} (E_3 - E_1 E_2^{-1} E_4) \\
T_2' &= -\frac{1}{t_{16}} E_2^{-1} E_4 \\
T_3' &= \frac{1}{t_{16}} E_1 E_2^{-1} \\
T_4' &= \frac{1}{t_{16}} E_2^{-1}
\end{aligned} \tag{5.10}$$

If only 8 error terms, the diagonal elements of E_1 , E_2 , E_3 , and E_4 , are involved, the non-diagonal elements of the corresponding T_1' , T_2' , T_3' , and T_4' matrices calculated using (5.10) are all zero.

$$E = \left[\begin{array}{cc|cc} e_{00} & 0 & e_{01} & 0 \\ 0 & e_{33} & 0 & e_{32} \\ \hline e_{10} & 0 & e_{11} & 0 \\ 0 & e_{23} & 0 & e_{22} \end{array} \right] \Rightarrow T' = \left[\begin{array}{cc|cc} t_1' & 0 & t_9' & 0 \\ 0 & t_6' & 0 & t_{14}' \\ \hline t_3' & 0 & t_{11}' & 0 \\ 0 & t_8' & 0 & 1 \end{array} \right] \tag{5.11}$$

The four linear equations for each measurement are

$$\left[\begin{array}{cccccc} S_{11}^A & -S_{11}^A S_{11}^{DUT} & 0 & -S_{21}^A S_{12}^{DUT} & 1 & -S_{11}^{DUT} & 0 \\ 0 & -S_{11}^A S_{21}^{DUT} & S_{21}^A & -S_{21}^A S_{22}^{DUT} & 0 & -S_{21}^{DUT} & 0 \\ S_{12}^A & -S_{12}^A S_{11}^{DUT} & 0 & -S_{22}^A S_{12}^{DUT} & 0 & 0 & 0 \\ 0 & -S_{12}^A S_{21}^{DUT} & S_{22}^A & -S_{22}^A S_{22}^{DUT} & 0 & 0 & 1 \end{array} \right] \left[\begin{array}{c} t_1' \\ t_3' \\ t_6' \\ t_8' \\ t_9' \\ t_{11}' \\ t_{14}' \end{array} \right] = \left[\begin{array}{c} 0 \\ 0 \\ S_{12}^{DUT} \\ S_{22}^{DUT} \end{array} \right]. \tag{5.12}$$

Three standards give 12 equations, which are sufficient to solve the 7 unknown elements in (5.12). However, the set of non-singular standards need to be carefully chosen. Some of on-wafer standards may lead to unphysical results. For example, with a

perfect matched load at Port 1, the second equation in (5.12) is $S_{21}^{DUT} t_{11} = 0$, which is obviously not true in practice. The singularity of the standards can be verified using condition number of the coefficient matrix. Considering the 5 available standards fabricated, there are six possible combinations of three standards if THRU is chosen as one standard. Among the six combinations, only three of them are non-singular, i.e. 1) SHORT, THRU, LEFT, 2) SHORT, THRU, RIGHT, and 3) THRU, LEFT, RIGHT. Fig. 5.5 shows the condition number of the coefficient matrix built using the six combinations of three standards. The condition number for 5 standards is also shown as reference. Fig. 5.6 compares the de-embedded Y-parameters using open-short, SVD based 16-term and 8-term solution. The 5 standards used to solve the 16-term solution are OPEN, SHORT, THRU, LEFT, and RIGHT. The 3 standards used for 8-term solution are SHORT, THRU, and LEFT. With non-singular standards, 8-term model can provide reasonably accurate results.

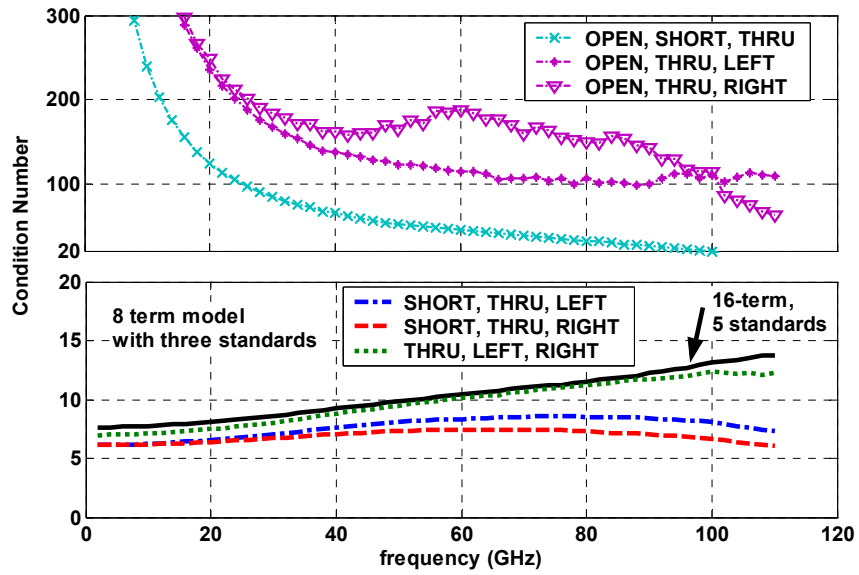


Fig. 5.5 Comparison of Y-parameters between open-short, SVD based 16-term solution, and SVD based 8-term solution.

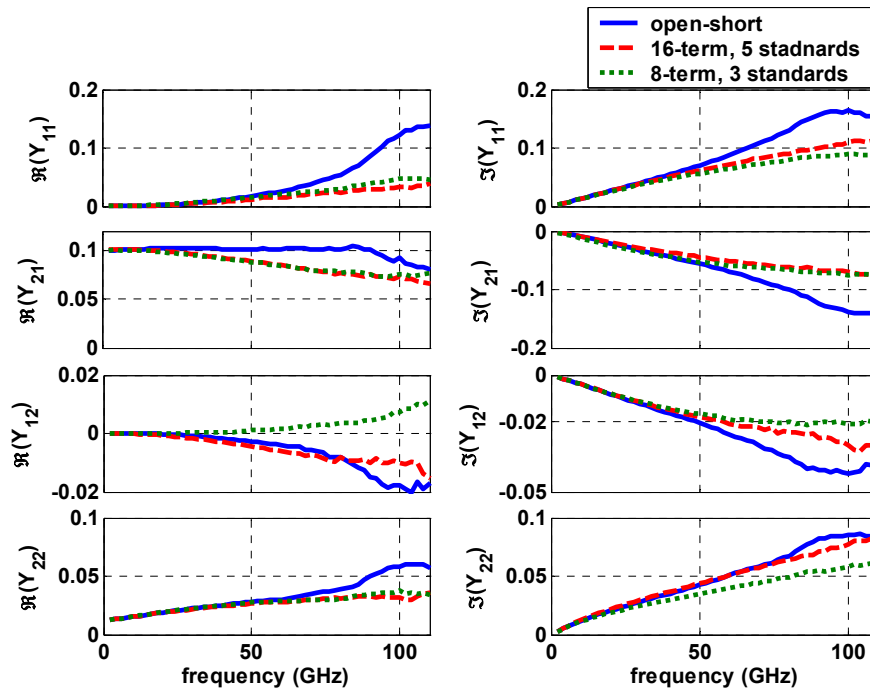


Fig. 5.6 Comparison of Y-parameters between open-short, SVD based 16-term solution, and SVD based 8-term solution.

5.5 Summary

On-wafer parasitics de-embedding using a SVD based numerical four-port solution is demonstrated on a $0.13\mu\text{m}$ RF CMOS process. The SVD four-port results are shown to be close to the analytical four-port results in Section 4.2. Redundancy and singularities are naturally handled by SVD. The leakage errors are much smaller than the directivity errors, frequency response errors, and port match errors. 8-term error model is sufficient for on-wafer parasitics de-embedding. Three standards are necessary for solving the 8-term error matrix. Non-singular standards must be carefully designed. The condition number of the coefficient matrix can be used as an indicator of the singularity of the sets of standards and thus the validity of the de-embedded results.

CHAPTER 6

SINGLE-STEP CALIBRATION

As discussed in Section 3.6, both system errors and on-wafer parasitics can be described as a four-port network. A significant difference between the four-port system error adaptor and the four-port on-wafer parasitics network is that system errors drift over time, and for this reason, ISS calibration must be performed at least once a day, and validated often, e.g. before measurement of each wafer lot. This inevitably requires an operator to load a special ISS holder on the wafer prober, making measurement time consuming and impossible to automate in production testing. A solution to this problem is to combine the four-port system error adaptor and the four-port on-wafer parasitics into one four-port network, and directly solve the combined four-port network without ISS calibration using the same on-wafer standards previously used for on-wafer four-port parasitics de-embedding. In practice, this single-step approach is rarely used, particularly for transistor measurements, for various reasons, the most important being that on-wafer standards are less accurately known compared to precision ISS calibration standards. In this work, we will compare the results from the two-step approach, i.e. ISS calibration plus four-port on-wafer parasitics de-embedding, and the results from the single-step approach to quantify the errors introduced in the much simpler and easier to automate single-step approach on a $0.13\mu\text{m}$ RF CMOS process. This is facilitated for two four-port de-embedding approaches, the analytical four-port solution in Section 4.2,

and the numerical four-port solution in Section 5.2. We will show that the single-step approach can give as accurate transistor Y-parameters as two-step calibration, from 2GHz to 110GHz. However, switch errors must be removed first, since switch errors are not involved in the four-port network. Switch errors are introduced by non-ideal Z_0 or non-ideal switch. The four-port error adaptor only lumps the linear errors between the measured four-waves, the four receivers, and the four desired waves.

6.1 Analytical four-port single-step calibration

Although the combined four-port network is no longer reciprocal, the intrinsic device parameters can still be retrieved from measured raw S-parameters using the general four-port solution in Section 4.2 as is, without performing ISS calibration. For best accuracy, the parasitic capacitance of the load resistor in LEFT and RIGHT can be included in the same way as in the two-step four-port calibration described in Section 4.2. The parasitic capacitance needs to be determined from ISS calibrated LEFT and RIGHT measurement. As the on-wafer parasitics do not drift a lot as the VNA system errors do, the capacitance only needs to be determined once. The value can then be used for all measurements sharing the same load resistor. It does not need to be frequently verified or recalibrated as VNA system error calibration does, which, in general, will cost at least 30 minutes for one full two-port calibration. Moreover, poor calibration associated with less accurate calibration standards can also degrade the overall accuracy of the measured results. Single-step calibration results, using the general four-port solutions, can save a lot of time and effort during RF on-wafer measurements.

To perform single-step calibration, the measured raw S-parameters without ISS calibration are directly used for all of the calculations from step 2) to step 8) in Section 4.2.4. As the on-wafer standards are all assumed to be ideal, e.g. ideal OPEN and SHORT, relatively ideal on-wafer load resistor with a capacitive parasitics, the single-step calibration results are expected to be less accurate than the two-step calibration results. Fig. 6.1 compares single-step and two-step four-port calibration results for the HP 8510XF system, from 2 GHz to 110 GHz. As expected, the Y-parameters from single-step calibration are not as well behaved as the Y-parameters from two-step calibration. However, the overall values of Y-parameters are still fairly accurate, particularly for critical parameters like imaginary part of Y_{11} , which indicates the gate capacitance. To further analysis the source of the small ripples, the same measurements are repeated using another system, a HP 8510C system, from 2 GHz to 26 GHz. The results are shown in Fig. 6.2. The 8510C results are much less noisy than the 8510XF result, even when compared over the same frequency range. Given the measurement system dependence and the frequency dependence, these ripples in single-step calibration are believed to be due to the system errors that are not calibrated out by the on-wafer standards. Since the SVD based numerical four-port solution can give information of the singularity of the solution, examining the condition number during single-step calibration using the SVD based solution in Section 6.2 may give some information of the ripples.

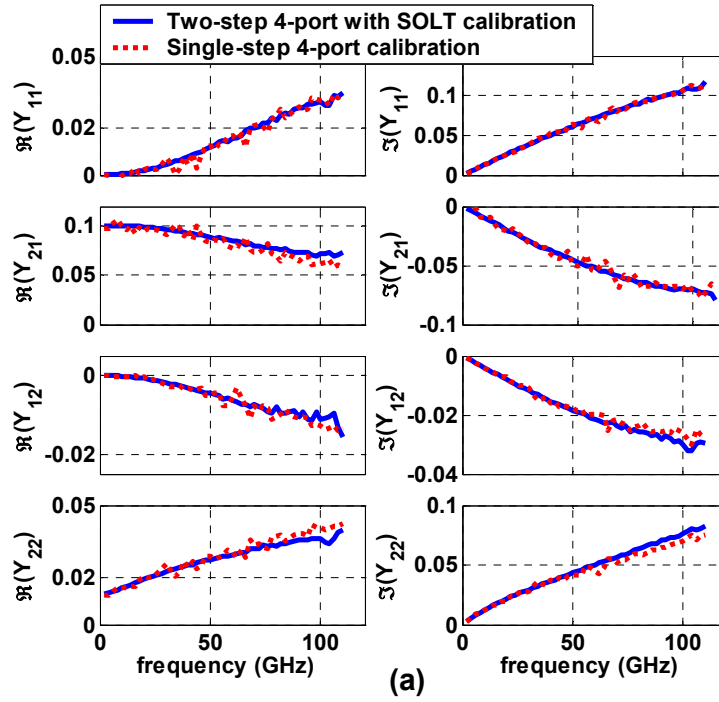


Fig. 6.1. Single-step versus two-step four-port using the analytical four-port solution with data measured using a HP 8510XF system from 2 GHz to 110 GHz.

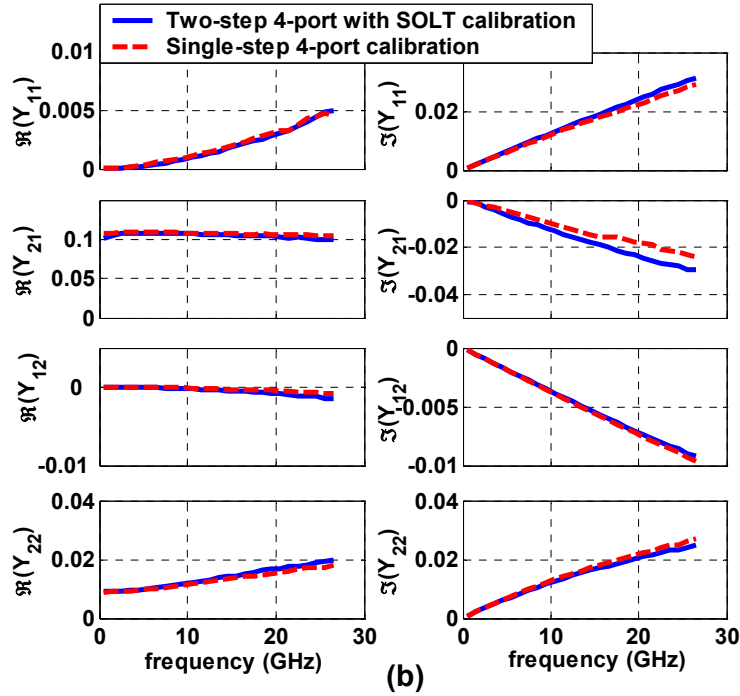


Fig. 6.2. Single-step versus two-step four-port using the analytical four-port solution with data measured using a HP 8510C system from 2 GHz to 26.5 GHz.

6.2 Numerical four-port single-step calibration

The same four-port SVD algorithm in Section 5.2 is applied on the raw S-parameters without ISS calibration to remove the combined four-port network including both system errors and on-wafer parasitics. S^{DUT} in (5.1) and (5.7) is replaced by S^M . Again, the on-wafer standards are all assumed to be ideal, e.g. ideal open and short, relatively ideal resistor loads with a capacitive parasitics.

Fig. 6.3 compares the single-step SVD four-port results with the two-step SVD four-port results. Also shown are the open-short results with ISS calibration, the most popular practice today. The pad-open-short results are similar to open-short results, and thus not repeated here. Similar to the single-step calibration using analytical four-port solution, the Y-parameters from single-step SVD four-port are noisier than the Y-parameters from two-step SVD four-port. To summarize, single-step calibration using four-port techniques has led to reasonably accurate transistor Y-parameters, despite the less accurate on-wafer standards compared to precision ISS standards. The ability to avoid ISS calibration makes this attractive for production testing, as ISS calibration needs to be performed and checked often, and involves a separate manual step of loading ISS holder.

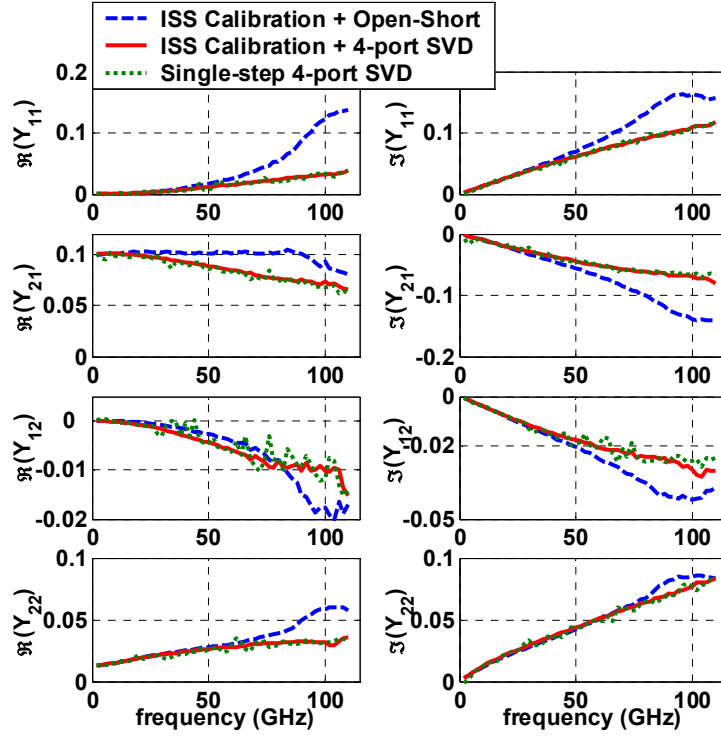


Fig. 6.3 Comparison between two-step open-short and four-port on-wafer parasitics de-embedding results with ISS calibration and single-step four-port calibration results without any ISS calibration.

Another advantage of using single-step calibration is that the distributive nature of on-wafer parasitics is naturally included, as evidenced by the closeness of the single-step results to the two-step results. Above 50GHz, the open-short results are much less accurate, simply because the lumped equivalent circuit used for open-short de-embedding fails. Even though system errors are removed more accurately with ISS standards when compared with single-step calibration, the failure of open-short on-wafer de-embedding makes the final result invalid.

The added advantage of using SVD is that it not only solves the system equations, but also gives valuable information about the system [36] [89]. The condition number of

the coefficient matrix is an indicator of the error sensitivity. For the same inaccurate on-wafer standards, the condition number is noticeably higher for single-step four-port calibration, as shown in Fig. 6.4, indicating less tolerance to measurement errors. This is another reason for the less accurate single-step result compared to the two-step result with ISS calibration.

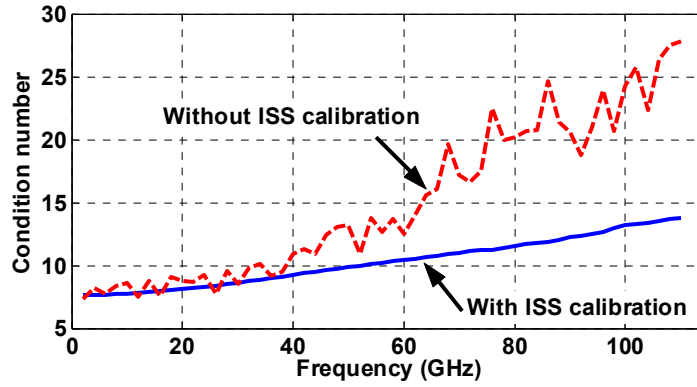


Fig. 6.4 Condition numbers of the coefficient matrix in on-wafer parasitics de-embedding and single-step calibration.

6.3 Impact of switch errors

One possible reason for the ripples is the switch errors. To investigate this, the switch errors are removed using the algorithm in Section 3.1 [80]. Fig. 6.5 and Fig. 6.6 compare the single-step calibrated results with and without switch error removal for the two four-port solutions. The two-step four-port calibrated results are also shown for comparison. Fig. 6.5 compares results using analytical four-port, while Fig. 6.6 compares results using SVD based numerical four-port. Both of them indicate that switch errors are the most important reason for the ripples in single-step calibration.

Switch error terms Γ_1 and Γ_2 are determined by the load impedance connected to the switch inside the VNA system, which does not change a lot even for months. Adding switch error removal will not cost a lot of labor for large volume measurements.

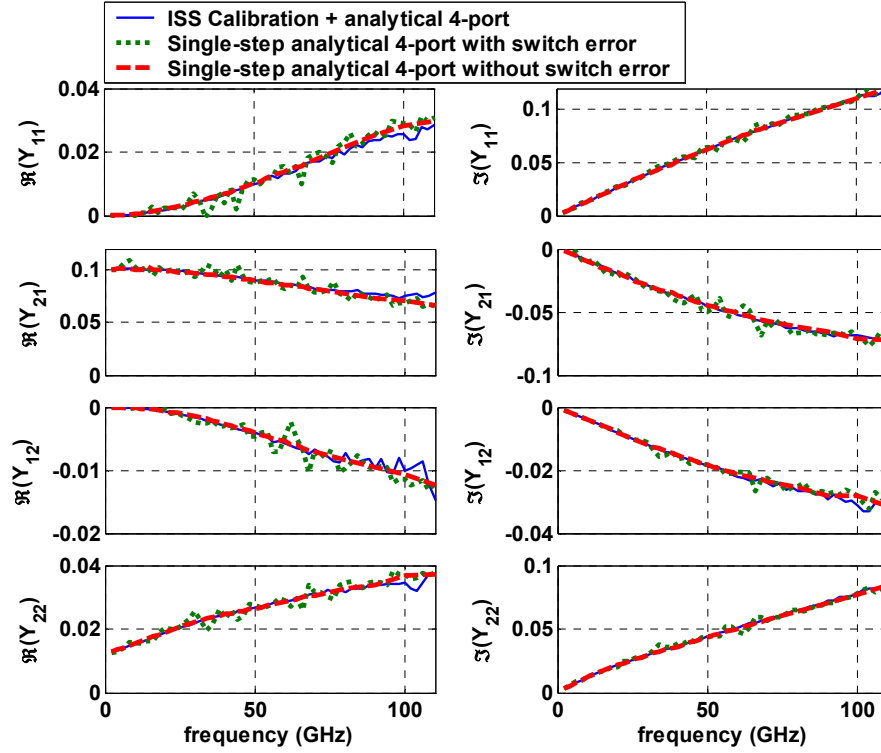


Fig. 6.5 Comparison of the single-step four-port calibrated results with and without switching error correction. The analytical four-port solution in Section 4.2 is applied.

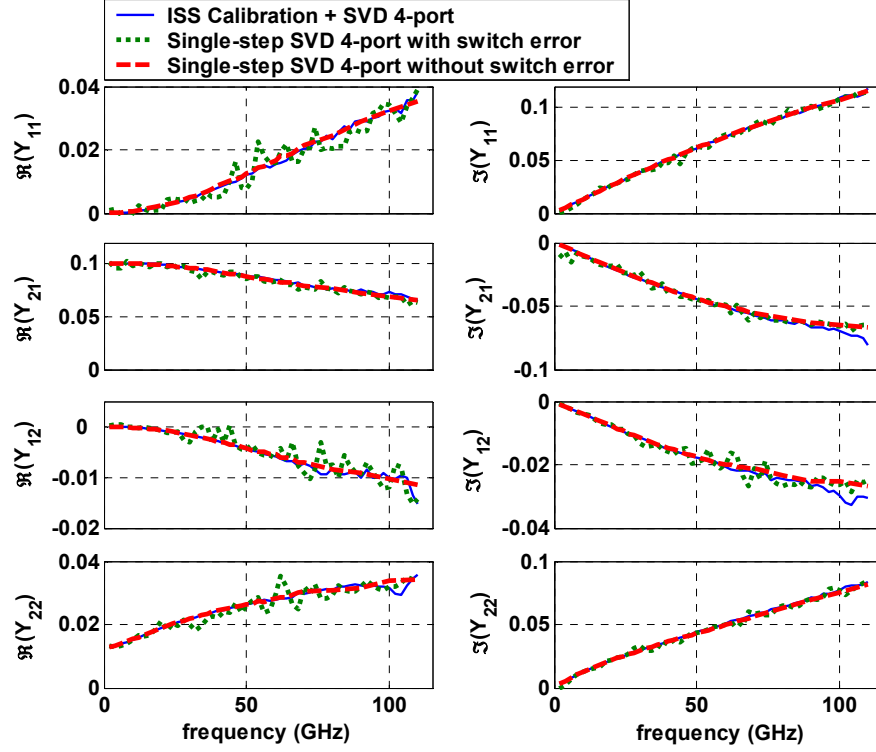


Fig. 6.6 Comparison of the single-step four-port calibrated results with and without switching error correction. The SVD based numerical four-port solution in Section 5.2 is applied.

6.3.1 Quantify error terms using S-parameters

As discussed in Section 5.4, although, E'_1 , E'_2 , E'_3 , and E'_4 are no longer the original S-parameters of the four-port network, it does not affect the relative importance of the error terms. Fig. 6.7 shows the magnitude of the error terms in E'_1 , E'_2 , E'_3 , and E'_4 . The normalization factor, t_{16} , does not affect the comparison between the diagonal elements and the non-diagonal elements in each 2×2 matrix. From Fig. 6.7, it is clear that this assumption limits the application of 8-term model on single-step calibration. First, (e_{30}, e_{03}) are not that small when compared with (e_{00}, e_{33}) even at low frequencies.

Similar situation exist when evaluating the elements in E_4 . Secondly, as frequency increases, the difference between the diagonal elements and the non-diagonal elements in E_2 and E_3 reduces, which means the leakage errors become comparable to the dominant errors. Note that although the magnitude of the diagonal elements of E_2 and E_3 are 10dB larger than the diagonal elements of E_1 and E_4 , it cannot be concluded that the diagonal elements of E_2 and E_3 are dominant elements, because the elements in E_2 and E_3 are normalized S-parameters.

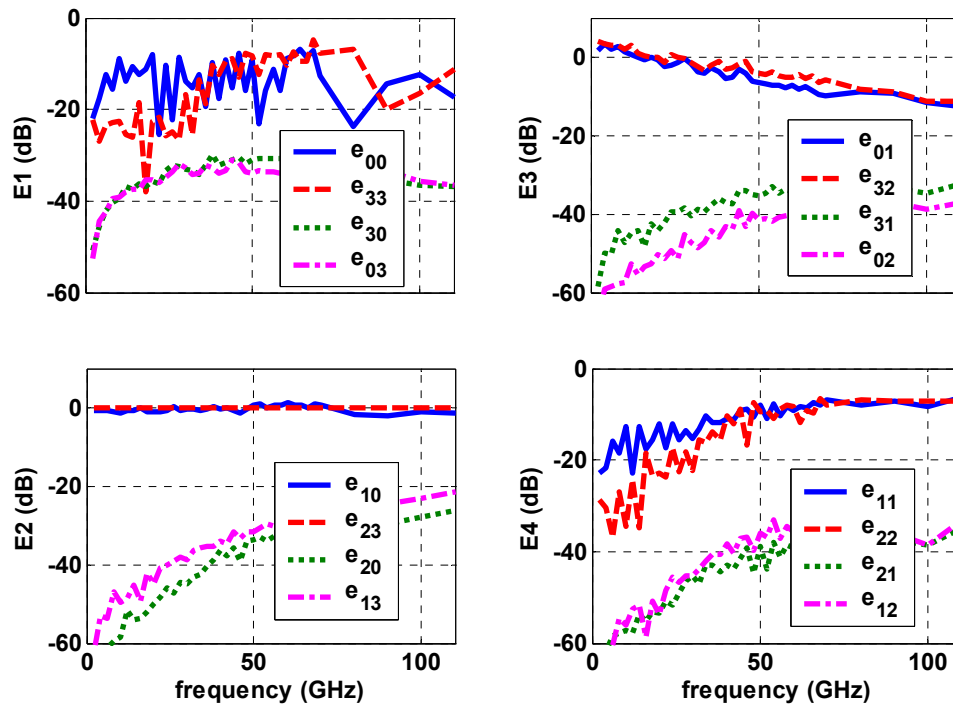


Fig. 6.7 The magnitude of the solved 16 error terms of the combined four-port network.

6.4 Summary

The accuracy of single-step calibration using two general four-port solutions is experimentally investigated on a $0.13\mu\text{m}$ RF CMOS process. In contrast to popular belief, single-step four-port calibration produces reasonably accurate and acceptable transistor Y-parameters from 2GHz to 110GHz, despite the less accurate on-wafer standards compared to precision ISS standards, which facilitates production testing and process monitoring. The distributive nature of on-wafer parasitics is also naturally included, due to the four-port description of the combined error adaptor. The single-step approach to transistor measurements is thus valuable as it does not require ISS calibration and thus facilitates production testing. The impact of switch errors on single-step measurement is also investigated. After removing switch error, single-step calibration provides practically the same results as two-step calibration for both the analytical four-port solution which was first developed for on-wafer parasitics and the numerical four-port solution based on SVD.

CHAPTER 7

VALIDITY OF BSIM4 MODEL FOR NONLINEAR RF MODELING

Once the model parameters are extracted from a set of DC, CV, and S-parameter measurements, it is important to verify the developed model by performing model validation experiments. The idea is to provide an environment as close as possible to the real measurement, and to verify whether the model can predict the measured results. Only after the model passes the validation test, the model can be transferred to designers. Note that the measurements used for model parameter extraction can be quite different from the measurements used for model validation. For example, distortion measurement with high input power is not used for parameter extraction, but it is necessary for verifying the linear model developed as distortion exists in real applications. In this chapter, the DC, AC, and nonlinear performance of a BSIM4 model is verified.

BSIM4 model is one of the widely used MOS transistor model for RF designs. In BSIM4, the moderate inversion region is modeled by mathematical smoothing functions interpolating between physics based approximations in the weak and strong inversion regions, instead of physics based surface potential approximation that can cover all levels of inversion. Its accuracy in linearity simulation, particularly in moderate inversion region, needs to be experimentally evaluated, since an IP3 sweet spot exists in this region.

7.1 Linearity measurement and simulation

The two-tone on-wafer system in Fig. 1.6 is used to measure the output spectrum at the drain of the examined NMOS transistor [22]. Fig. 7.1 shows the simplified block diagram for this 50Ω system. The gate and the drain of the NMOS transistor in linearity simulation are also terminated at 50Ω ports. Fig. 7.2 shows the schematic for two-tone intermodulation distortion simulation. A “psin” component from the “analoglib” of Cadence generates the two-tone excitation at the gate. Since linearity measurement is made at the probe tips, and it is impossible to calibrate on-wafer parasitics, a passive RLC network, which models the low frequency on-wafer parasitics, is inserted between bias tees and device terminals to make the environment as close as possible to real measurement. An array of transistors with number of fingers (N_f) ranging from 5 to 64 are measured and simulated with sweeping biasing voltages at different fundamental frequencies. Although QPSS analysis is selected to speed the linearity simulation, it may still take days even with a high performance computer.

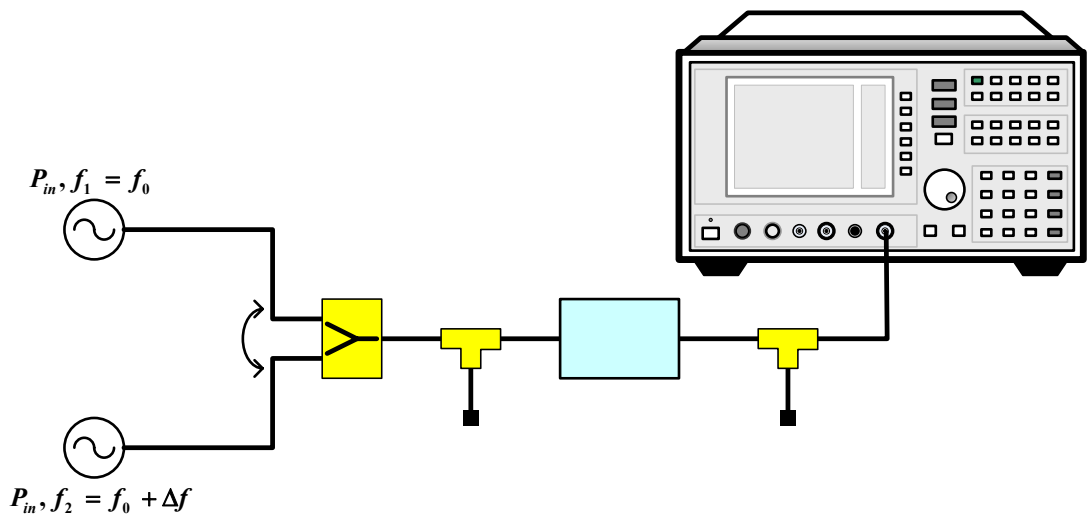


Fig. 7.1 Block diagram for two-tone intermodulation linearity measurement.

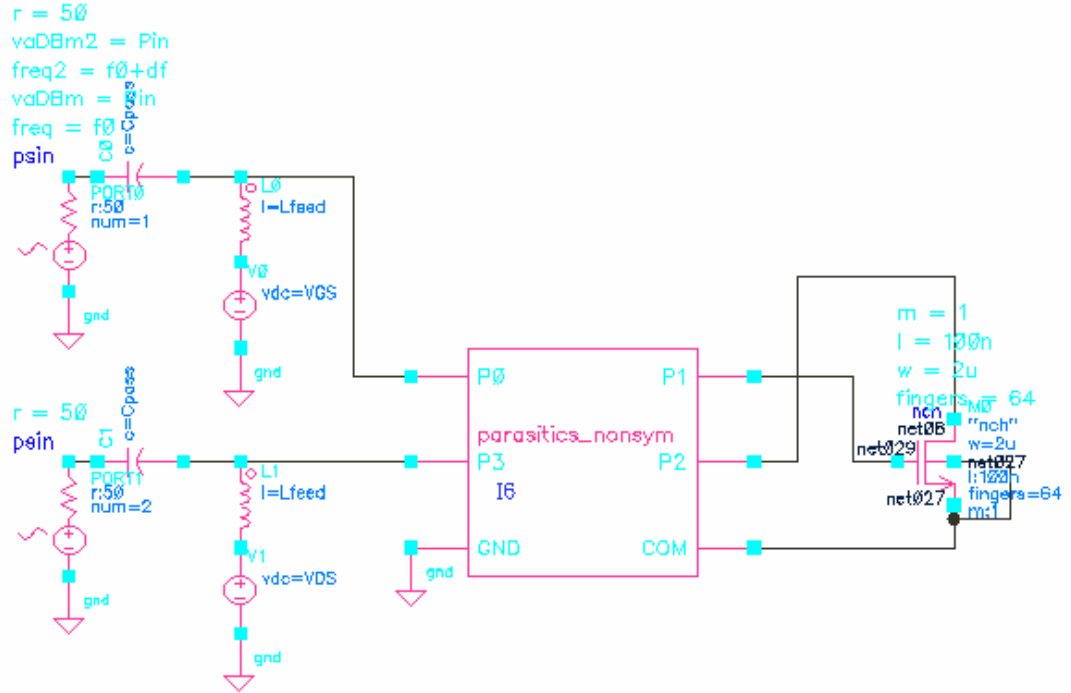


Fig. 7.2 Schematic for two-tone intermodulation linearity simulation in Cadence.

Each finger of the NMOS transistor is $2\mu\text{m}$ wide and 90nm long. For $N_f = 64$, the total width is $128\mu\text{m}$, which is close to those found in 5GHz low power 90nm CMOS LNAs [49]. Therefore, the analysis will be focused on this device size. In practice, because of low gain and low levels of intermodulation products in small width transistors as well as dynamic range limits of the spectrum analyzer, the minimum gate width that IP_3 can be measured reliably is $10 \times 2\mu\text{m}$ for this 90nm CMOS technology. IIP_3 , OIP_3 , and power gain are measured at different V_{GS} , V_{DS} , and fundamental frequencies to examine the biasing and frequency dependence of IP_3 sweet spot. For each measurement, the power amplitude at the signal generators, P_{in} , is swept, and the

output power level for the 1st order output and the 3rd order intermodulation product are monitored, $P_{out,1st}$ and $P_{out,3rd}$. After power calibration, the third order intercept point is obtained using linear extrapolation illustrated in Section 1.2. $P_{in,ref}=-25\text{dBm}$.

7.2 DC and linear characteristics

Fig. 7.3 shows $I_{DS}-V_{GS}$ curves for $V_{DS}=0.6, 0.8$ and 1.0V and Fig. 7.4 shows $I_{DS}-V_{DS}$ curves for $V_{GS}=0.4$ and 0.8V . Fig. 7.5 (a) shows S_{21} versus frequency at one fixed biasing point, $V_{GS}=0.4\text{V}$ and $V_{DS}=1.0\text{V}$, while Fig. 7.5 (b) shows S_{21} versus gate biasing voltage at a fixed frequency, 5GHz . Fig. 7.6 compares the cut-off frequency, f_T , extracted from S-parameters versus gate voltage. Fig. 7.7 shows all Y-parameters versus frequency at $V_{GS}=0.4\text{V}$ and $V_{DS}=1.0\text{V}$, a representative moderate inversion bias. Fig. 7.8 shows all Y-parameters at 5GHz versus V_{GS} for $V_{DS}=1.0\text{V}$. Both simulation and measurement data are shown in Fig. 7.3-Fig. 7.8. The Y-parameters here include pad parasitics by design, as IP3 is measured on-wafer including probing pads. Overall, the BSIM4 based subcircuit model does a good job in modeling $I_{DS}-V_{GS}$, $I_{DS}-V_{DS}$, S-parameters, f_T , and both frequency and bias dependence of most Y-parameters over the whole V_{GS} region, including the moderate inversion region.

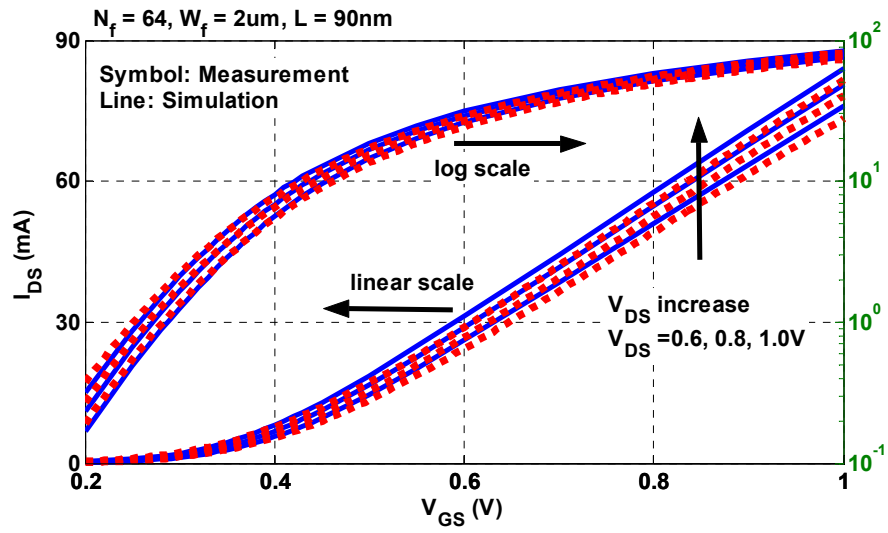


Fig. 7.3 Measured and simulated $I_{DS} - V_{GS}$ for $V_{DS} = 0.6, 0.8$, and 1.0V .

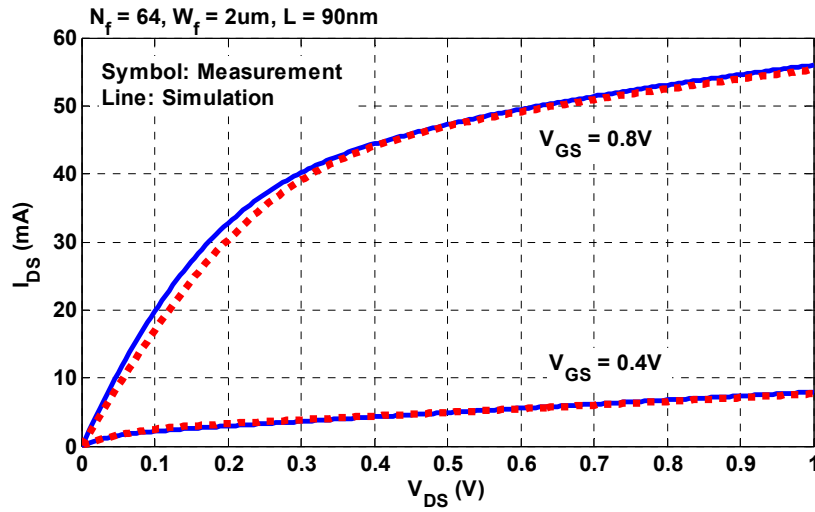


Fig. 7.4 Measured and simulated $I_{DS} - V_{DS}$ for $V_{GS} = 0.4\text{V}$ and 0.8V .

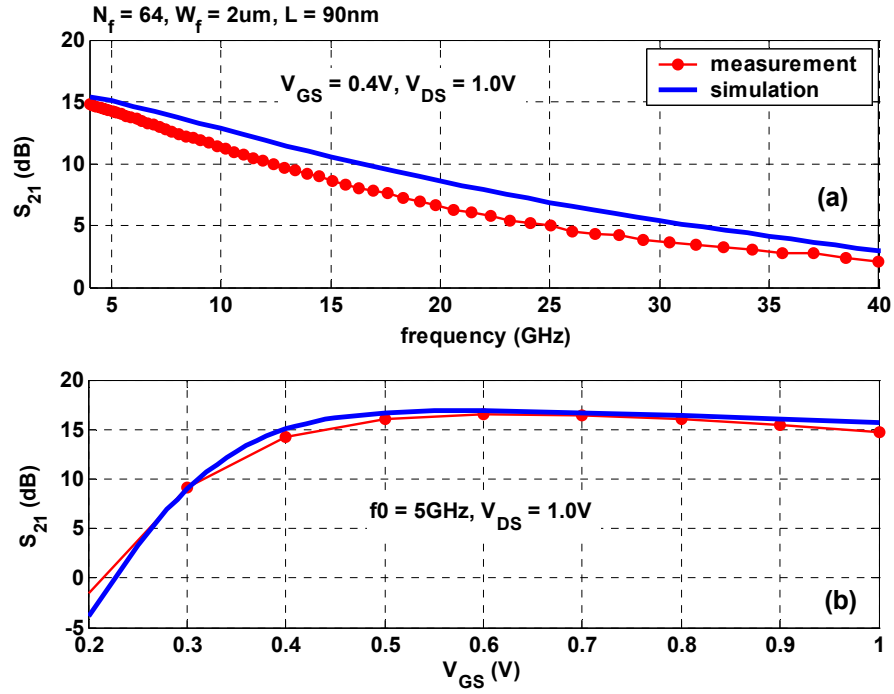


Fig. 7.5 (a) S_{21} in dB versus frequency at $V_{GS} = 0.4\text{V}$ and $V_{DS} = 1.0\text{V}$. (b) S_{21} in dB versus V_{GS} at 5GHz and $V_{DS} = 1.0\text{V}$.

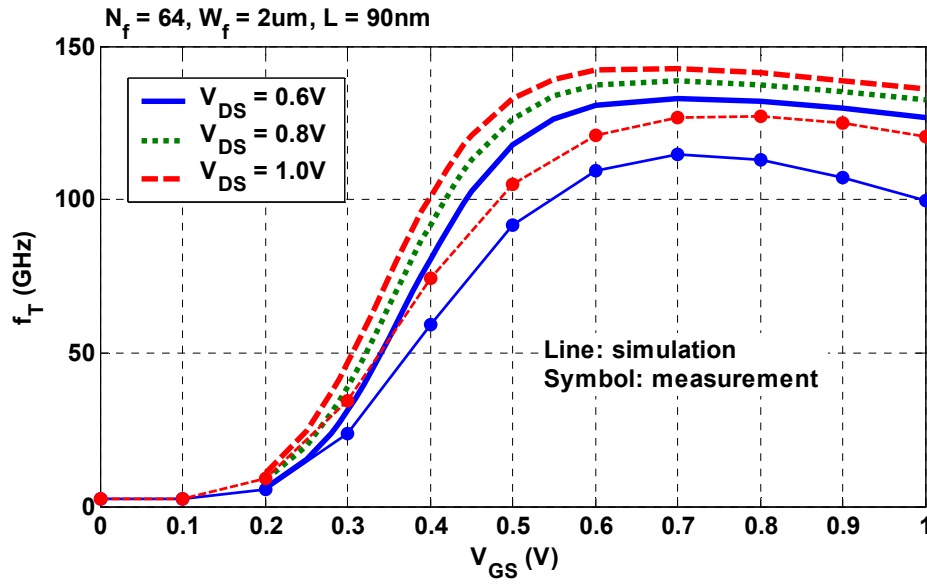


Fig. 7.6 f_T extracted from measured and simulated S-parameters.

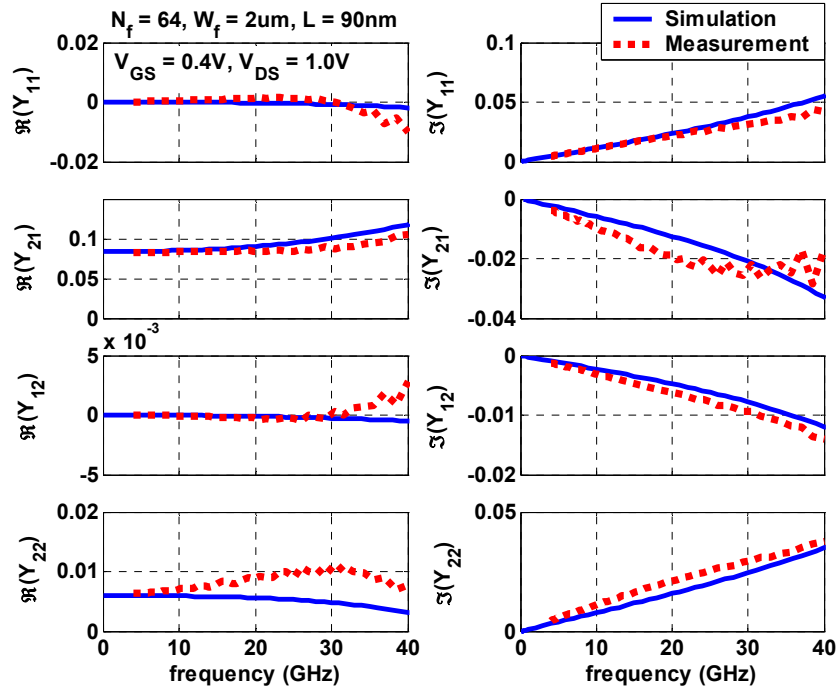


Fig. 7.7 Y-parameters versus frequency at $V_{GS} = 0.4\text{V}$ and $V_{DS} = 1.0\text{V}$. \Re and \Im stand for real and imaginary parts.

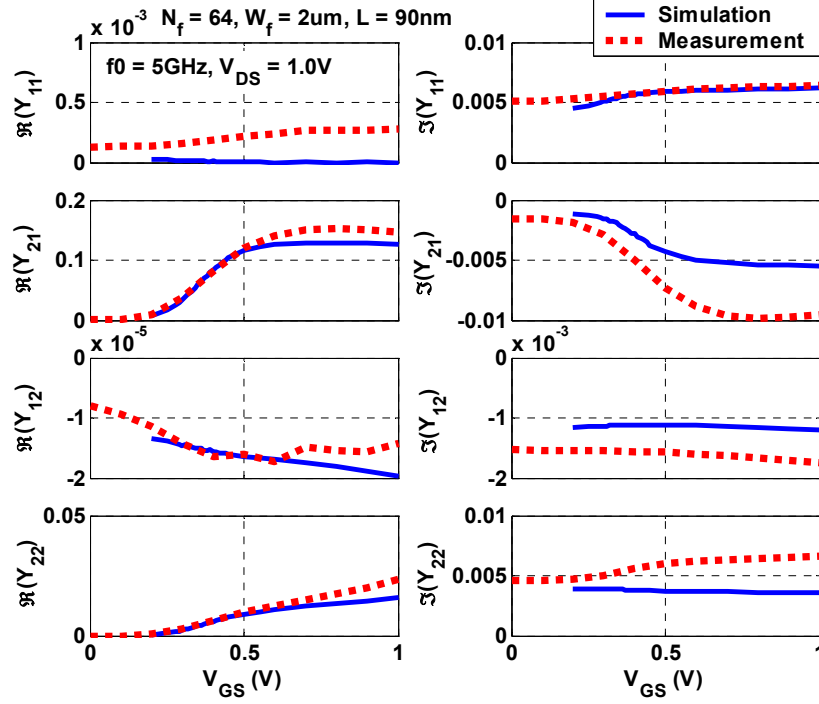


Fig. 7.8 Y-parameters at 5GHz versus V_{GS} . $V_{DS}=1.0V$. \Re and \Im stand for real and imaginary parts.

7.3 Nonlinear characteristics

In real applications, the nonlinearities of the transistors and other components can introduce undesired harmonic products to the output signal. Since the amplitudes of harmonics and intermodulation products are higher order functions of the amplitude of the input signal, and the amplitude of the fundamental signal is a linear function of the input power level, the amplitudes of the harmonics and intermodulation products increase in a much faster way than the fundamental signal amplitudes as input power increases. Fig. 7.9 compares the amplitudes of the measured and simulated output signals at the fundamental frequency. The transistor is biased in moderate inversion

region with $V_{GS}=0.4V$, $V_{DS}=0.8V$. At low input power, the higher order products are much less than the fundamental signal, so the power gain from measurement and simulation are both approximately constant. As P_{in} increases, power gain starts to drop because of the amplitude of the harmonics and intermodulation products added to the desired signal are negative. The gain drop in the simulated result is clearly observed in Fig. 7.9. Because of the limitation of the maximum power level that can be generated by the signal generators, the P_{in} in measurement is not high enough to show this gain drop obviously. The 1dB compression point, where power gain drops by 1dB, can be determined as illustrated in Fig. 7.9.

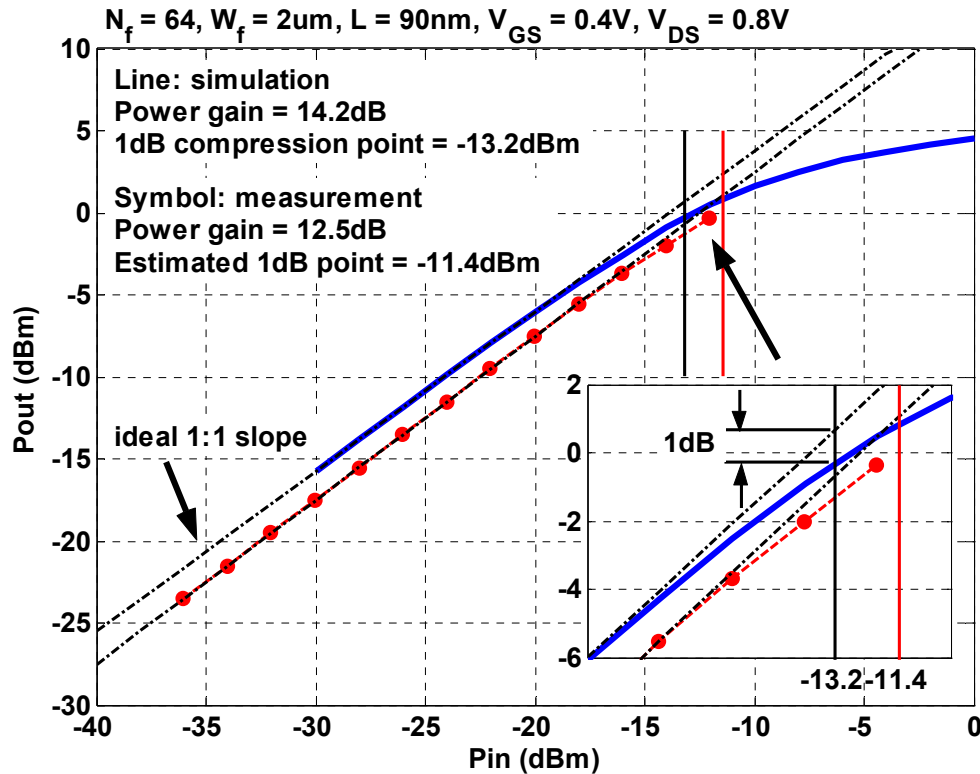


Fig. 7.9 The amplitude of the fundamental output signal versus input power level at $V_{GS}=0.4V$, $V_{DS}=0.8V$.

Fig. 7.10 shows the power level for the fundamental signal and the third order intermodulation product versus drain current density for multiple drain biasing voltages. Due to the limitation of the maximum power level can be generated in experiments, the results are for $P_{in} = -28\text{dBm}$, which means the transistor is operated in linear mode. Only when P_{in} is larger than -10dBm , the transistor is driven into nonlinear mode. However, as shown in Fig. 7.9, it is hard to realize in experiments. Therefore, the nonlinear RF modeling performance is not directly evaluated by comparing the output power level. Instead, IP3 is extracted from low P_{in} measurement and used as an indicator for the linearity of the transistor.

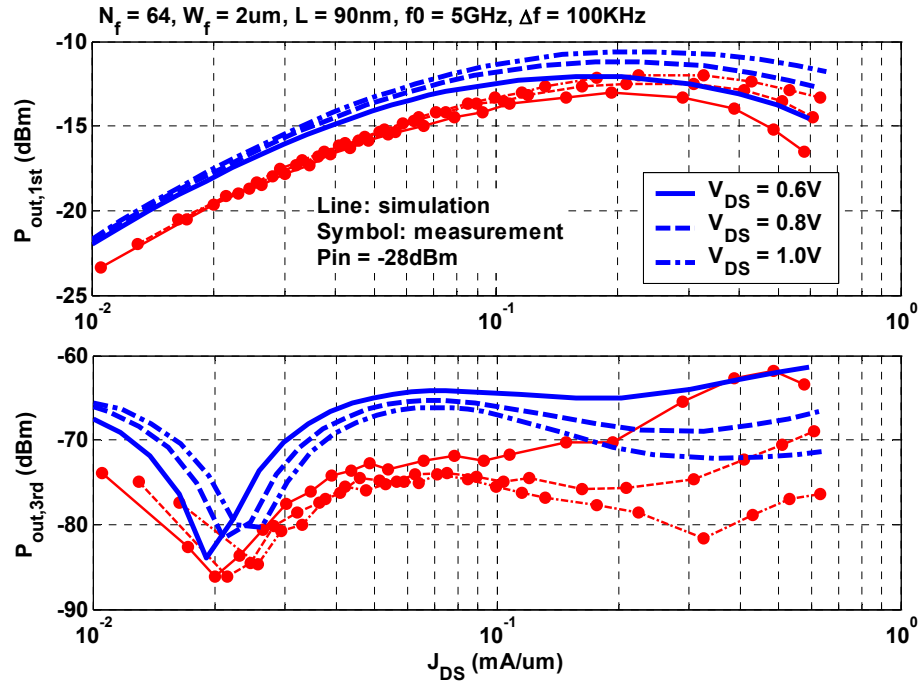


Fig. 7.10 The amplitude of the fundamental output signal and the third order intermodulation product versus J_{DS} .

Fig. 7.11 shows input IP3, IIP3, from measurement and simulation for multiple V_{DS} . $N_f = 64$. With V_{DS} increasing from 0.6 to 1.0V, I_{DS} increases only slightly, but IIP3 increases by a much larger factor, particularly at higher V_{GS} when the device is biased closer to linear operation region. And, the IP3 sweet spot V_{GS} decreases as V_{DS} increases. The V_{DS} dependence of IP3 sweet spot is determined by the threshold voltage change due to DIBL, which will be verified using simulation results with and without DIBL induced V_{th} change in Section 8.5.

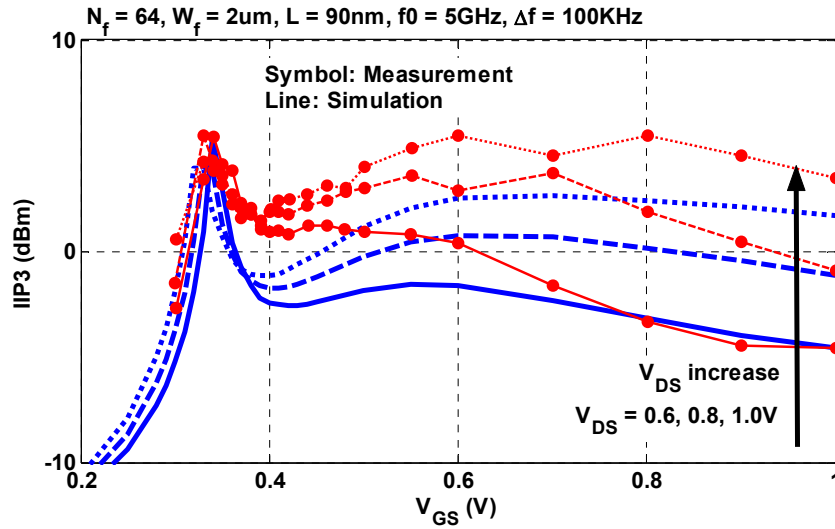


Fig. 7.11 Measured and simulated IIP3 versus V_{GS} at multiple V_{DS} .

Fig. 7.12 shows IIP3 from measurement and simulation versus J_{DS} for $N_f = 10, 20$, and 64. The peak of measured IIP3 is not perfect because IP3 is not measured in a very fine biasing step. Linearity simulation can predict IP3 sweet spot accurately for devices with $N_f = 10, 20$, and 64. Note that the sweet spot J_{DS} decreases from $35\mu\text{A}/\mu\text{m}$ for

$N_f=10$ to $20\mu\text{A}/\mu\text{m}$ for $N_f=64$. Note that, the zero K_{3g_m} point is marked as it is the IP3 sweet spot estimated using first order IP3 theory. The zero K_{3g_m} points for the three transistors are practically the same since the devices scale well. So, just the zero K_{3g_m} point for $N_f=64$ is shown in Fig. 7.12. The deviation between the actual IP3 sweet spot and the zero K_{3g_m} point increases as device size increases.

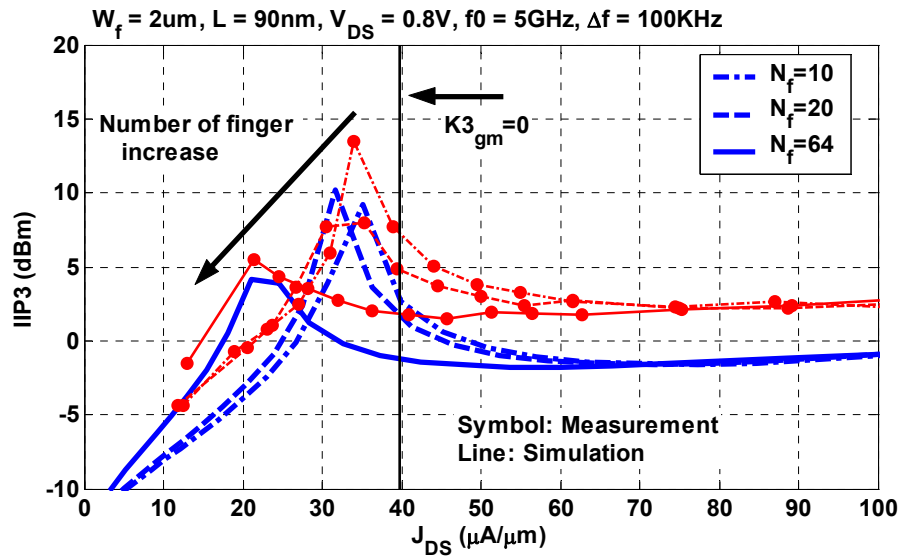


Fig. 7.12 Measured and simulated IIP3 versus J_{DS} for devices with $N_f=10, 20$, and 64 .

Fig. 7.13 and Fig. 7.14 show the measured IIP3 at 2, 5, and 10GHz for devices with $N_f=10$ and 64 (a total width of $20\mu\text{m}$ and $128\mu\text{m}$). For $N_f=10$, IIP3 at 2, 5, and 10GHz are practically identical. For $N_f=64$, IIP3 increases as frequency increases. This frequency dependence can be attributed to capacitive components in the transistor as

detailed in Section 9.2.3. The frequency dependence of simulated IIP3 in Fig. 7.13 and Fig. 7.14 is similar to the frequency dependence of measured IIP3.

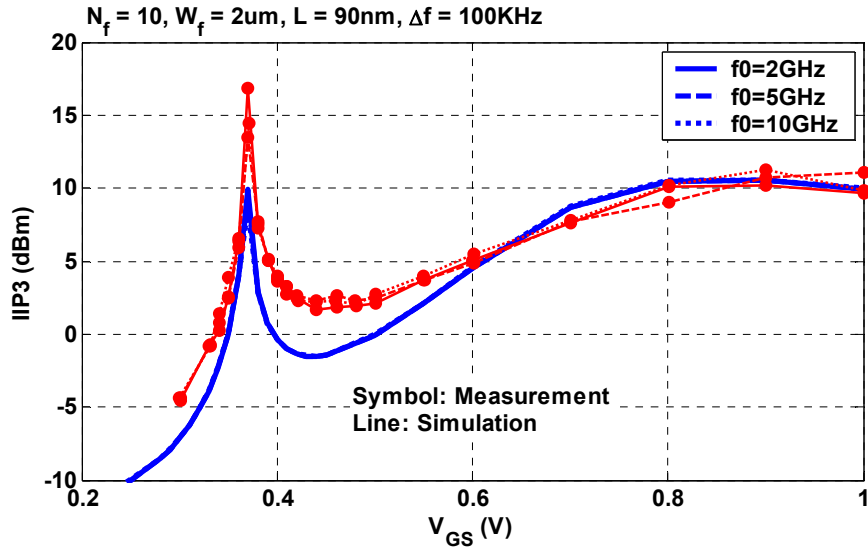


Fig. 7.13 Measured and simulated IIP3 versus V_{GS} at multiple frequencies for $N_f=10$ ($W=20\mu\text{m}$).

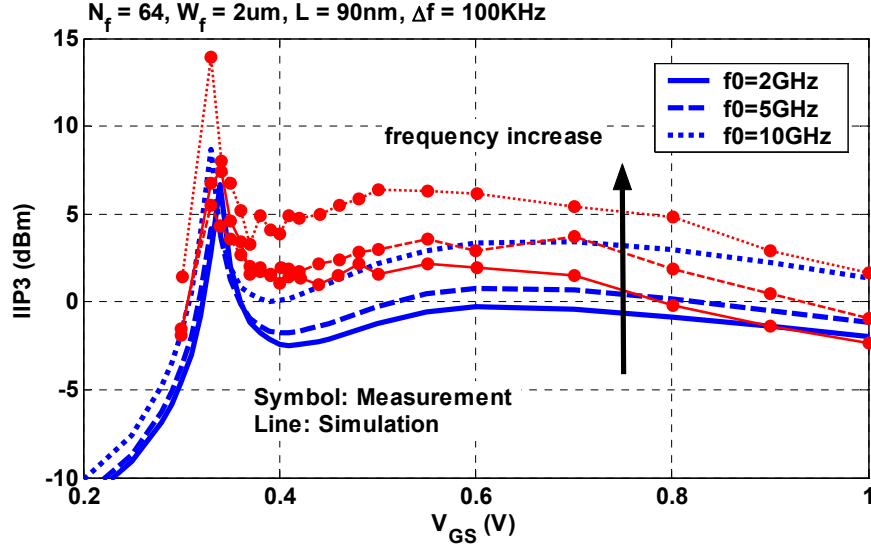


Fig. 7.14 Measured and simulated IIP3 versus V_{GS} at multiple frequencies for $N_f=64$ ($W=128\mu m$).

7.4 Summary

This chapter evaluates the BSIM4 model of the NMOS transistor for linear and nonlinear performance using a set of DC, S-parameter, and power spectrum measurements, especially in the moderate inversion region. The results demonstrate good fittings on both DC and AC characteristics. Despite its interpolating nature of moderate inversion modeling, the BSIM4 model can accurately describe I-V, and Y-parameters in moderate inversion region. The subcircuit based BSIM4 model can predict the distortion behavior of a CMOS transistor correctly, which enables distortion optimization of RFIC circuits using mathematical models and simulators. The linearity sweet spot is investigated to be deviated significantly from the widely accepted zero K_{3g_m} point for a practically large device size found in LNAs.

CHAPTER 8

MODELING OF INTERMODULATION LINEARITY

An important consideration in RFIC design is linearity, which sets the upper limit of spurious free dynamic range. Among various linearity measures, the two-tone third order intermodulation distortion (IM3) is the most widely used, and is typically characterized by the third order intercept point (IP3) [11]. Using either measured or simulated I-V data, IP3 sweet spot biasing current can be determined from zero K_{3g_m} point based on first order IP3 theory [11]. K_{3g_m} is defined as the 3rd order coefficient of the nonlinear transconductance. Circuits have been published to utilize this zero K_{3g_m} point for high linearity LNA designs [48] [52].

However, it was shown in [53] that the IP3 sweet spot from measurement and simulation both shift to a lower V_{GS} than the zero K_{3g_m} point and the first order IP3 theory cannot correctly model the biasing and device scaling dependence of IIP3. More accurate analytical IP3 expressions need to be developed. The complete IP3 expression in this work is developed using the nonlinear current source method based on Volterra series. The nonlinear drain current source includes nonlinear transconductance, output conductance and the cross terms. The IP3 expression published in [22] [25] [26] are just special cases of this complete IP3 expression. It will be shown later that the cross

terms ignored in [22] [25] are important for accurate IP3 modeling and are responsible for the V_{DS} dependence of IP3 sweet spot V_{GS} to drain induced barrier lowering (DIBL).

Linearity simulation results using BSIM4 model are compared with calculated results. For the frequencies examined in this work, 2GHz, 5GHz, and 10GHz, open-short de-embedding is valid for the layout design used. Thus, the pads and interconnections are modeled using open-short equivalent circuit consisting of three series and three shunt elements in Cadence. However, this added parasitics network does not affect IIP3 that much.

8.1 First order IP3 theory

Fig. 8.1 shows the small signal equivalent circuit used for analytical IP3 analysis. $v_s = V_s (\cos \omega_1 t + \cos \omega_2 t)$ is the two tone input signal. $\omega_1 = 2\pi f_1$ and $\omega_2 = 2\pi f_2$. f_1 and f_2 are the frequencies of the two-tone excitation spacing by $\Delta f = 100\text{KHz}$. R_s is the source resistance, while R_L is the load resistance. Here R_s and R_L are both 50Ω . C_{gs} and C_d are the small-signal gate to source capacitance and drain to substrate capacitance with values extracted from S-parameters. First order IP3 theory considers nonlinear transconductance only. The linear and the second- and third- order nonlinear transconductance can be identified with the coefficients of Taylor expansion as

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}}, K_{2g_m} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2}, K_{3g_m} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3}. \quad (8.1)$$

The small-signal nonlinear current source i_{ds} can then be approximated by the first three order nonlinearities as

$$i_{ds} = g_m v_{gs} + K_{2g_m} v_{gs}^2 + K_{3g_m} v_{gs}^3. \quad (8.2)$$

The first order input referred IP3 (IIP3) for the small-signal equivalent circuit in Fig. 8.1 is then calculated as

$$IIP3 = \frac{1}{6R_s} \frac{1 + (\omega C_{gs} R_s)^2}{\left| \frac{K_{3g_m}}{g_m} \right|}. \quad (8.3)$$

The derivation is detailed in Appendix J.

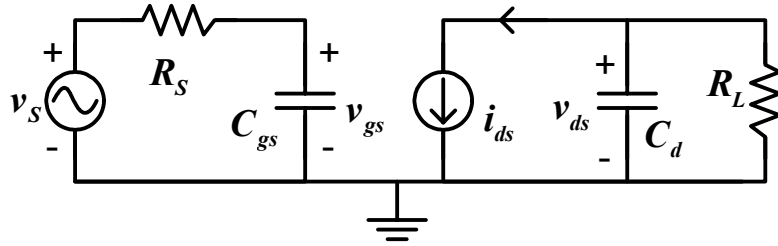


Fig. 8.1 The small signal equivalent circuit used for IP3 analysis.

When $K_{3g_m} = 0$, first order IP3 gives the maximum IIP3, which is the well known IP3 sweet spot used to improve linearity in circuit designs. Fig. 8.2 plots g_m , K_{3g_m} , and the first order IP3 calculated using (8.3). A sharp IIP3 peak is observed near the threshold voltage, during the transition from subthreshold to strong inversion when K_{3g_m} becomes zero. However, the sweet spot IIP3 is not necessarily the highest. The calculated IIP3 can be higher in strong inversion since g_m saturates and K_{3g_m} is very

small as V_{GS} increases. Experimental results also show that IIP3 varies strongly with V_{DS} at sweet spot position and high V_{GS} as detailed in Section 9.2.

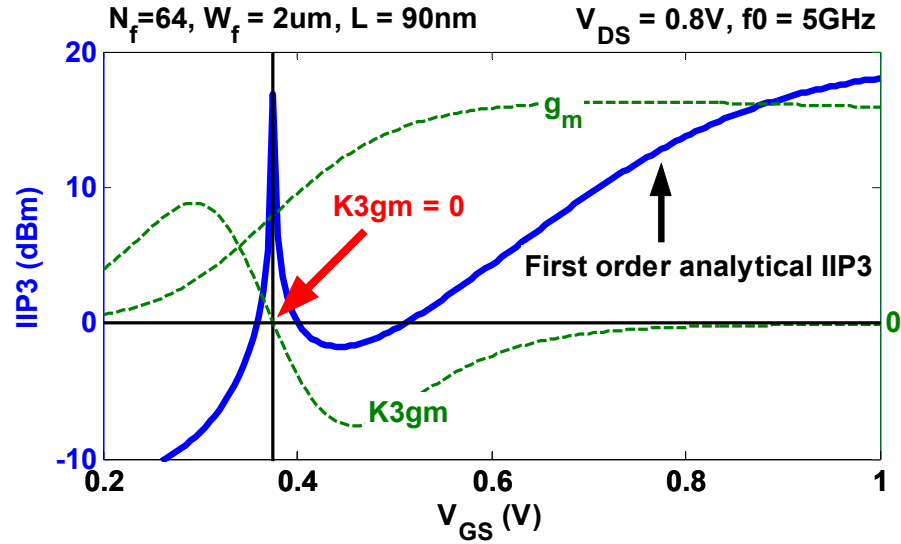


Fig. 8.2 First order IP3 with a sweet spot at $K_{3g_m}=0$.

8.2 Complete IP3 expression

An IP3 expression including all of the nonlinear coefficients of order 3 and below is derived using Volterra series. The nonlinear current source method together with the small signal equivalent circuit in Fig. 8.1 is used to calculate the Volterra kernels [24]. Although C_{gd} is not included in Fig. 8.1, Volterra series analysis with C_{gd} can also be done. The expression with C_{gd} is too complicated and thus not shown. The IP3 expression without C_{gd} is sufficient for understanding the biasing and device size dependence of IP3. Only when frequency is high and device size is large, C_{gd} is needed as illustrated in Section 9.2.3.

8.2.1 Two dimension nonlinear current source

The nonlinear current source i_{ds} in Fig. 8.1 controlled by gate-source and drain-source voltages is written as [24]

$$\begin{aligned} i_{ds} = & g_m v_{gs} + K_{2g_m} v_{gs}^2 + K_{3g_m} v_{gs}^3 \\ & + g_o v_{ds} + K_{2g_o} v_{ds}^2 + K_{3g_o} v_{ds}^3 \\ & + K_{2g_m g_o} v_{gs} v_{ds} + K_{32g_m g_o} v_{gs}^2 v_{ds} + K_{3g_m 2g_o} v_{gs} v_{ds}^2 \end{aligned} \quad (8.4)$$

g_m and g_o are the linear transconductance and output conductance. K_{2g_m} and K_{3g_m} are the 2nd and 3rd order nonlinear transconductance, while K_{2g_o} and K_{3g_o} are the 2nd and 3rd order nonlinear output conductance. $K_{2g_m g_o}$, $K_{32g_m g_o}$, and $K_{3g_m 2g_o}$ are the 2nd, and 3rd order cross terms. The nonlinearity coefficients are defined in Table 8.1.

Table 8.1 Definition of nonlinearity coefficients of nonlinear drain current.

Transconductance	Output Conductance	Cross term
$g_m = \frac{\partial I_{DS}}{\partial V_{GS}}$	$g_o = \frac{\partial I_{DS}}{\partial V_{DS}}$	$K_{2g_m g_o} = \frac{\partial^2 I_{DS}}{\partial V_{GS} \partial V_{DS}}$
$K_{2g_m} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2}$	$K_{2g_o} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{DS}^2}$	$K_{32g_m g_o} = \frac{\partial^3 I_{DS}}{\partial V_{GS}^2 \partial V_{DS}}$
$K_{3g_m} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3}$	$K_{3g_o} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{DS}^3}$	$K_{3g_m 2g_o} = \frac{\partial^3 I_{DS}}{\partial V_{GS} \partial V_{DS}^2}$

Fig. 8.3 shows all of the nonlinear coefficients versus V_{GS} at $V_{DS}=0.8V$ for a large device width used in practical circuits, $W=128\mu m$. The zero K_{3g_m} V_{GS} is marked because it is the IP3 sweet spot estimated from first order IP3 theory. All of the

derivatives are calculated from simulated I-V data. The I-V data are simulated using BSIM4 model in Cadence and exported with 12 digits to ensure accurate numerical evaluation of 2nd and 3rd order derivatives. Fig. 8.3 (a) and (d) show the linear transconductance and output conductance versus V_{GS} . Fig. 8.3 (b) and (c) show the 2nd and 3rd order nonlinear transconductance, while Fig. 8.3 (e) and (f) show the 2nd and 3rd order nonlinear output conductance. Fig. 8.3 (g)-(i) are the 2nd and 3rd order cross terms. Compared with the cross terms, the output conductance nonlinearities are much smaller, especially at the sweet spot. The impact of cross terms on IIP3 sweet spot location should be negligible, which is evaluated numerically in Section 8.3.

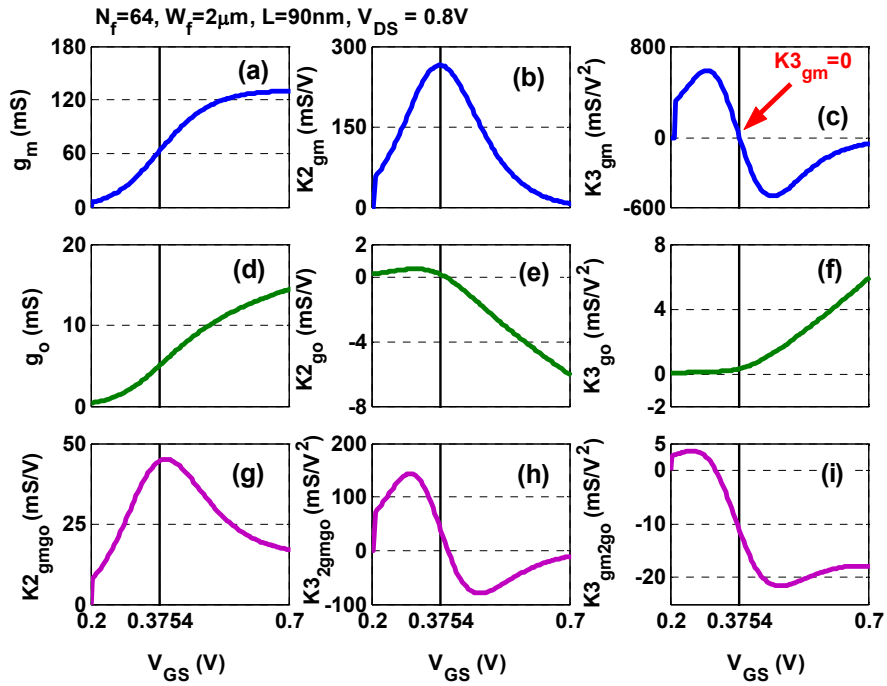


Fig. 8.3 The nonlinear coefficients versus V_{GS} .

8.2.2 Input IP3 expression

Using Volterra series analysis, the complete IIP3 expression with i_{ds} in (8.4) is derived as

$$IIP3 = \frac{1}{6R_s} \frac{1 + (\omega C_{gs} R_s)^2}{\left| \frac{K_{3g_m}}{g_m} + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 \right|}. \quad (8.5)$$

The derivation of (8.5) is detailed in Appendix J. The first term in the denominator, K_{3g_m} / g_m , is due to the 3rd order transconductance as found in first order IP3 expression (8.3). The other terms containing nonlinear output conductance and cross terms are grouped as Δ_1 , Δ_2 , Δ_3 , and Δ_4 .

$$\Delta_1 = -\frac{1}{3} K_{2g_m g_o} \frac{K_{2g_m}}{g_m} Z_1 - \frac{1}{3} (K_{3g_m g_o}) Z_2, \quad (8.6)$$

$$\Delta_2 = \frac{2}{3} K_{2g_m} K_{2g_o} Z_3 + \frac{1}{3} (K_{3g_m^2 g_o}) g_m Z_4 + \frac{1}{3} (K_{2g_m g_o})^2 Z_5, \quad (8.7)$$

$$\Delta_3 = -K_{3g_o} g_m^2 Z_6 - \frac{1}{3} K_{2g_m g_o} K_{2g_o} g_m Z_7, \quad (8.8)$$

$$\Delta_4 = \frac{2}{3} (K_{2g_o})^2 g_m^2 Z_8, \quad (8.9)$$

where

$$Z_1 = Z_L (2\omega_1) + 2Z_L (\omega_1 - \omega_2),$$

$$Z_2 = Z_L (\omega_1) [Y_s (-\omega_2) Y_s^{-1} (\omega_1) + 2],$$

$$Z_3 = 2Z_L (\omega_1 - \omega_2) Z_L (\omega_1) + Z_L (2\omega_1) Z_L (-\omega_2),$$

$$Z_4 = Z_L^2 (\omega_1) [2Y_s (-\omega_2) Y_s^{-1} (\omega_1) + 1],$$

$$Z_5 = 2Z_L(\omega_1 - \omega_2)Z_L(-\omega_2) + Z_L(2\omega_1)Z_L(\omega_1),$$

$$Z_6 = Z_L^2(\omega_1)Z_L(-\omega_2),$$

$$Z_7 = Z_L^2(\omega_1)Z_L(2\omega_1) + 2Z_L(\omega_1)Z_L(\omega_2)Z_L(2\omega_1) + 6Z_L(\omega_1)Z_L(\omega_2)Z_L(\omega_1 - \omega_2),$$

$$Z_7 = Z_L^2(\omega_1)[Z_L(2\omega_1) + 2Z_L(\omega_2) + 6Z_L(\omega_1 - \omega_2)],$$

$$Z_8 = Z_L^2(\omega_1)Z_L(-\omega_2)[Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2)],$$

$$Z_L(\omega) = \frac{1}{g_o + j\omega C_d + 1/R_L}, \quad Y_s(\omega) = \frac{1}{R_s} + j\omega C_{gs}.$$

Δ_1 , Δ_2 , Δ_3 , and Δ_4 are functions of the 2nd and 3rd order nonlinear output conductance and cross terms. The values of the cross terms, especially $K_{2g_m g_o}$ and $K_{3g_m g_o}$ in Δ_1 , are comparable to g_m in moderate inversion region as shown in Fig. 8.3. In strong inversion region, g_m saturates, and K_{3g_m} reduces to zero. Δ_1 , Δ_2 , Δ_3 , and Δ_4 will be comparable to K_{3g_m} / g_m even if they are close to zero as we will show below. This indicates that Δ_1 , Δ_2 , Δ_3 , and Δ_4 are all important for IP3 modeling. Therefore, the IP3 expressions without cross terms in [22] [25] are not accurate enough. Note that the complete IP3 expression derived in [26] is a special case of (8.5) at low frequencies. Furthermore, the numerical results in [26] were calculated by neglecting various nonlinear terms and the derivatives in the nonlinear coefficients were calculated from an approximate drain current function instead of a complete MOSFET model. Here, the numerical results are all calculated using the complete IIP3 expression in (8.5), and all of the derivatives are calculated using simulated I-V data from a BSIM4 model.

An inspection of (8.6)-(8.9) shows that Δ_1 , Δ_2 , Δ_3 , and Δ_4 depend on Z_L and Y_S as well. The load impedance will thus affect IP3 sweet spot when it dominates Z_L . In this work, IP3 is only examined for a 50Ω load due to its practical relevance and straightforward measurement. When $\Delta_1=\Delta_2=\Delta_3=\Delta_4=0$, the complete IP3 expression of (8.5) reduces to the first order IP3 of (8.3).

First order IP3 does not scale as device size scales because the scaling factors of K_{3g_m} and g_m are cancelled and $(\omega C_{gs} R_s)^2$ is much smaller than 1 for the devices used. Δ_1 , Δ_2 , Δ_3 , and Δ_4 have quite different scaling factors as device size scales as shown in Section 8.4.

8.3 Impact of the additional terms

Instead of the zero K_{3g_m} point in (8.3), IIP3 peaks at the point where the denominator of (8.5) is zero. Fig. 8.4 (a) shows the denominator in (8.5) versus V_{GS} , which is the sum of K_{3g_m}/g_m , Δ_1 , Δ_2 , Δ_3 , and Δ_4 , while Fig. 8.4 (b) shows K_{3g_m}/g_m , Δ_1 , Δ_2 , Δ_3 , and Δ_4 as a function of V_{GS} individually. $V_{DS}=0.8V$. The IP3 sweet spot from (8.5), 0.327V, is much lower than the zero K_{3g_m} point, 0.3754V. Δ_1 and Δ_2 are the two largest terms that affect the shift of the IP3 sweet spot. Δ_3 and Δ_4 have very little impact on IP3 sweet spot since they are much smaller than Δ_1 and Δ_2 near the zero K_{3g_m} point. Since Δ_1 is negative and Δ_2 is positive around the zero K_{3g_m} V_{GS} , the impact of Δ_1 and Δ_2 on the shift of IP3 sweet spot are opposite. Δ_1 moves IP3 sweet spot

to lower V_{GS} , and Δ_2 moves IP3 sweet spot to higher V_{GS} when compared with the zero $K_{3g_m} V_{GS}$. The Δ_3 and Δ_4 terms, however, are important at higher V_{GS} .

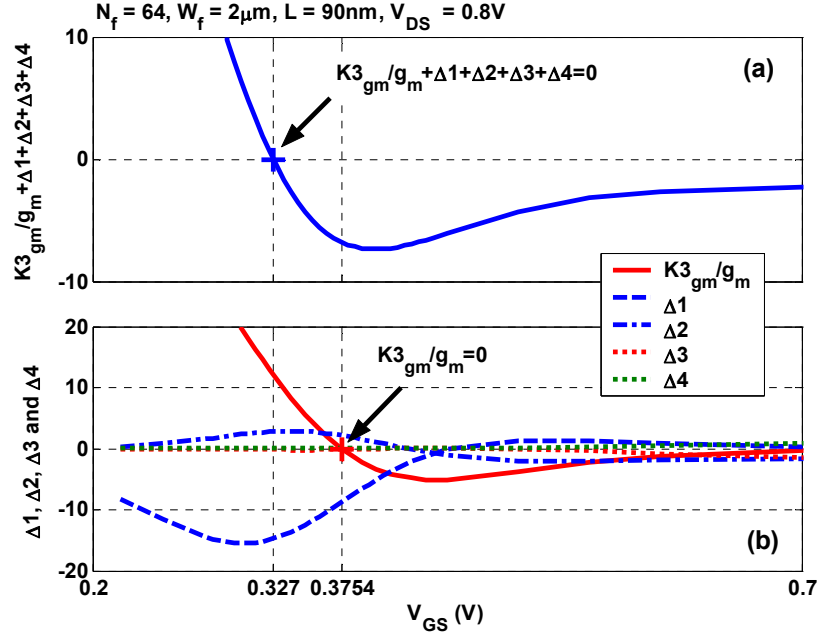


Fig. 8.4 (a) The denominator in (8.5) versus V_{GS} . (b) Each term in the denominator of (8.5) versus V_{GS} . $V_{DS}=0.8V$.

Fig. 8.5 shows the impacts of Δ_1 , Δ_2 , Δ_3 , and Δ_4 on IIP3. Since Δ_1 is much larger than Δ_2 , a V_{GS} lower than the zero $K_{3g_m} V_{GS}$ is observed at IP3 sweet spot. Although the deviation between IP3 sweet spot V_{GS} and zero $K_{3g_m} V_{GS}$ is dominated by Δ_1 , adding the other three elements can model IP3 sweet spot better. At high V_{GS} , K_{3g_m}/g_m is close to zero, and is comparable with the value of Δ_1 , Δ_2 , Δ_3 , and Δ_4 as shown in Fig. 8.4 (b). Thus, Δ_1 , Δ_2 , Δ_3 , and Δ_4 all affect the value of IP3 at high V_{GS} .

significantly as shown in Fig. 8.5. Therefore, all of the nonlinear coefficients are important for IP3 modeling including cross terms.

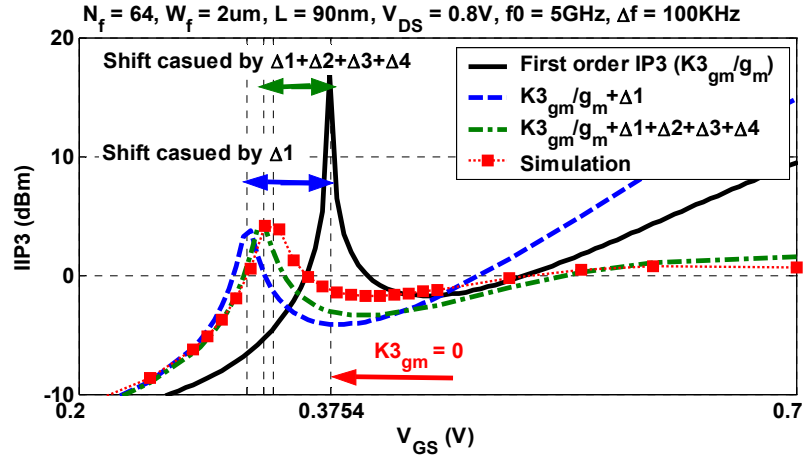


Fig. 8.5 IIP3 versus V_{GS} from simulation, first order IP3 expression in (8.3), and complete IP3 expression in (8.5) with different nonlinearities included. $V_{DS}=0.8V$.

8.4 Device width scaling

The linear and nonlinear coefficients in (8.5) all scale by a factor of K as device size scales by K . The scaling factors of the Z terms are complicated. For very small devices, Z_L is approximately R_L , and Y_S is approximately $1/R_S$. Therefore, the scaling factors for Δ_1 , Δ_2 , Δ_3 , and Δ_4 are K , K^2 , K^3 , and K^4 respectively. In the extreme case, if Z_L and Y_S are dominated by g_o , C_d and C_{gs} , Δ_1 , Δ_2 , Δ_3 , and Δ_4 do not scale as device size scales. For the device sizes and frequencies examined in this work, the C_d and C_{gs} terms are relatively small, and the scaling factors for Δ_1 , Δ_2 , Δ_3 , and Δ_4 are close to K , K^2 , K^3 , and K^4 . This indicates that the impact of Δ_1 , Δ_2 , Δ_3 , and Δ_4 on IP3 sweet spot is

much stronger for large devices. As N_f increases, a decrease of IP3 sweet spot V_{GS} is expected.

Fig. 8.6 shows the calculated IIP3 using (8.5) versus J_{DS} for devices with multiple finger numbers. Drain current density J_{DS} is defined as I_{DS}/W . First order IIP3 is shown for comparison. As $J_{DS} - V_{GS}$ is nearly identical for varying N_f , the calculated first order IP3 for varying N_f are perfectly overlapped. For $N_f = 1$, IP3 from (8.5) is almost the same as first order IP3, and the sweet spot is at the zero K_{3g_m} point. As device size increases, the deviation between IP3 calculated using (8.5) and first order IP3 increases. J_{DS} of the IP3 sweet spot is the lowest for the largest N_f .

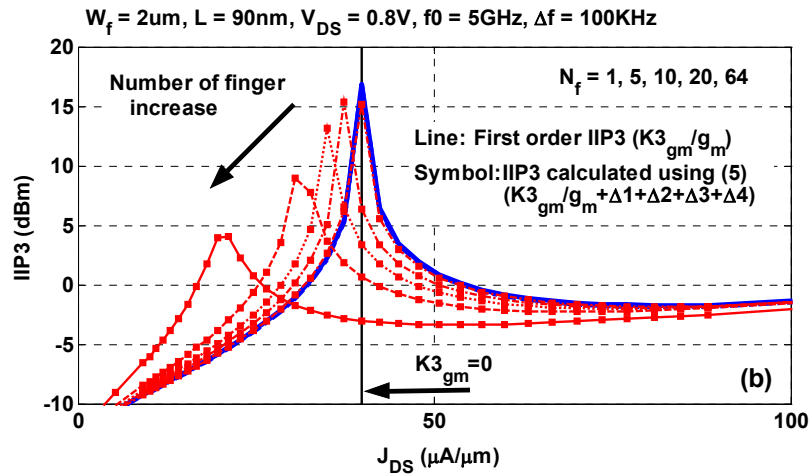


Fig. 8.6 IIP3 calculated using (8.5) and (8.3) versus J_{DS} for devices with multiple finger numbers.

8.5 DIBL effect

The V_{DS} dependence of IP3 sweet spot V_{GS} is a direct result of the DIBL introduced threshold voltage change from previous analysis. The threshold voltage change caused by DIBL is modeled using ΔV_{th} (DIBL) in BSIM4 model [2]. To further investigate the impact of DIBL, simulation results with and without ΔV_{th} (DIBL) are compared. ΔV_{th} (DIBL) is turned off by setting corresponding model parameters to zero.

Fig. 8.7 (a) shows I_{DS} - V_{GS} results from simulation with and without ΔV_{th} (DIBL). Without ΔV_{th} (DIBL), I_{DS} at low V_{GS} for different V_{DS} are close. The zero K_{3g_m} points are therefore close for different V_{DS} as shown in Fig. 8.7 (b). Fig. 8.8 shows simulated IIP3 with and without ΔV_{th} (DIBL) for $N_f=64$. Without DIBL, the V_{DS} dependence of IP3 sweet spot V_{GS} is dramatically reduced.

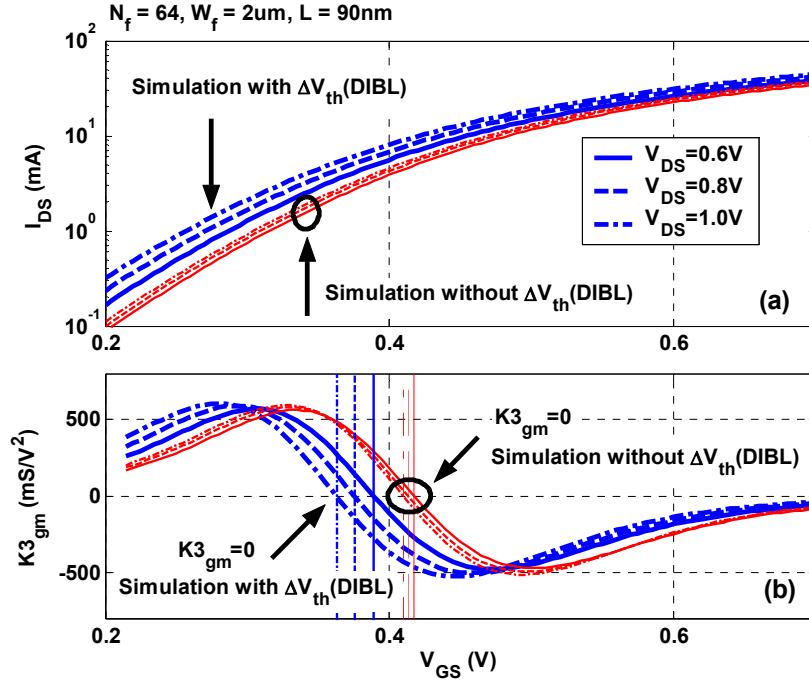


Fig. 8.7 (a) I_{DS} , (b) K_{3gm} versus V_{GS} at multiple V_{DS} for simulation with and without V_{th} shift due to ΔV_{th} (DIBL).

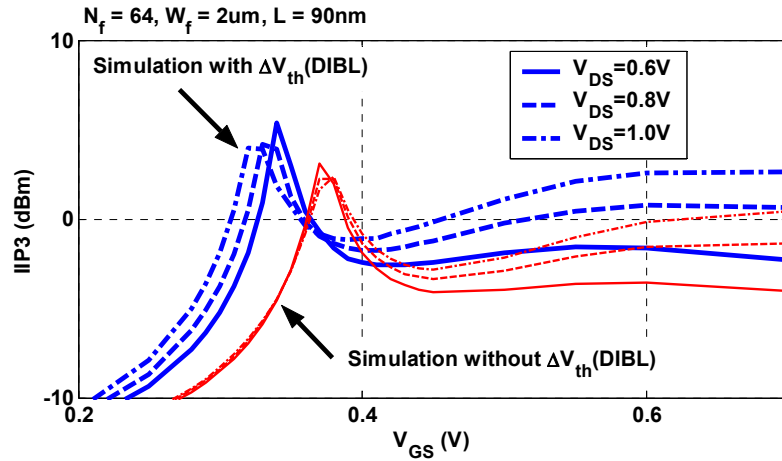


Fig. 8.8 IIP3 calculated using (8.5) versus V_{GS} at multiple V_{DS} for simulation with and without ΔV_{th} (DIBL).

Fig. 8.9 (a) compares the denominator of (8.5) for simulations with and without ΔV_{th} (DIBL). Fig. 8.9 (b) and (c) compare the most important terms in the denominator of (8.5), K_{3g_m} / g_m and $\Delta_1 + \Delta_2$ individually. Δ_3 and Δ_4 are not shown here because they have very little impact on IP3 sweet spot. The variation of $\Delta_1 + \Delta_2$ with V_{DS} is approximately the same for simulation with and without ΔV_{th} (DIBL). Thus, the impact of $\Delta_1 + \Delta_2$ on the V_{DS} dependence of the deviation of IP3 sweet spot V_{GS} from zero K_{3g_m} point is approximately the same for simulation with and without ΔV_{th} (DIBL). The V_{DS} dependence of IP3 sweet spot V_{GS} is thus mainly a result of the V_{DS} dependence of the zero K_{3g_m}

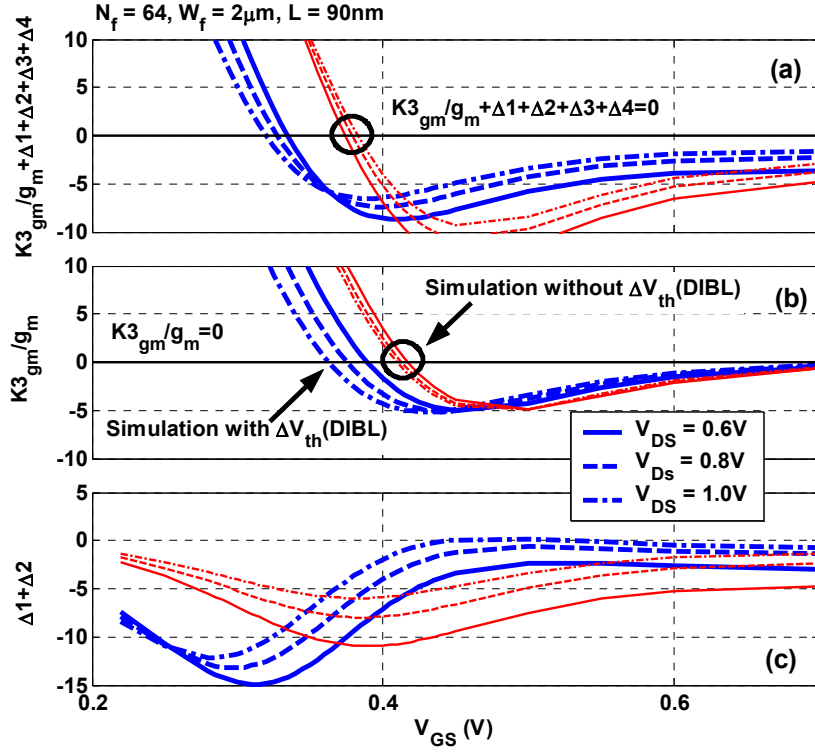


Fig. 8.9 (a) $K_{3g_m}/g_m + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$, (b) K_{3g_m}/g_m , and (c) $\Delta_1 + \Delta_2$ versus V_{GS} at multiple V_{DS} for Cadence simulation with and without ΔV_{th} (DIBL).

8.6 Summary

This chapter develops a complete analytical IP3 expression which involves not only nonlinear transconductance but also nonlinear output conductance and cross terms. The deviation of the sweet spot V_{GS} from the widely accepted zero K_{3g_m} point for a practically large device size found in LNAs is attributed to output conductance nonlinearities and cross terms through this expression. The impacts of these additional terms added to K_{3g_m}/g_m are examined individually to figure out the dominant factor for IP3 sweet spot shift. The significance of the additional terms scales with device

width. Therefore, the deviation of IP3 sweet spot from zero K_{3g_m} point is the largest, and IP3 sweet spot J_{DS} is the lowest for the largest device. In the 90nm CMOS technology used, the sweet spot J_{DS} decreases from 40 to $20\mu\text{A}/\mu\text{m}$ as gate width increases from 2 to $128\mu\text{m}$. For large devices of interest to RFIC design, simulation using a good model and measurement of IP3 must be used to accurately identify the sweet spot biasing current density. Simulation results without ΔV_{th} (DIBL) indicate that the V_{DS} dependence of IP3 sweet spot V_{GS} is dominated by the threshold voltage shift caused by DIBL effect.

CHAPTER 9

CHARACTERIZATION OF RF INTERMODULATION LINEARITY

After developing the complete IP3 expression in Section 8.2, IP3 sweet spot of single transistor and simple circuits can be estimated using measured I-V and S-parameters of single transistor, instead of two-tone intermodulation measurement. However, because of the I-V data measured using Agilent 4155 only has 5 digits, which is not sufficient to provide smooth 2nd and 3rd order nonlinearity coefficients in Table 8.1, the calculated IP3 are all from simulated I-V and S-parameters using the BSIM4 model examined in Chapter 7. The same set of equations can be applied on measured I-V and S-parameters once the measured data has enough digits.

This chapter compares the measured and calculated IP3 results for a 90nm RF CMOS technology. The complete IP3 expression developed in Section 8.2 is used. I-V data and device small signal parameters are extracted from DC and S-parameter simulations using the BSIM4 model validated in Chapter 7. The IP3 expression can accurately predict the biasing and device size dependence of IP3 sweet spot. The frequency dependence of IP3 is determined by the small signal capacitance. Thus, the frequency dependence is very weak and negligible for small device. For large device, not only gate-source capacitance and drain-bulk capacitance but also gate-drain capacitance are important. To determine the value of IP3 accurately, more complete equivalent circuit of MOS transistor must be used in Volterra series analysis.

9.1 Power gain measurement

Of particular interest to linearity measurement is the power gain. Since 50Ω source and load are used in IP3 measurement, at low input power, the power gain obtained from sweeping input power linearity measurement should agree with the small signal power gain S_{21} from S-parameter measurement, which involves much more systematic error correction. Therefore, the power gain (at low P_{in}) from linearity measurement using spectrum analyzer with S_{21} from S-parameter measurement using VNA, are compared in Fig. 9.1, as a means of assuring power calibration accuracy for linearity measurement. The power gains extracted from intermodulation measurement are close to S_{21} from S-parameter measurement within 0.5dB for most measurements in this dissertation.

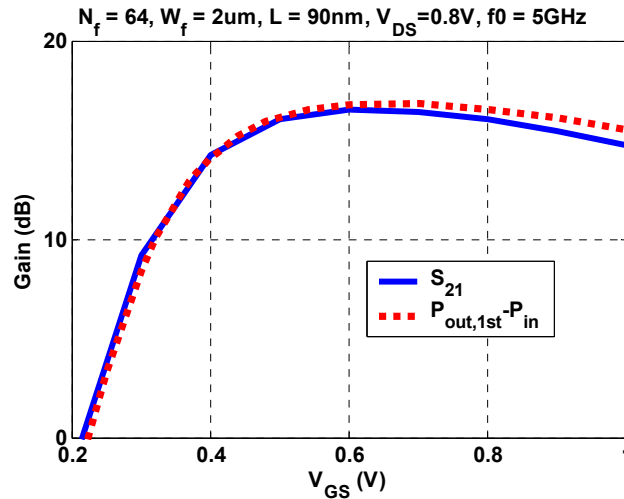


Fig. 9.1 Gain from linearity measurement ($P_{out,1st} - P_{in}$) and gains-parameter measurement (S_{21}) versus V_{GS} .

9.2 Linearity Characteristics

From analysis in Section 8.3, the IP3 sweet spot is not only determined by K_{3g_m} , but also the 2nd and 3rd order cross terms through Δ_1 and Δ_2 in (8.5). Here, the accuracy of (8.5) is examined against measured IP3 for different biasing voltages, different device sizes, and different fundamental frequencies to further verify the impact of the additional terms. Overall, the analytical expression is not bad for this 90nm RF CMOS technology.

9.2.1 Drain voltage dependence

Fig. 9.2 shows IIP3 from measurement and (8.5) for multiple V_{DS} . $N_f = 64$. With V_{DS} increasing from 0.6 to 1.0V, I_{DS} increases only slightly, IIP3 increases by a much larger factor, particularly at higher V_{GS} , and the IP3 sweet spot V_{GS} decreases.

To explain the V_{DS} dependence of IP3 sweet spot and high V_{GS} IP3, K_{3g_m}/g_m and $K_{3g_m}/g_m + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$ are plotted in Fig. 9.3. The zero K_{3g_m} point shifts to lower V_{GS} as V_{DS} increases because of the threshold voltage change due to DIBL. The V_{GS} gaps between IP3 sweet spot and zero K_{3g_m} point for different V_{DS} are approximately the same. Thus, the V_{DS} dependence of IP3 sweet spot is determined by the threshold voltage change due to DIBL, and the impacts of the Δ terms on IP3 sweet spot are similar for different V_{DS} . This was verified using simulation results with and without DIBL induced V_{th} change in Section 8.5. At high V_{GS} , K_{3g_m}/g_m are close while $K_{3g_m}/g_m + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$ split for different V_{DS} . Both K_{3g_m} and g_m do not show great

V_{DS} dependence at high V_{GS} because the transistor is biased in saturation region for all three V_{DS} . However, the V_{DS} dependence of the 2nd and 3rd order output conductance nonlinearities and cross terms is noticeable. This directly leads to highly V_{DS} dependent Δ terms at high V_{GS} . Therefore, the denominator of (8.5) and thus the calculated IIP3 are V_{DS} dependent at high V_{GS} .

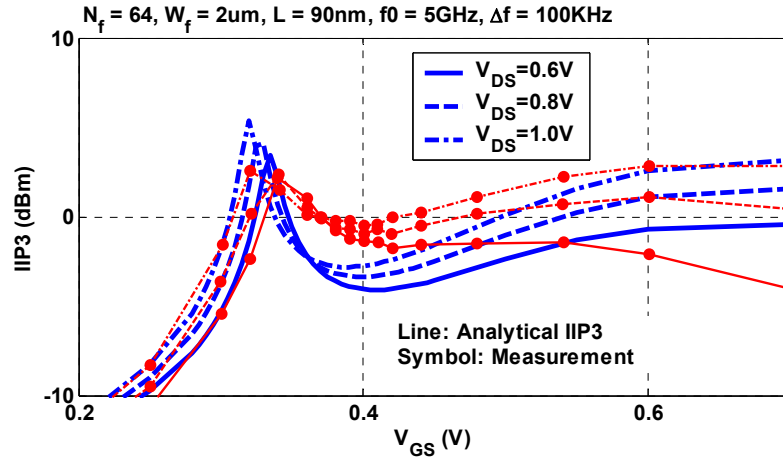


Fig. 9.2 Measured and analytical IIP3 versus V_{GS} at multiple V_{DS} . Analytical IIP3 is calculated using (8.5).

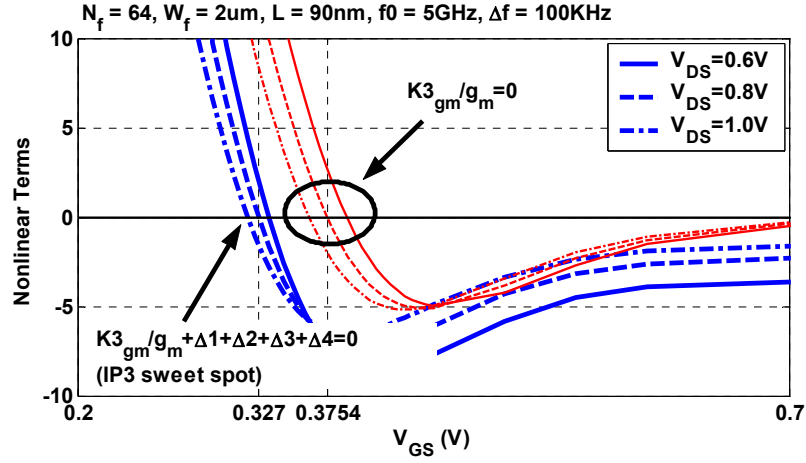


Fig. 9.3 K_{3g_m}/g_m and $K_{3g_m}/g_m + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$ versus V_{GS} at multiple V_{DS} .

9.2.2 Finger number dependence

To verify the analysis of device scaling in Section 8.4, Fig. 9.4 shows IIP3 from measurement and (8.5) versus J_{DS} for $N_f=10, 20$, and 64. The peak of measured IIP3 is not perfect because IP3 is not measured in a very fine biasing step. (8.5) can predict IP3 sweet spot accurately for devices with $N_f=10, 20$, and 64. Note that the sweet spot J_{DS} decreases from $35\mu\text{A}/\mu\text{m}$ for $N_f=10$ to $20\mu\text{A}/\mu\text{m}$ for $N_f=64$. K_{3g_m}/g_m for $N_f=10, 20$, and 64 are so close that only the zero K_{3g_m} point for $N_f=64$ is shown in Fig. 9.4 for reference.

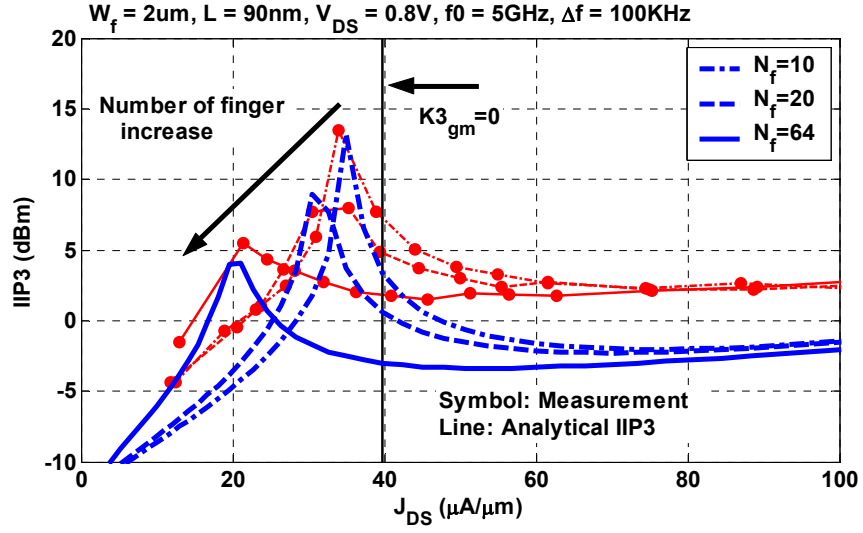


Fig. 9.4 Measured and analytical IIP3 versus J_{DS} for devices with $N_f=10, 20$, and 64 . Analytical IIP3 is calculated using (8.5).

9.2.3 Frequency dependence

Fig. 9.5 shows measured IIP3 at 2, 5, and 10GHz for devices with $N_f=10$ and 64 (a total width of $20\mu m$ and $128\mu m$). For $N_f=10$, IIP3 at 2, 5, and 10GHz are practically identical. For $N_f=64$, IIP3 increases as frequency increases. This frequency dependence can only be attributed to C_{gs} and C_d in (8.5). However, IIP3 calculated using (8.5) does not show a strong frequency dependence for $N_f=64$ as shown in Fig. 9.6 (a). To further explore this, C_{gd} is added to the small signal equivalent circuit. Fig. 9.6 (b) shows IIP3 calculated using Volterra series with C_{gd} added at multiple frequencies. The strong frequency dependence of calculated IIP3 with C_{gd} is similar to the frequency dependence of measured IIP3.

For small devices, C_{gs} , C_d , and C_{gd} are small, and thus the terms containing these capacitances are relatively small. For frequencies below 10GHz, the IIP3 calculated are practically the same for different frequencies. Fig. 9.7 shows IIP3 calculated using (8.5) and Volterra series with C_{gd} for $N_f=10$ at 2, 5, and 10GHz. The results are overlapped.

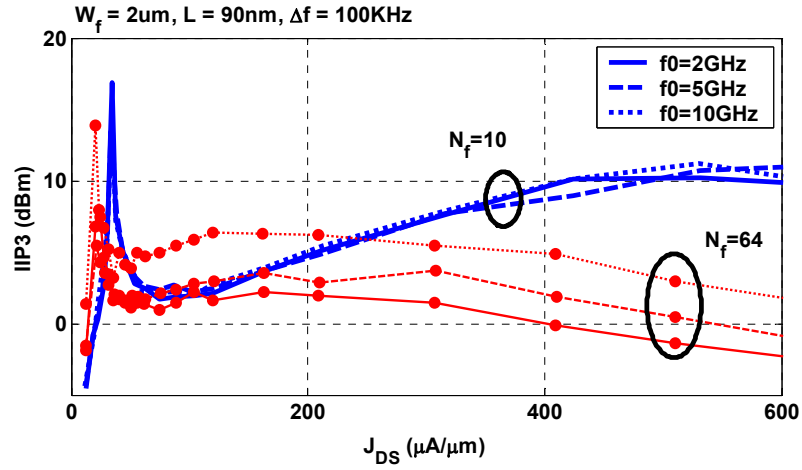


Fig. 9.5 Measured IIP3 versus J_{DS} at multiple frequencies for $N_f=10$ and 64 ($W=20\mu\text{m}$ and $128\mu\text{m}$).

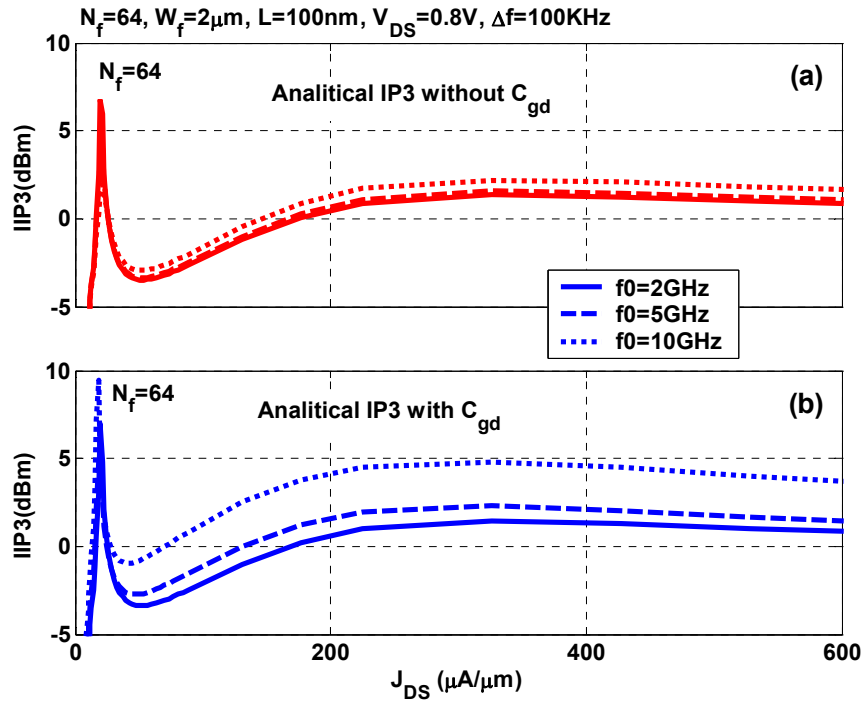


Fig. 9.6 Analytical IIP3 (a) without C_{gd} and (b) with C_{gd} at multiple frequencies for $N_f=64$ ($W=128\mu m$). Analytical IIP3 without C_{gd} is calculated using (8.5).

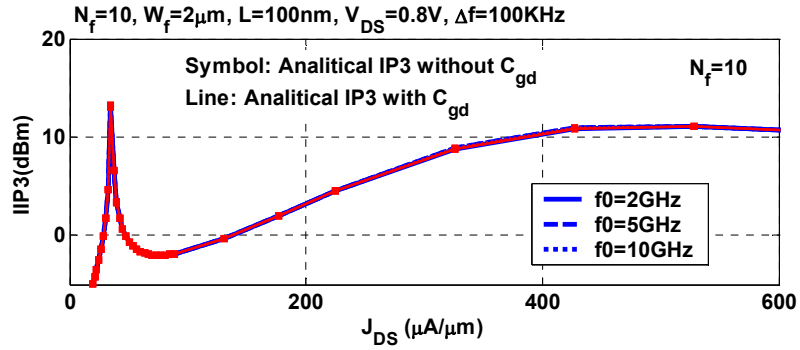


Fig. 9.7 Analytical IIP3 with and without C_{gd} at multiple frequencies for $N_f=10$ ($W=20\mu m$). Analytical IIP3 without C_{gd} is calculated using (8.5).

9.2.4 Large signal linearity

The solid lines in Fig. 9.8 represents the typical fundamental frequency output power, $P_{out,1st}$, and the third order intermodulation (IM3) output power, $P_{out,3rd}$, versus input power, P_{in} , curves for a moderate inversion gate bias. By observing the two curves, an unexpected minimum IM3 output power point at certain input power level is investigated. A better output signal power to distortion ratio can be achieved by selecting this P_{in} level as the circuits working point [23]. This large-signal IM3 sweet spot is not defined for small-signal operation, and cannot be evaluated using the extrapolated IP3 point. Note that instead of a gain compression at certain input power as shown in Fig. 1.7, the output signal first linearly follows the input power, then it experiences a faster rate of rise before it saturates. This phenomena is named as gain expansion, and can be observed sometimes [23].

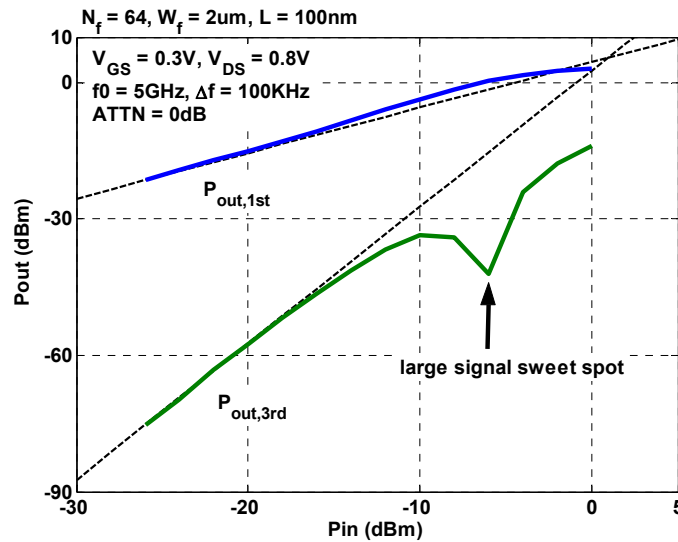


Fig. 9.8 The output power amplitude for fundamental and 3rd order intermodulation products versus input power.

Fig. 9.9 shows a contour plot for sweeping gate bias and input power. The deep valley marked using square symbols are the IM3 sweet spots [90]. Below -10dBm input power, the sweet spots appear at around gate bias 0.33V, which is the IP3 sweet spot. As the input power increases, the IM3 sweet spot shifts to lower gate bias voltage obviously.

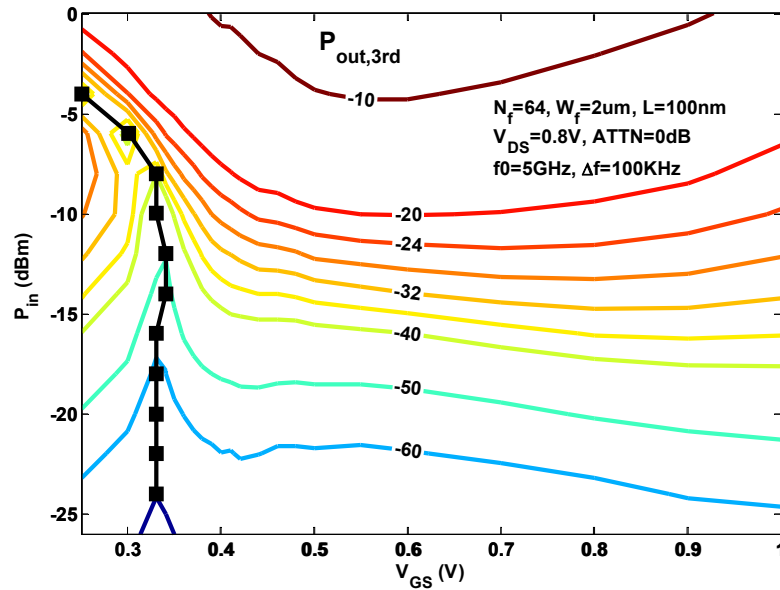


Fig. 9.9 Contour of 3rd order intermodulation output power with sweeping gate bias and input power.

9.3 Summary

In this chapter, the measured IIP3 is compared with calculated IIP3 using I-V and S-parameters from BSIM4 based simulation. The complete IP3 expression can correctly model the biasing, frequency, and device size dependence of IIP3 even with simulated I-V and S-parameters as long as the model is valid in DC I-V and S-parameters. In the

90nm CMOS technology used, the sweet spot J_{DS} decreases from 40 to 20 $\mu\text{A}/\mu\text{m}$ as gate width increases from 2 to 128 μm . The V_{DS} dependence and its device width dependence are also investigated using experimental results. These results provide useful guidelines to linearity characterization, simulation as well as optimal biasing and sizing for high linearity in RFIC design.

BIBLIOGRAPHY

- [1] W. Liu, X. Jin, X. Xi, et al., "BSIM3V3.3 MOSFET Model - User's Manual," University of California, Berkeley, <http://www-device.eecs.berkeley.edu/~bsim3/get.html>.
- [2] M. V. Dunga, W. Yang, X. Xi, et al., "BSIM4.6.1 MOSFET Model - User's Manual," University of California, Berkeley, http://www-device.eecs.berkeley.edu/~bsim3/bsim4_get.html.
- [3] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "MOS model 11 - Level 1102," Philips Electronics N.V. 2003/2004, http://www.nxp.com/acrobat_download/other/models/nltn2004_00085.pdf.
- [4] X. Li, W. Wu, G. Gildenblat, et al., "PSP 102.3," NXP Semiconductors, http://www.nxp.com/acrobat_download/other/models/psp102p3_summary.pdf.
- [5] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "MOS Model 11," in *Compact Model Council Meeting*, 2000.
- [6] C. Galup-Montoro and M. C. Schneider, *MOSFET modeling for circuit analysis and design*. Singapore: World Scientific Publishing Company, 2007.
- [7] M. Golio and J. Couie, "Who Pays for Characterization?: The Final Dilemma for MESFET Modeling," in *Dig. 48th ARFTG Conference-Fall*, pp. 87-93, 1996.
- [8] F. Sischka, "Applying Nonlinear RF Device Modeling To Verify S-Parameter Linearity," in *European Microwave Week*, 2001.
- [9] J. Civello, "Adressing the challenges of RF device modeling for successful high-frequency design," *Microwave Engineering Europe*, pp. 21-28, 2004.
- [10] R. van Langevelde, L. F. Tiemeijer, R. J. Havens, et al., "RF-distortion in deep-submicron CMOS technologies," in *Tech. Dig. of Int. Electron Devices Meeting (IEDM)*, 2000, pp. 807-810.
- [11] P. H. Woerlee, M. J. Knitel, R. v. Langevelde, et al., "RF-CMOS performance trends," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776-1782, 2001.
- [12] M. Hiebel, *Fundamentals of vector network analysis*. Germany: Rohde & Schwarz, 2007.

- [13] S. A. Wartenberg, *RF measurements of Die and Packages*. Boston, USA: Artech House, 2002.
- [14] M. C. A. M. Koolen, J. A. M. Geelen, and M. P. J. G. Versleijen, "An improved de-embedding technique for on-wafer high frequency characterization," in *Proc. of IEEE Bipolar Circuits and Technology Meeting*, pp. 188-191, 1991.
- [15] H. Cho and D. E. Burk, "A three-step method for the de-embedding of high-frequency S-parameter measurements," *IEEE Trans. Electron Devices*, vol. 38, no. 6, pp. 1371-1375, 1991.
- [16] E. P. Vandamme, D. M. M. P. Schreurs, and G. Van Dinther, "Improved three-step de-embedding method to accurately account for the influence of pad parasitics in silicon on-wafer RF test-structures," *IEEE Trans. Electron Devices*, vol. 48, no. 4, pp. 737-742, 2001.
- [17] L. F. Tiemeijer, R. J. Havens, A. B. M. Jansman, et al., "Comparison of the "pad-open-short" and "open-short-load" deembedding techniques for accurate on-wafer RF characterization of high-quality passives," *IEEE Trans. Microwave Theory and Techniques*, vol. 53, no. 2, pp. 723-729, 2005.
- [18] X. Jin, J.-J. Ou, C.-H. Chen, et al., "An effective gate resistance model for CMOS RF and noise modeling," in *Dig. Int. Electron Device Meeting*, pp. 961-964, 1998.
- [19] X. Wei, K. Xia, G. Niu, et al., "An Improved On-chip 4-Port Parasitics De-embedding Method with Application to RF CMOS," in *Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 24-27, 2007.
- [20] Q. Liang, J. D. Cressler, G. Niu, et al., "A simple four-port parasitic deembedding methodology for high-frequency scattering parameter and noise characterization of SiGe HBTs," *IEEE Trans. Microwave Theory and Techniques*, vol. 51, no. 11, pp. 2165-2174, 2003.
- [21] S. A. Wartenberg, "Selected topics in RF coplanar probing," *IEEE Trans. Microwave Theory and Techniques*, vol. 51, no. 4, pp. 1413-1421, 2003.
- [22] G. Niu, J. Pan, X. Wei, et al., "Intermodulation linearity characteristics of CMOS transistors in a 0.13 μ m process," in *Dig. of IEEE Radio Frequency Integrated Circuits Symp.*, pp. 65-68, 2005.
- [23] N. B. De Carvalho and J. C. Pedro, "Large- and small-signal IMD behavior of microwave power amplifiers," *IEEE Trans. Microwave Theory and Techniques*, vol. 47, no. 12, pp. 2364-2374, 1999.
- [24] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*: Kluwer Academic Publisher, 2001.

- [25] S. Kang, B. Choi, and B. Kim, "Linearity analysis of CMOS for RF application," *IEEE Trans. Microwave Theory and Tech.*, vol. 51, no. 3, pp. 972-977, 2003.
- [26] B. Toole, C. Plett, and M. Cloutier, "RF circuit implications of moderate inversion enhanced linear region in MOSFETs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 2, pp. 319-328, 2004.
- [27] L. Tzung-Yin and C. Yuhua, "High-frequency characterization and modeling of distortion behavior of MOSFETs for RF IC design," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1407-1414, 2004.
- [28] C. H. Doan, S. Emami, D. A. Sobel, et al., "Design considerations for 60 GHz CMOS radios," *IEEE Communications Magazine*, vol. 42, no. 12, pp. 132-140, 2004.
- [29] T. Yao, T. Yao, M. Gordon, et al., "60-GHz PA and LNA in 90-nm RF-CMOS," in *Dig. of IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, 2006, pp. 4 pages.
- [30] T. LaRocca and M. C. F. Chang, "60GHz CMOS differential and transformer-coupled power amplifier for compact design," in *Dig. IEEE Radio Frequency Integrated Circuits Symp.*, pp. 65-68, 2008.
- [31] S. T. Nicolson, A. Tomkins, K. W. Tang, et al., "A 1.2V, 140GHz receiver with on-die antenna in 65nm CMOS," in *Dig. IEEE Radio Frequency Integrated Circuits Symp.*, pp. 229-232, 2008.
- [32] B. Oldfield, "Wafer Probe Calibration Accuracy Measurements," in *Dig. 48th ARFTG Conference-Fall*, pp. 54-58, 1996.
- [33] T. E. Kolding, "A four-step method for de-embedding gigahertz on-wafer CMOS measurements," *IEEE Trans. Electron Devices*, vol. 47, no. 4, pp. 734-740, 2000.
- [34] A. Rumiantsev and N. Ridler, "VNA calibration," *IEEE Microwave Magazine*, vol. 9, no. 3, pp. 86-99, 2008.
- [35] R. A. Speciale, "A Generalization of the TSD Network-Analyzer Calibration Procedure, Covering n-Port Scattering-Parameter Measurements, Affected by Leakage Errors," *IEEE Trans. Microwave Theory and Techniques*, vol. 25, no. 12, pp. 1100-1115, 1977.
- [36] J. V. Butler, D. K. Rytting, M. F. Iskander, et al., "16-term error model and calibration procedure for on-wafer network analysis measurements," *IEEE Trans. Microwave Theory and Techniques*, vol. 39, no. 12, pp. 2211-2217, 1991.

- [37] R. A. Speciale and N. R. Franzen, "Super - TSD, A Generalization of the TSD Network Analyzer Calibration Procedure, Covering n - Port Measurements with Leakage," in *Dig. IEEE MTT-S Int. Microwave Symp.*, pp. 114-117, 1977.
- [38] A. Gronefeld and B. Schiek, "Network-analyzer self-calibration with four or five standards for the 15-term error-model," in *Dig. IEEE MTT-S Int. Microwave Symposium*, vol. 3, 1997, pp. 1655-1658 vol.3.
- [39] K. Silvonen, "LMR 16-a self-calibration procedure for a leaky network analyzer," *IEEE Trans. Microwave Theory and Techniques*, vol. 45, no. 7, pp. 1041-1049, 1997.
- [40] R. A. Pucel, W. Struble, R. Hallgren, et al., "A general noise de-embedding procedure for packaged two-port linear active devices," *IEEE Trans. Microwave Theory and Techniques*, vol. 40, no. 11, pp. 2013-2024, 1992.
- [41] K. J. Silvonen, "Calibration of 16-term error model," *Electronics Lett.*, vol. 29, no. 17, pp. 1544-1545, 1993.
- [42] H. Heuermann and B. Schiek, "Results of network analyzer measurements with leakage errors-corrected with direct calibration techniques," *IEEE Trans. Instrumentation and Measurement*, vol. 46, no. 5, pp. 1120-1127, 1997.
- [43] H. Heuermann and B. Schiek, "15-term self-calibration methods for the error-correction of on-wafer measurements," *IEEE Trans. Instrumentation and Measurement*, vol. 46, no. 5, pp. 1105-1110, 1997.
- [44] Q. Liang, W. Kuo, J. D. Cressler, et al., "Accurate AC transistor characterization to 110 GHz using a new four-port self-calibrated extraction technique," in *Dig. 2004 IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, vol. 51, 2004, pp. 282-285.
- [45] X. Wei, G. Niu, S. Sweeney, et al., "A general 4-port solution for 110 GHz on-wafer transistor measurements with or without impedance standard substrate (ISS) calibration," *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2706-2714, 2007.
- [46] T. Jamneala, D. A. Feld, D. Blackham, et al., "Why Reciprocal Procedure Works?," in *Dig. IEEE Radio Frequency Integrated Circuits Symposium*, 2006, pp. 465-469.
- [47] X. Wei, G. Niu, S. L. Sweeney, et al., "Singular-Value-Decomposition Based Four Port De-embedding and Single-step Error Calibration for On-chip Measurement," in *IEEE/MTT-S Int. Microwave Symposium*, pp. 1497-1500, 2007.

- [48] V. Aparin, G. Brown, and L. E. Larson, "Linearization of CMOS LNA's via optimum gate biasing," in *Proc. of Int. Symp. on Circuits and Systems*, pp. IV-748-51, 2004.
- [49] D. Linten, L. Aspemyr, W. Jeamsaksiri, et al., "Low-power 5 GHz LNA and VCO in 90 nm RF CMOS," in *Dig. of IEEE Symp. on VLSI Circuits*, pp. 372-375, 2004.
- [50] R. Sung, P. Bendix, and M. B. Das, "Extraction of high-frequency equivalent circuit parameters of submicron gate-length MOSFET's," *IEEE Trans. Electron Devices*, vol. 45, no. 8, pp. 1769-1775, 1998.
- [51] C. Fager, J. C. Pedro, N. B. de Carvalho, et al., "A comprehensive analysis of IMD behavior in RF CMOS power amplifiers," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 24-34, 2004.
- [52] V. Aparin and L. E. Larson, "Modified derivative superposition method for linearizing FET low-noise amplifiers," *IEEE Trans. Microwave Theory and Tech.*, vol. 53, no. 2, pp. 571-581, 2005.
- [53] X. Wei, G. Niu, Y. Li, et al., "Experimental characterization and simulation of RF intermodulation linearity in a 90 nm RF CMOS technology," in *Dig. of IEEE Radio Frequency Integrated Circuits Symp.*, pp. 251-254, 2008.
- [54] X. Wei, G. Niu, Y. Li, et al., "Modeling and characterization of intermodulation linearity on a 90nm RF CMOS technology," *IEEE Trans. Microwave Theory and Techniques*, submitted.
- [55] Cascade Microtech, "Mechanical Layout Rules for Infinity Probes," Application Note, 2004.
- [56] C. S. Kim, J.-W. Park, H. K. Yu, et al., "Gate layout and bonding pad structure of a RF n-MOSFET for low noise performance," *IEEE Electron Device Lett.*, vol. 21, no. 12, pp. 607-609, 2000.
- [57] T. E. Kolding, O. K. Jensen, and T. Larsen, "Ground-shielded measuring technique for accurate on-wafer characterization of RF CMOS devices," in *Proc. of IEEE Int. Conf. on Microelectronic Test Structures*, pp. 246-251, 2000.
- [58] T. E. Kolding, "Shield-based microwave on-wafer device measurements," *IEEE Trans. Microwave Theory and Techniques*, vol. 49, no. 6, pp. 1039-1044, 2001.
- [59] H. Tsuno, K. Anzai, M. Matsumura, et al., "Advanced Analysis and Modeling of MOSFET Characteristic Fluctuation Caused by Layout Variation," in *IEEE Symp. on VLSI Technology*, pp. 204-205, 2007.

- [60] T. Tanaka, A. Satoh, M. Yamaji, et al., "Fully considered layout variation analysis and compact modeling of MOSFETs and its application to circuit simulation," in *IEEE Int. Conf. on Microelectronic Test Structures*, pp. 223-227, 2008.
- [61] T. E. Kolding, "Test structure for universal estimation of MOSFET substrate effects at gigahertz frequencies," in *Proc. of IEEE Int. Conf. on Microelectronic Test Structures*, pp. 106-111, 2000.
- [62] A. Cathignol, S. Mennillo, S. Bordez, et al., "Spacing impact on MOSFET mismatch," in *IEEE Int. Conf. on Microelectronic Test Structures*, pp. 90-95, 2008.
- [63] J. T. Colvin, S. S. Bhatia, and K. K. O, "A bond-pad structure for reducing effects of substrate resistance on LNA performance in a silicon bipolar technology," in *Proc. of IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 109-112, 1998.
- [64] J. d. Cressler and G. Niu, *Silicon-germanium heterojunction bipolar transistors*. Boston, MA/USA: Artech House, INC., 2002.
- [65] K. Ickjin, J. Minkyu, L. Kwyro, et al., "A simple and analytical parameter-extraction method of a microwave MOSFET," *IEEE Trans. Microwave Theory and Techniques*, vol. 50, no. 6, pp. 1503-1509, 2002.
- [66] J. Tabuchi, B. Hughes, and J. Perdomo, "On-Wafer Millimeter-Wave Network Analysis for Device and Circuit Design," in *Dig. 38th ARFTG Conference-Winter*, pp. 53-61, 1991.
- [67] S. Rehnmark, "On the Calibration Process of Automatic Network Analyzer Systems," *IEEE Trans. Microwave Theory and Techniques*, vol. 22, no. 4, pp. 457-458, 1974.
- [68] D. K. Rytting, "Network Analyzer Error Models and Calibration Methods," RF & Microwave Measurements for Wireless Applications, ARFTG/NIST Short Course Notes, 1996.
- [69] R. A. Hackborn, "An automatic network analyzer system," *Microwave Journal*, vol. 11, pp. 45-52, 1968.
- [70] W. Kruppa and K. F. Sodomsy, "An Explicit Solution for the Scattering Parameters of a Linear Two-Port Measured with an Imperfect Test Set (Correspondence)," *IEEE Trans. Microwave Theory and Techniques*, vol. 19, no. 1, pp. 122-123, 1971.

- [71] N. R. Franzen and R. A. Speciale, "A New Procedure for System Calibration and Error Removal in Automated S-Parameter Measurements," in *Proc. of 5th European Microwave Conf.*, pp. 69-73, 1975.
- [72] H. V. Shurmer, "Calibration procedure for computer-corrected s parameter characterisation of devices mounted in microstrip," *Electronics Letters*, vol. 9, no. 14, pp. 323-324, 1973.
- [73] D. F. Williams, R. B. Marks, and A. Davidson, "Comparison of On-Wafer Calibrations," in *Dig. 38th ARFTG Conference-Winter*, pp. 68-81, 1991.
- [74] R. B. Marks, J. A. Jargon, and J. R. Juroshek, "Calibration Comparison Method for Vector Network Analyzers," in *Dig. 48th ARFTG Conference-Fall*, pp. 38-45, 1996.
- [75] R. B. Marks, "Formulations of the basic vector network analyzer error model including switch-terms," in *ARFTG Conference Digest-Fall, 50th*, pp. 115-126, 1997.
- [76] S. Vandenberghe, D. Schreurs, G. Carchon, et al., "Identifying error-box parameters from the twelve-term vector network analyzer error model," in *Dig. 60th ARFTG Microwave Measurements Conf.*, pp. 157-165, 2002.
- [77] K. H. Wong, "Characterization of Calibration Standards by Physical Measurements," in *ARFTG Conference Digest-Spring, 39th*, pp. 53-62, 1992.
- [78] Cascade Microtech, "Impedance standard substrates to support all of your high-frequency probing applications," Data Sheet, 2005.
- [79] H. J. Eul and B. Schiek, "A generalized theory and new calibration procedures for network analyzer self-calibration," *IEEE Trans. Microwave Theory and Techniques*, vol. 39, no. 4, pp. 724-731, 1991.
- [80] D. K. Rytting, "An Analysis of Vector Measurement Accuracy Enhancement Techniques," Hewlett-Packard RF & Microwave Measurement Symposium and Exhibition, 1987.
- [81] H. Heuermann and B. Schiek, "Robust algorithms for Txx network analyzer self-calibration procedures," *IEEE Trans. Instrumentation and Measurement*, vol. 43, no. 1, pp. 18-23, 1994.
- [82] H. Heuermann, "Calibration procedures with series impedances and unknown lines simplify on-wafer measurements," *IEEE Trans. Microwave Theory and Techniques*, vol. 47, no. 1, pp. 1-5, 1999.

- [83] W. R. Deal and D. S. Farkas, "A simple CAD-based method to develop and evaluate calibration standards," *IEEE Microwave Magazine*, vol. 7, no. 3, pp. 70-84, 2006.
- [84] S. Padmanabhan, P. Kirby, J. Daniel, et al., "Accurate broadband on-wafer SOLT calibrations with complex load and thru models," in *Dig. 61st ARFTG Conference-Spring*, pp. 5-10, 2003.
- [85] D. Blackham and K. Wong, "Latest advances in VNA accuracy enhancements," *Microwave Journal*, vol. 48, no. 7, pp. 78-94, 2005.
- [86] F. Williams, "Characterization of Thin-Film Calibration Elements," in *Dig. 38th ARFTG Conference-Winter*, pp. 25-35, 1991.
- [87] D. K. Rytting, "Appendix to An Analysis of Vector Measurement Accuracy Enhancement Techniques ": Hewlett-Packard RF & Microwave Measurement Symposium and Exhibition, 1987.
- [88] D. A. Frickey, "Conversions between S, Z, Y, H, ABCD, and T parameters which are valid for complex source and load impedances," *IEEE Trans. Microwave Theory and Techniques*, vol. 42, no. 2, pp. 205-211, 1994.
- [89] G. A. Forsythe, *Computer Methods for Mathematical Computations*: Englewood Cliffs, NJ: Prentice Hall, 1977.
- [90] L. Seung-Yup, J. Kye-Ik, L. Yong-Sub, et al., "The IMD sweet spots varied with gate bias voltages and input powers in RF LDMOS power amplifiers," in *Proc. 33rd European Microwave Conf.*, pp. 1353-1356 Vol.3, 2003.

APPENDIX A

ABBREVIATIONS AND SYMBOLS

A.1 Abbreviations

DUT: device under test. (Everything probed by the probes is a DUT.)

GSG: ground-signal-ground.

IM3: third order intermodulation.

IP3: third order intercept point. (IIP3: input IP3) and (OIP3: output IP3)

ISS: impedance standard substrate.

PSA: performance spectrum analyzer.

PSG: performance signal generator.

SOLT: short-open-load-thru.

SVD: singular-value-decomposition.

TRL: thru-reflection-line

VNA: vector network analyzer.

A.2 Matrix symbols and matrix index

$Y^{OS, left}$: Y-parameters of LEFT after open-short de-embedding.

$Y^{OS, right}$: Y-parameters of RIGHT after open-short de-embedding.

S^M : Measured S-parameter without switch errors.

S^{DUT} : S-parameter of the whole on-wafer test structure being probed.

S^A : The actual S-parameter of the desired two-port.

Comments:

1) If Q is a two-dimension matrix, Q_{mn} or q_{mn} is the (m,n) element in the matrix.

The subscript is the row and column index of the element.

2) If Q is the name of a test structure or a multi-port network, then $S^Q(E^Q)$, Y^Q ,

Z^Q , T^Q , and A^Q are the S-, Y-, Z-, T-, and ABCD- parameters of the structure

or the network. The transformation between them is listed in Appendix B.

APPENDIX B

TWO PORT NETWORK REPRESENTATIONS

Two port network can be represented using S-, H-, Y-, Z-, and ABCD-parameters. The transformation between these representations is important for system error calibration, on-wafer parasitics de-embedding and model parameter extraction. Table B.1 gives the transformation between H, Y, Z, and ABCD parameters. The transformation from S to Y and Z are given as

$$Y = Y_0 (I - S)(I + S)^{-1} \leftrightarrow S = (Y + Y_0)^{-1} (Y_0 - Y), \quad (\text{B.1})$$

$$Z = Z_0 (I + S)(I - S)^{-1} \leftrightarrow S = (Z_0 + Z)^{-1} (Z - Z_0), \quad (\text{B.2})$$

where Z_0 is system characteristic impedance. $Z_0 = 50\Omega$. $Y_0 = Z_0^{-1}$. The 2×2 matrices for the transformation from S to Y and Z are listed below:

$$\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} Y_0 \frac{1 - S_{11} + S_{22} - \Delta_S}{1 + S_{11} + S_{22} + \Delta_S} & Y_0 \frac{-2S_{12}}{1 + S_{11} + S_{22} + \Delta_S} \\ Y_0 \frac{-2S_{21}}{1 + S_{11} + S_{22} + \Delta_S} & Y_0 \frac{1 + S_{11} - S_{22} - \Delta_S}{1 + S_{11} + S_{22} + \Delta_S} \end{bmatrix}, \quad (\text{B.3})$$

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = \begin{bmatrix} Z_0 \frac{1 + S_{11} - S_{22} - \Delta_S}{1 - S_{11} - S_{22} + \Delta_S} & Z_0 \frac{2S_{21}}{1 - S_{11} - S_{22} + \Delta_S} \\ Z_0 \frac{2S_{12}}{1 - S_{11} - S_{22} + \Delta_S} & Z_0 \frac{1 - S_{11} + S_{22} - \Delta_S}{1 - S_{11} - S_{22} + \Delta_S} \end{bmatrix}, \quad (\text{B.4})$$

where $\Delta_S = S_{11}S_{22} - S_{12}S_{21}$.

Table B.1 Transformation between two port H, Y, Z, and ABCD representations

	H	Y	Z	ABCD
H	$\begin{array}{cc} H_{11} & H_{12} \\ H_{21} & H_{22} \end{array}$	$\begin{array}{cc} \frac{1}{Y_{11}} & \frac{-Y_{12}}{Y_{11}} \\ \frac{Y_{21}}{Y_{11}} & \frac{\Delta_Y}{Y_{11}} \end{array}$	$\begin{array}{cc} \frac{\Delta_Z}{Z_{22}} & \frac{Z_{12}}{Z_{22}} \\ \frac{-Z_{21}}{Z_{22}} & \frac{1}{Z_{22}} \end{array}$	$\begin{array}{cc} \frac{B}{D} & \frac{\Delta_{ABCD}}{D} \\ \frac{-1}{D} & \frac{C}{D} \end{array}$
Y	$\begin{array}{cc} \frac{1}{H_{11}} & \frac{-H_{12}}{H_{11}} \\ \frac{H_{21}}{H_{11}} & \frac{\Delta_H}{H_{11}} \end{array}$	$\begin{array}{cc} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{array}$	$\begin{array}{cc} \frac{Z_{22}}{\Delta_Z} & \frac{-Z_{12}}{\Delta_Z} \\ \frac{-Z_{21}}{\Delta_Z} & \frac{Z_{11}}{\Delta_Z} \end{array}$	$\begin{array}{cc} \frac{D}{B} & \frac{-\Delta_{ABCD}}{B} \\ \frac{-1}{B} & \frac{A}{B} \end{array}$
Z	$\begin{array}{cc} \frac{\Delta_H}{H_{22}} & \frac{H_{12}}{H_{22}} \\ \frac{-H_{21}}{H_{22}} & \frac{1}{H_{22}} \end{array}$	$\begin{array}{cc} \frac{Y_{22}}{\Delta_Y} & \frac{-Y_{12}}{\Delta_Y} \\ \frac{-Y_{21}}{\Delta_Y} & \frac{Y_{11}}{\Delta_Y} \end{array}$	$\begin{array}{cc} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{array}$	$\begin{array}{cc} \frac{A}{C} & \frac{\Delta_{ABCD}}{C} \\ \frac{1}{C} & \frac{D}{C} \end{array}$
ABCD	$\begin{array}{cc} \frac{-\Delta_H}{H_{21}} & \frac{-H_{11}}{H_{21}} \\ \frac{-H_{22}}{H_{21}} & \frac{-1}{H_{21}} \end{array}$	$\begin{array}{cc} \frac{-Y_{22}}{Y_{21}} & \frac{-1}{Y_{21}} \\ \frac{-\Delta_Y}{Y_{21}} & \frac{-Y_{11}}{Y_{21}} \end{array}$	$\begin{array}{cc} \frac{Z_{11}}{Z_{21}} & \frac{\Delta_Z}{Z_{21}} \\ \frac{1}{Z_{21}} & \frac{Z_{22}}{Z_{21}} \end{array}$	$\begin{array}{cc} A & B \\ C & D \end{array}$
$\Delta_H = H_{11}H_{22} - H_{12}H_{21}, \Delta_Y = Y_{11}Y_{22} - Y_{12}Y_{21}, \Delta_Z = Z_{11}Z_{22} - Z_{12}Z_{21}, \Delta_{ABCD} = AD - BC.$				

APPENDIX C

REVIEW OF ON-WAFER DE-EMBEDDING METHODS

C.1 Open-Short de-embedding

Fig. C.1 shows the equivalent circuit for open-short de-embedding. Fig. C.2 shows the equivalent circuits and layouts of the OPEN and SHORT standards. Denoting

$$Y_E = \begin{bmatrix} Y_1 + Y_3 & -Y_3 \\ -Y_3 & Y_2 + Y_3 \end{bmatrix}, \quad (\text{C.1})$$

and

$$Z_S = \begin{bmatrix} Z_4 + Z_6 & Z_6 \\ Z_6 & Z_5 + Z_6 \end{bmatrix}, \quad (\text{C.2})$$

the measured Y-parameters of OPEN and SHORT are $Y^{M,open} = Y_E$ and

$Y^{M,short} = Y_E + Z_S^{-1}$. That leads to $Z_S = (Y^{M,short} - Y^{M,open})^{-1}$.

Using the properties of shunt and series connected two-port networks, the measured Y-parameter of any DUT, Y^M , in Fig. C.1 is

$$Y^M = Y_E + \left\{ Z_S + [Y^A]^{-1} \right\}^{-1}. \quad (\text{C.3})$$

Thus, the actual Y-parameters, Y^A , can be obtained as

$$Y^A = \left((Y^M - Y^{M,open})^{-1} - (Y^{M,short} - Y^{M,open})^{-1} \right)^{-1}. \quad (\text{C.4})$$

Open-short de-embedding is valid as long as the parallel parasitics is mainly located at the probing pads. It is still an industry standard de-embedding method, and can provide valuable device parameters below 30GHz.

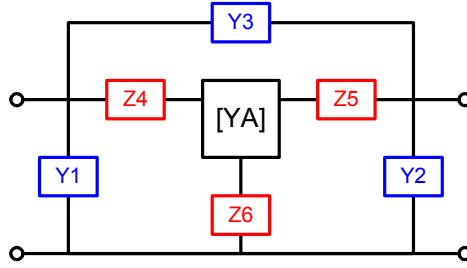


Fig. C.1 Equivalent circuit of on-wafer parasitics for open-short de-embedding.

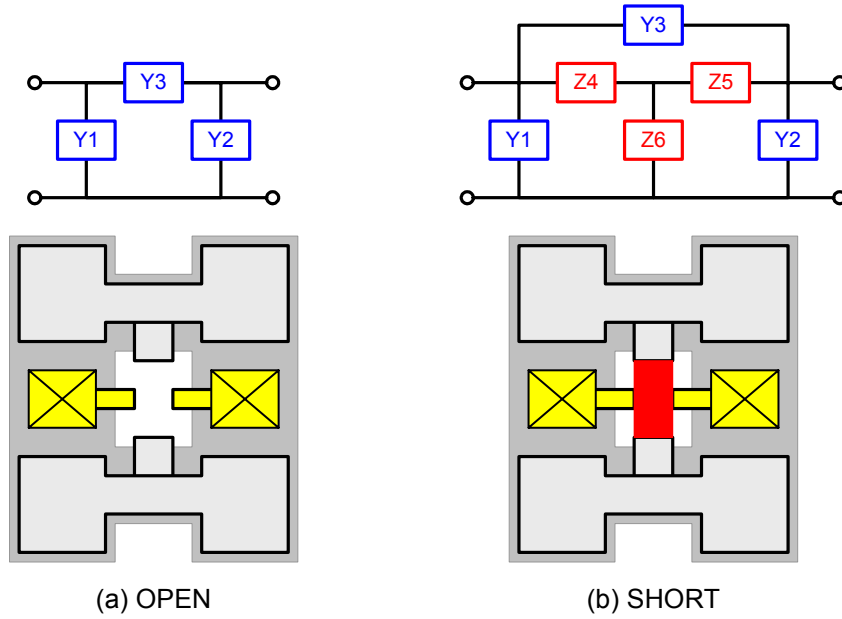


Fig. C.2 Equivalent circuits and layouts of (a) OPEN, and (b) SHORT standards.

C.2 Pad-open-Short de-embedding

Fig. C.3 shows the equivalent circuit for pad-open-short de-embedding. Fig. C.4 shows the equivalent circuits and layouts of the PAD, OPEN, and SHORT standards.

Denoting Y_E as (C.1), Z_S as (C.2), and Y_I as

$$Y_I = \begin{bmatrix} Y_7 + Y_9 & -Y_9 \\ -Y_9 & Y_8 + Y_9 \end{bmatrix}, \quad (\text{C.5})$$

Thus, $Y^{M,pad} = Y_E$, $Y^{M,short} = Y_E + Z_S^{-1}$, and $Y^{M,open} = Y_E + (Z_S + Y_I^{-1})^{-1}$. Z_S and Y_I can be solved as

$$Z_S = (Y^{M,short} - Y^{M,pad})^{-1}, \quad (\text{C.6})$$

$$Y_I = \left[(Y^{M,open} - Y^{M,pad})^{-1} - (Y^{M,short} - Y^{M,pad})^{-1} \right]^{-1} \approx Y^{M,open} - Y^{M,pad}. \quad (\text{C.7})$$

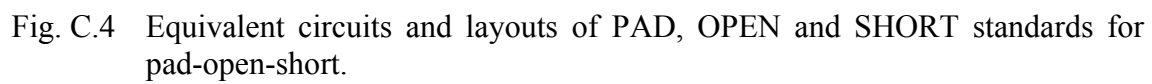
The equivalent circuit that shown in Fig. C.3 gives

$$Y^M = Y_E + \left[Z_S + (Y_I + Y^A)^{-1} \right]^{-1}, \quad (\text{C.8})$$

Y^A can then be obtained as

$$Y^A = \left((Y^M - Y_E)^{-1} - Z_S \right)^{-1} - Y_I. \quad (\text{C.9})$$

Pad-open-short de-embedding lumps the distributive parasitics along the connections at the pad and the end of connections, which can work up to 50GHz.



C.3 Three-step de-embedding

Fig. C.5 shows the equivalent circuit for three-step de-embedding. Fig. C.6 shows the equivalent circuits and layouts of the OPEN, SHORT1, SHORT2 and THRU standards. Denoting

$$Y_E = \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}, \quad (C.10)$$

$$Z_S = \begin{bmatrix} Z_4 + Z_6 & Z_6 \\ Z_6 & Z_5 + Z_6 \end{bmatrix}, \quad (C.11)$$

$$Y_I = \begin{bmatrix} Y_3 & -Y_3 \\ -Y_3 & Y_3 \end{bmatrix}, \quad (C.12)$$

Y^M and Y^A can be related through (C.8) and (C.9). The elements of Y_E , Z_S , and Y_I can be solved from on-wafer standards since

$$Y^{M,open} = Y_E + \begin{bmatrix} Y'_3 & -Y'_3 \\ -Y'_3 & Y'_3 \end{bmatrix}, \quad Y'_3 = (Y_3^{-1} + Z_4 + Z_5)^{-1} \quad (C.13)$$

$$Y^{M,short1} = Y_E + Z_{S1}^{-1}, \quad Z_{S1} = \begin{bmatrix} Z_4 + Z_6 & Z_6 \\ Z_6 & Z_5 + Y_3^{-1} + Z_6 \end{bmatrix}, \quad (C.14)$$

$$Y^{M,short2} = Y_E + Z_{S2}^{-1}, \quad Z_{S2} = \begin{bmatrix} Z_4 + Y_3^{-1} + Z_6 & Z_6 \\ Z_6 & Z_5 + Z_6 \end{bmatrix}, \quad (C.15)$$

$$Y^{M,thru} = \begin{bmatrix} Y_1 + Y_3'' & -Y_3'' \\ -Y_3'' & Y_2 + Y_3'' \end{bmatrix}, \quad Y_3'' = (Z_4 + Z_5)^{-1}. \quad (C.16)$$

The elements of Y_E , Z_S , and Y_I are

$$Y_1 = Y_{11}^{M,open} + Y_{12}^{M,open}, \quad (C.17)$$

$$Y_2 = Y_{22}^{M,open} + Y_{12}^{M,open}, \quad (C.18)$$

$$Z_4 = \frac{1}{2} \left[(Z_{S1,11} - Z_{S2,22}) - (Y_{12}^{M,thru})^{-1} \right], \quad (C.19)$$

$$Z_5 = \frac{1}{2} \left[(Z_{S2,22} - Z_{S1,11}) - (Y_{12}^{M,thru})^{-1} \right], \quad (C.20)$$

$$Z_6 = \frac{1}{2} \left[(Z_{S1,11} + Z_{S2,22}) + (Y_{12}^{M,thru})^{-1} \right], \quad (C.21)$$

$$Y_3 = \left[- (Y_{12}^{M,open})^{-1} + (Y_{12}^{M,thru})^{-1} \right]^{-1}. \quad (C.22)$$

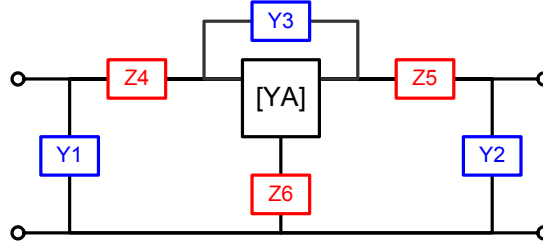


Fig. C.5 Equivalent circuit for improved three step de-embedding.

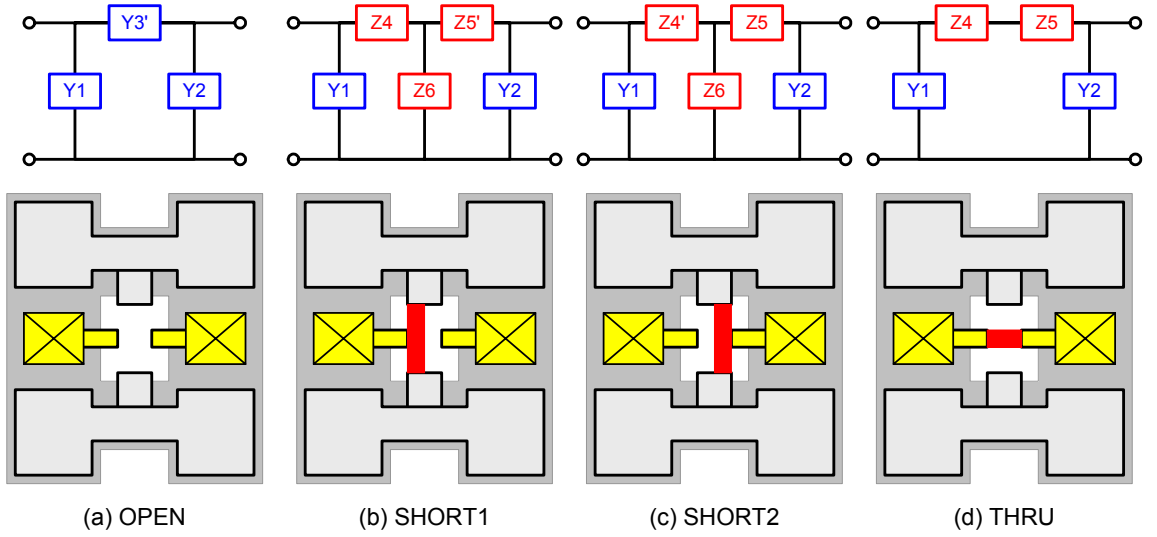


Fig. C.6 Equivalent circuits and layouts of OPEN, SHORT1, SHORT2, and THRU standards for improved three step.

C.4 Transmission line de-embedding

The on-wafer parasitics and the desired device can be represented as a cascade of several two port networks as shown in Fig. C.7. The input and output networks, which are composed of the probe pads and the interconnections leading to the device, are represented using ABCD parameters, A^{IN} and A^{OUT} . $A^{IN} = A^{PAD1} A^X$ and $A^{OUT} = A^Y A^{PAD2}$, where, A^{PAD1} and A^{PAD2} are the ABCD parameters of the probe pads at input and output, A^X and A^Y are ABCD parameters of input and output interconnections. Fig. C.8 shows the equivalent circuits and layouts for the THRU1 and THRU2 standards. The two transmission line structures have different length ℓ_s and ℓ_L .

The measured ABCD parameters of the desired device is

$$A^M = A^{IN} A^A A^{OUT}. \quad (C.23)$$

Thus, the measured ABCD parameters of THRU1 and THRU2 are

$$A^{M,thru1} = A^{IN} A^{A,thru1} A^{OUT} = A^{PAD1} A^{A,\ell_s} A^{PAD2}, \quad (C.24)$$

$$A^{M,thru2} = A^{IN} A^{A,thru2} A^{OUT} = A^{PAD1} A^{A,\ell_L} A^{PAD2}. \quad (C.25)$$

The ABCD parameters of a transmission line with length $\ell_L - \ell_s$ can be calculated as

$A^{A,\ell_L-\ell_s} = A^{A,\ell_L} (A^{A,\ell_s})^{-1}$ because the ABCD parameters of a transmission line of length ℓ are given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cosh \gamma \ell & Z_C \sinh \gamma \ell \\ \frac{1}{Z_C} \sinh \gamma \ell & \cosh \gamma \ell \end{bmatrix}, \quad (C.26)$$

where Z_C is the characteristic impedance, and γ is the propagation constant. Z_C and γ are the same for THRU1 and THRU2. Denoting $A^{M,\ell_L-\ell_S} = A^{A,thru2} \left(A^{A,thru1} \right)^{-1}$, one will have

$$A^{M,\ell_L-\ell_S} = A^{PAD1} A^{A,\ell_L-\ell_S} \left(A^{PAD1} \right)^{-1} \quad (C.27)$$

In special case, $\ell_L = 2\ell_S$, the ABCD parameters of symmetric pads can be determined from THRU1 and THRU2. $A^{PAD1} = A^{PAD2} = \sqrt{\left(A^{M,thru1} \right)^{-1} A^{M,thru2} \left(A^{M,thru1} \right)^{-1}}$. If $\ell_L \neq 2\ell_S$, it is hard to solve the ABCD parameters without a PAD standard. However, it is not necessary to solve PAD parameters for de-embedding purpose because the PAD parameters are cancelled out during de-embedding as shown below.

To solve $A^{A,\ell_L-\ell_S}$ without solve A^{PAD1} and A^{PAD2} , $A^{M,\ell_L-\ell_S}$ is transformed to $Y^{M,\ell_L-\ell_S}$ using equations in Appendix B. From pad-open-short de-embedding, the Y-parameters of a symmetric PAD standard can be represented as

$$Y^{M,PAD} = \begin{bmatrix} Y_P & 0 \\ 0 & Y_P \end{bmatrix} \quad (C.28)$$

Thus, using Y-parameter representation, we have

$$Y^{A,\ell_L-\ell_S} = Y^{M,\ell_L-\ell_S} - \begin{bmatrix} Y_P & 0 \\ 0 & -Y_P \end{bmatrix} \quad (C.29)$$

Because transmission line is a symmetric structure, the Y-parameters of PAD can be cancelled out using $Y^{A,\ell_L-\ell_S} = \left[Y^{M,\ell_L-\ell_S} + \text{swap} \left(Y^{M,\ell_L-\ell_S} \right) \right] / 2$. $\text{swap} \left(Y^{M,\ell_L-\ell_S} \right)$ swaps the two ports of $Y^{M,\ell_L-\ell_S}$. $A^{A,\ell_L-\ell_S}$ can then be obtained from $Y^{A,\ell_L-\ell_S}$.

The Y-parameters of the left and right half of THRU1 both contains one probing pad and a transmission line with length $\ell_s/2$. Denoting the left half as Y^{IN*} , and the right half as Y^{OUT*} , Y^{IN*} and Y^{OUT*} are calculated as

$$Y^{IN*} = \begin{bmatrix} y_{11}^{M,thru1} - y_{12}^{M,thru1} - \frac{\gamma \ell_s}{4Z_C} & 2y_{12}^{M,thru1} \\ 2y_{12}^{M,thru1} & \frac{\gamma \ell_s}{4Z_C} - 2y_{12}^{M,thru1} \end{bmatrix}, \quad (C.30)$$

$$Y^{OUT*} = P Y^{IN*} P, \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (C.31)$$

Z_C and γ are extracted from $A^{A, \ell_L - \ell_s}$ as

$$Z_C = \sqrt{\frac{A_{12}^{A, \ell_L - \ell_s}}{A_{21}^{A, \ell_L - \ell_s}}} \quad \text{and} \quad \gamma = \frac{\cosh^{-1} A_{11}^{A, \ell_L - \ell_s}}{\ell_L - \ell_s}. \quad (C.32)$$

Thus, the ABCD parameters of the input and output networks are then given by $A^{IN} = A^{IN*} A^{\ell_X - \ell_s/2}$ and $A^{OUT} = A^{OUT*} A^{\ell_Y - \ell_s/2}$. The ABCD parameters of the desired device is obtained as

$$A^A = (A^{IN})^{-1} A^M (A^{OUT})^{-1}. \quad (C.33)$$

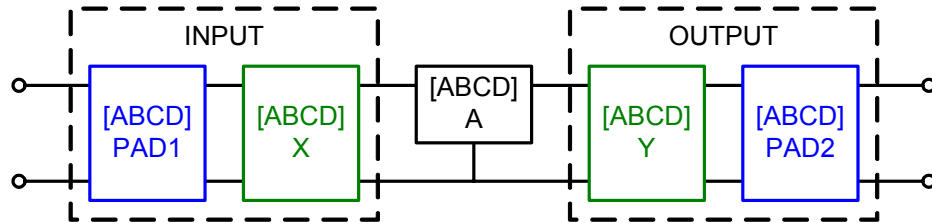


Fig. C.7 Equivalent circuit for transmission line de-embedding.

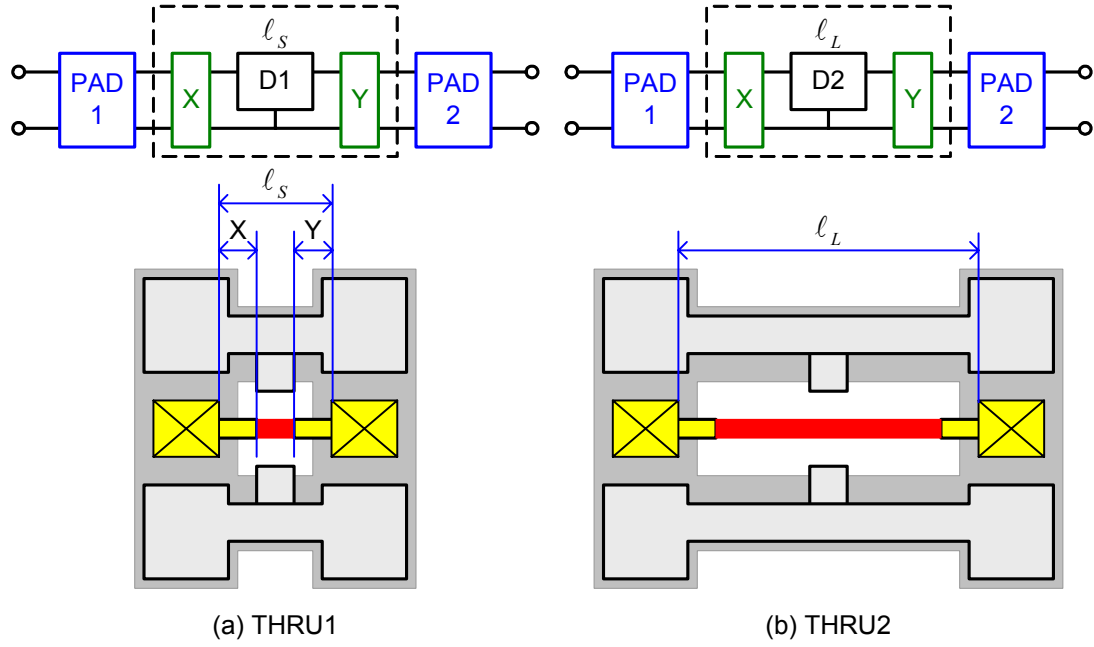


Fig. C.8 Equivalent circuits and layouts of THRU1 and THRU2 for transmission line de-embedding. The length of transmission line is not to scale.

APPENDIX D

SWITCH ERROR REMOVAL

D.1 Switch error removal equations

Fig. 3.2 shows a two-port measurement system with four receivers. The characteristics of the switch can be removed by making no assumption of Z_0 . For each two-port measurement, $S_{11} = b_0 / a_0$ and $S_{21} = b_3 / a_0$ are calculated in forward mode, while $S_{12} = b'_0 / a'_3$ and $S_{22} = b'_3 / a'_3$ are calculated in reverse mode. The subscript is the port number where the wave is monitored, while the superscript “'” means reverse mode.

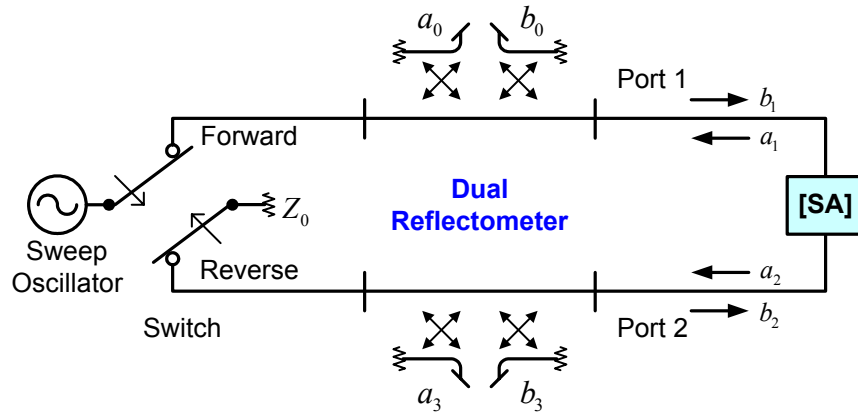


Fig. D.1 A two-port S-parameter measurement system with four receivers.

Under forward mode, if $a_3 = 0$, matched Z_0 termination, then the calculated S_{11} and S_{21} are the measured S_{11}^M and S_{21}^M of the two-port. If $a_3 \neq 0$, not matched Z_0 termination, the S-parameters of the two-port are defined using

$$\begin{bmatrix} b_0 \\ b_3 \end{bmatrix} = \begin{bmatrix} S_{11}^M & S_{12}^M \\ S_{21}^M & S_{22}^M \end{bmatrix} \begin{bmatrix} a_0 \\ a_3 \end{bmatrix}. \quad (\text{D.1})$$

Similarly, under reverse mode, if $a_0' = 0$, the calculated S_{12} and S_{22} are the measured S_{12}^M and S_{22}^M of the two-port. If $a_0' \neq 0$, the S-parameters of the two-port are defined using

$$\begin{bmatrix} b_0' \\ b_3' \end{bmatrix} = \begin{bmatrix} S_{11}^M & S_{12}^M \\ S_{21}^M & S_{22}^M \end{bmatrix} \begin{bmatrix} a_0' \\ a_3' \end{bmatrix}. \quad (\text{D.2})$$

Combining the forward and reverse mode configurations,

$$\begin{bmatrix} b_0 & b_0' \\ b_3 & b_3' \end{bmatrix} = \begin{bmatrix} S_{11}^M & S_{12}^M \\ S_{21}^M & S_{22}^M \end{bmatrix} \begin{bmatrix} a_0 & a_0' \\ a_3 & a_3' \end{bmatrix}. \quad (\text{D.3})$$

Therefore, the S-parameter of the two-port, S^M , is calculated as

$$\begin{bmatrix} S_{11}^M & S_{12}^M \\ S_{21}^M & S_{22}^M \end{bmatrix} = \begin{bmatrix} b_0 & b_0' \\ b_3 & b_3' \end{bmatrix} \begin{bmatrix} a_0 & a_0' \\ a_3 & a_3' \end{bmatrix}^{-1}, \quad (\text{D.4})$$

which can be rewritten as

$$S^M = \begin{bmatrix} \frac{b_0 a_3' - b_0' a_3}{\Delta} & \frac{b_0' a_0 - b_0 a_0'}{\Delta} \\ \frac{b_3 a_3' - b_3' a_3}{\Delta} & \frac{b_3' a_0 - b_3 a_0'}{\Delta} \end{bmatrix}, \Delta = a_0 a_3' - a_3 a_0' \quad (\text{D.5})$$

Substituting $S_{11} = b_0 / a_0$, $S_{21} = b_3 / a_3$, $S_{12} = b_3' / a_3'$ and $S_{22} = b_0' / a_0'$ into (D.5), S^M with switch error removed is calculated from the raw S-parameters exported from VNA as

$$S^M = \begin{bmatrix} \frac{S_{11} - S_{12}S_{21}\Gamma_1}{D} & \frac{S_{12} - S_{11}S_{12}\Gamma_2}{D} \\ \frac{S_{21} - S_{22}S_{21}\Gamma_1}{D} & \frac{S_{22} - S_{21}S_{12}\Gamma_2}{D} \end{bmatrix}, \Gamma_1 = \left. \frac{a_3}{b_3} \right|_{forward}, \Gamma_2 = \left. \frac{a'_0}{b'_0} \right|_{reverse} \quad (D.6)$$

Γ_1 and Γ_2 are user functions defined above for forward and reverse mode, which can only be measured using four-receiver VNA. $D = 1 - S_{21}S_{12}\Gamma_1\Gamma_2$. S^M is the measured S-parameters of the DUT after removing switch errors, while S_{11} , S_{21} , S_{12} , and S_{22} are the raw S-parameters directly saved from the VNA without switch error removal.

D.2 Step-by-step guide to measure the switch errors

1. Setup VNA

Define the frequency list, the input power level, and the averaging factor as the same as the setup used for on-wafer standards and transistor measurement.

2. Define user functions in VNA

Press the **MENU** key in **PARAMETER** block to bring the user parameter menu onto the CRT/LCD screen.

Define $\Gamma_1 = a_3 / b_3$ under forward mode first. a_3 and b_3 are the waves monitored by the receivers at Port 2. Thus, in VNA, they are named as a2 and b2.

- Select USER1.
- Press REDEFINE PARAMETER.
- Press DRIVE, PORT1.
- Press PHASE LOCK, a1.
- Press NUMERATOR, b2.
- Press DENOMINATOR, a2.

- Press CONVERSION, 1/S.
- Press PARAMETER LABEL, then enter a2/b2, then press TITLE DOWN, REDEFINE DONE.

Define $\Gamma_2 = a_0 / b_0$ under reverse mode. a_0 and b_0 are the waves monitored by the receivers at Port 1. Thus, in VNA, they are named as a1 and b1.

- Select USER2.
- Press REDEFINE PARAMETER.
- Press DRIVE, PORT2.
- Press PHASE LOCK, a2.
- Press NUMERATOR, b1.
- Press DENOMINATOR, a1.
- Press CONVERSION, 1/S.
- Press PARAMETER LABEL, then enter a1/b1, then press TITLE DOWN, REDEFINE DONE.

3. Measure Γ_1 and Γ_2

Display all of the four S-parameters on the screen first. Press the **DISPLAY** key in **MENUS** block. Select DISPLAY MODE, FOUR PARAM SPLIT. All of the four S-parameters, S11, S21, S12, and S22 are displayed on the screen.

Then, replace two of the S-parameters with the defined user functions, USER1 and USER2. Press the **MENU** key in **PARAMETER** block, select USER1 and USER2. The four parameters displayed on the screen are now 1/USER1, 1/USER2, S12, and S22.

Probe THRU standard on the Cascade ISS 101-190 substrate. Press the **MENU** key in **STIMULUS** block to bring the stimulus control menu onto the CRT/LCD screen. Select the MORE, then SINGLE to make a single measurement. Wait until the measurement is finished and HOLD is marked with underline.

4. Export data as a CITI file

Insert a floppy disk. Press the **DISK** key in the **AUXILIARY MENUS** block. Press STORE, MORE, DATA, enter the name of the file DD_ERR, and then press STORE FILE.

5. Example CITI file exported

```

CITIFILE A.01.01
#NA VERSION HP8510XF.01.02
NAME RAW_DATA
#NA REGISTER 6
VAR FREQ MAG 35
DATA USER[1] RI
DATA USER[2] RI
DATA S[1,2] RI
DATA S[2,2] RI
#NA DUPLICATES 0
#NA ARB_SEG 2000000000 70000000000 35
VAR_LIST_BEGIN
2000000000
4000000000
6000000000
8000000000
10000000000
12000000000
... ..
64000000000
66000000000
68000000000
70000000000
VAR_LIST_END
COMMENT      YEAR MONTH DAY HOUR MINUTE SECONDS
CONSTANT TIME 2007   08   10   15     00    03.0
BEGIN
-2.35986E0,1.20932E1
7.40478E0,-3.54687E0
-6.18872E0,-6.42651E0
-6.24414E0,1.01318E1
2.61523E1,-9.20996E0
-2.70185E1,1.02001E1

```

```

... ..*
-1.75854E0,-5.03784E0
3.19091E0,-3.38647E0
6.30932E0,4.04101E0
1.53857E1,3.04736E1
END
BEGIN
-3.03686E0,7.43920E0
6.72070E0,-3.75439E0
-4.42919E0,-5.90649E0
-8.03173E0,3.51220E0
2.05234E1,-0.71484E0
... ..*
-5.20507E0,-1.43994E1
2.87963E0,-3.32666E0
3.16540E0,2.25952E-1
3.81811E0,4.58984E0
END
BEGIN
-1.07403E0,6.15478E-1
3.27758E-1,-1.46313E0
1.61621E-1,1.29162E0
-7.56713E-1,-1.14001E0
9.71008E-1,7.04650E-1
4.18945E-1,-9.37377E-1
... ..*
-3.78418E-1,3.74939E-1
5.19653E-1,1.14196E-1
0.25878E-1,-5.13000E-1
-3.73977E-1,2.49313E-1
END
BEGIN
-1.68396E-1,-0.82214E-1
2.12471E-1,-0.45509E-1
-0.3302E-1,2.36198E-1
-2.22015E-1,0.30609E-1
1.02417E-1,1.53465E-2
0.18722E-1,1.90048E-1
... ..*
1.18751E-1,1.54907E-1
-0.9021E-1,-1.97334E-1
0.75592E-1,-1.50070E-1
-0.87142E-1,-3.28628E-1
END

```

APPENDIX E
CALIBRATION KIT SETUP

This is a step-by-step tutorial for calibration kit setup on Agilent VNA 8510C. **MENUS** means the block's name of a group of the keys which is printed on the front panel of the equipment. **CAL** means hardkey which is the button on the front panel under each block. The number and unit keys on the right side of the screen are not included. MORE means softkey on the screen which can be selected using the buttons on the right side of the screen. The following steps are for Cascade RF infinity probe with pitch size of 100 μ m, and Cascade ISS 101-190 substrate. The values entered are from the data sheet of the probes and the substrate, which are also listed in Table E.1.

1. Modify CalKit

Press hardkey **CAL** in the **MENUS** block, then select MORE, MODIFY 1 to modify the calibration coefficients for CalKit 1. Select MODIFY 2 to modify the calibration coefficients for CalKit 2.

2. Define calibration standards

Select DEFINE STANDARD.

Press 1, x1. Make sure the OPEN is underlined.

Press OPEN.

Select C0, enter -6.5, x1. Enter 0, x1 for C2, C3, and C4.

Select SPECIFY OFFSET, enter 0, x1.

Press STD OFFSET DONE.

Press LABEL STD, enter the name of the standard, e.g. OPEN-6.5.

Press TITLE DONE.

Press STD DONE.

3. Repeat Step 2 for SHORT, LOAD, and THRU standards

SHORT: Press DEFINE STANDARD, enter 2, x1, and then select SHORT.

LOAD: Press DEFINE STANDARD, enter 3, x1, and then select LOAD.

THRU: Press DEFINE STANDARD, enter 4, x1, and then DELAY/THRU.

4. Class assignment

Press SPECIFY CLASS

Select S11A, enter 1, x1. OPEN is defined as standard 1 in step 2.

Select S11B, enter 2, x1. SHORT is standard 2 in step 3.

Select S11C, enter 3, x1. LOAD is standard 3 in step 3.

Do the same for S22A, S22B, S22C, FWD TRANS, REV TRANS, FWD MATCH, REV MATCH, FWD ISOL`N, REV ISOL`N, using the corresponding class assignment values from Table E.1 (c).

Press SPECIFY CLASS DONE.

5. Label classes

Press LABEL CLASS

Select S11A, enter `OPEN-6.5`, and then LABEL DONE.

Select S11B, enter `SHORT 3.3`, and then LABEL DONE.

Select S11C, enter `LOAD 50`, and then LABEL DONE.

Do the same for S22A, S22B, S22C, FWD TRANS, REV TRANS, FWD MATCH, REV MATCH, FWD ISOL`N, REV ISOL`N, using the corresponding standard labels from Table E.1 (c).

Press LABEL CLASS DONE.

6. Label the calibration kit

Press LABEL KIT, enter the title of the calibration kit, e.g. `ISS100UM`, and then TITLE DONE, KIT DONE (MODIFIED).

7. Save calibration kit in VNA and floppy disk

Press SAVE. To save the calibration kit to CALKIT 1 in the VNA memory.

Press DISC, in the **AUXILIARY MENUS** block, the select STORE, CALKIT1-2, CALKIT 1, enter a filename, e.g. `CK_100`. The calibration kit will be saved on a floppy disk with name CK_100, which can be loaded into VNA later using LOAD, CALKIT1-2, CALKIT 1.

Table E.1 Calibration Kit Coefficients

(a) SOLT calibration standard definitions for Cascade RF infinity GSG probe, 100um pitch, and Cascade ISS 101-190.

STANDARD NO	TYPE	C0 10 ⁻¹⁶ F	C1 10 ⁻²⁷ F/Hz	C2 10 ⁻³⁶ F/Hz	C3 10 ⁻⁴⁶ F/Hz	FIXED or SLIDING	OFFSET		FREQUENCY GHz		COAX or WAVEGUIDE	STANDARD LABEL
							DELAY psec	LOSS MΩ/s	Z ₀ Ω	MIN	MAX	
1	OPEN	-6.5	0	0	0		0	0	50	0	999	OPEN-6.5
2	SHORT	3.3					0	0	50	0	999	SHORT 3.3
3	LOAD					Fixed	-0.0008	0	500	0	999	LOAD 50
4	THRU						1	0	50	0	999	THRU 1P

(b) Constant for Cascade RF infinity GSG probes

Pitch	C Open	L Short	L Term
100um	-6.5fF	3.3pH	-0.4pH
150um	-6.7fF	8.2pH	3.7pH

- 1) ISS substrate : Cascade ISS 101-190.
- 2) Probe : Cascade RF infinity GSG probe.
- 3) L Term : Modeled as a high impedance transmission line offset with $Z_0=500\Omega$, $\text{delay}=L/Z_0$.
- 4) THRU delay is pitch size independent, which depends on the signal to signal pad distance of the test structure only. For transistor test structures, THRU with 200um length and 1psec delay is always used.

(c) Standard Class Assignments

	A	B	C	D	E	F	G	STANDARD CLASS LABEL
S ₁₁ A	1							OPEN-6.5
S ₁₁ B	2							SHORT 3.3
S ₁₁ C	3							LOAD 50
S ₂₂ A	1							OPEN-6.5
S ₂₂ B	2							SHORT 3.3
S ₂₂ C	3							LOAD 50
Forward Transmission	4							THRU 1P
Reverse Transmission	4							THRU 1P
Forward Match	4							THRU 1P
Reverse Match	4							THRU 1P
Forward Isolation	3							LOAD 50
Reverse Isolation	3							LOAD 50

APPENDIX F

THE RELATIONSHIP BETWEEN OPEN-SHORT AND FOUR-PORT

Fig. 4.1 illustrates the four-port network for on-wafer parasitics. There are two external ports, and two internal ports. V_m , I_m , V_m^* , and I_m^* , are the voltages and currents at each port. The subscript m is the port number, $m=1,2$. The superscript $*$ means internal ports. Based on the definition of Y-parameters, the voltages and currents can be related through the Y-parameters of the four-port network as

$$\begin{bmatrix} V_1 \\ V_2 \\ V_1^* \\ V_2^* \end{bmatrix} = \begin{bmatrix} y_{11}^{ee} & y_{12}^{ee} & y_{11}^{ei} & y_{12}^{ei} \\ y_{21}^{ee} & y_{22}^{ee} & y_{21}^{ei} & y_{22}^{ei} \\ y_{11}^{ie} & y_{12}^{ie} & y_{11}^{ii} & y_{12}^{ii} \\ y_{21}^{ie} & y_{22}^{ie} & y_{21}^{ii} & y_{22}^{ii} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_1^* \\ I_2^* \end{bmatrix}. \quad (\text{F.1})$$

To make the following derivations easier to read, voltage and current vectors are defined as

$$V_e = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, I_e = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, V_i = \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix}, I_i = \begin{bmatrix} I_1^* \\ I_2^* \end{bmatrix}. \quad (\text{F.2})$$

The superscript e means external ports, while i means internal ports. The relationship in (F.1) can be rewritten as

$$\begin{bmatrix} I_e \\ I_i \end{bmatrix} = \begin{bmatrix} Y^{ee} & Y^{ei} \\ Y^{ie} & Y^{ii} \end{bmatrix} \begin{bmatrix} V_e \\ V_i \end{bmatrix}, \quad (\text{F.3})$$

where

$$Y_{ee} = \begin{bmatrix} y_{11}^{ee} & y_{12}^{ee} \\ y_{21}^{ee} & y_{22}^{ee} \end{bmatrix}, Y_{ei} = \begin{bmatrix} y_{11}^{ei} & y_{12}^{ei} \\ y_{21}^{ei} & y_{22}^{ei} \end{bmatrix}, Y_{ie} = \begin{bmatrix} y_{11}^{ie} & y_{12}^{ie} \\ y_{21}^{ie} & y_{22}^{ie} \end{bmatrix}, Y_{ii} = \begin{bmatrix} y_{11}^{ii} & y_{12}^{ii} \\ y_{21}^{ii} & y_{22}^{ii} \end{bmatrix}. \quad (\text{F.4})$$

The two-port network between the external ports gives

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} y_{11}^{DUT} & y_{12}^{DUT} \\ y_{21}^{DUT} & y_{22}^{DUT} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, Y^{DUT} = \begin{bmatrix} y_{11}^{DUT} & y_{12}^{DUT} \\ y_{21}^{DUT} & y_{22}^{DUT} \end{bmatrix}. \quad (\text{F.5})$$

With the direction of currents defined in Fig. 4.1, the two-port network between the internal ports gives

$$\begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} = \begin{bmatrix} y_{11}^A & y_{12}^A \\ y_{21}^A & y_{22}^A \end{bmatrix} \begin{bmatrix} -I_1^* \\ -I_2^* \end{bmatrix}, Y^A = \begin{bmatrix} y_{11}^A & y_{12}^A \\ y_{21}^A & y_{22}^A \end{bmatrix}. \quad (\text{F.6})$$

Through the Y-parameters of the four-port network, Y^{DUT} and Y^A are related as

$$Y^{DUT} = Y^{ee} - Y^{ei} (Y^A + Y^{ii})^{-1} Y^{ie}, \quad (\text{F.7})$$

or,

$$Y^A = -Y^{ii} - Y^{ie} (Y^{DUT} - Y^{ee})^{-1} Y^{ei}. \quad (\text{F.8})$$

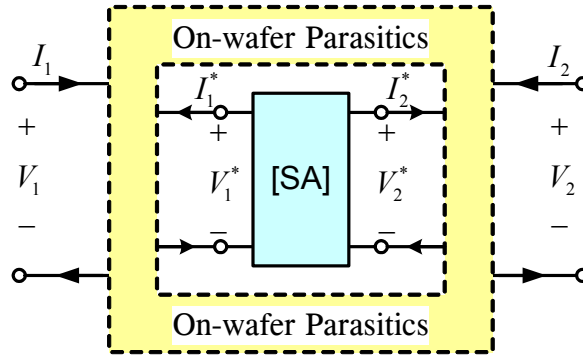


Fig. F.1. Block diagram of the 4-port network for on-wafer parasitics using I-V representation.

Since the Y-parameters of ideal OPEN and SHORT are $Y^{A,open} = [0]_{2 \times 2}$ and

$(Y^{A,short})^{-1} = [0]_{2 \times 2}$, using (F.7), $Y^{DUT,open}$ and $Y^{DUT,short}$ are

$$Y^{DUT,open} = Y_{ee} - Y_{ei} (Y_{ii})^{-1} Y_{ie} , \quad (F.9)$$

$$Y^{DUT,short} = Y_{ee} . \quad (F.10)$$

Recall the open-short de-embedded Y-parameters of the device, Y^{OS} ,

$$Y^{OS} = \left[(Y^{DUT} - Y^{DUT,open})^{-1} - (Y^{DUT,short} - Y^{DUT,open})^{-1} \right]^{-1} . \quad (F.11)$$

Substituting (F.9) and (F.10) into the Y^{OS} expression above,

$$Y^{DUT} - Y^{DUT,open} = Y^{ei} \left[(Y^{ii})^{-1} - (Y^A + Y^{ii})^{-1} \right] Y^{ie} , \quad (F.12)$$

$$Y^{DUT,short} - Y^{DUT,open} = Y^{ei} (Y^{ii})^{-1} Y^{ie} , \quad (F.13)$$

and thus

$$Y^{OS} = \left[\left(Y^{ei} \left[(Y^{ii})^{-1} - (Y^A + Y^{ii})^{-1} \right] Y^{ie} \right)^{-1} - \left(Y^{ei} (Y^{ii})^{-1} Y^{ie} \right)^{-1} \right]^{-1} . \quad (F.14)$$

The equation is too complicated to give any clue of the relationship between Y^{OS} and

Y^A . It must be simplified. The first thing can be done is taking the Y^{ei} and Y^{ie} out.

$$Y^{OS} = Y^{ei} \left\{ \left[(Y^{ii})^{-1} - (Y^A + Y^{ii})^{-1} \right]^{-1} - Y^{ii} \right\}^{-1} Y^{ie} . \quad (F.15)$$

It is difficult to further simplify the equation because of the plus-minus operators inside the brace. To eliminate the plus-minus operations, two identity matrices,

$(Y^A + Y^{ii})^{-1} (Y^A + Y^{ii})$ and $(Y^{ii})(Y^{ii})^{-1}$ are added to (F.15),

$$Y^{OS} = Y^{ei} \left\{ \left[\left(Y^A + Y^{ii} \right)^{-1} \left(Y^A + Y^{ii} \right) \left(Y^{ii} \right)^{-1} - \left(Y^A + Y^{ii} \right)^{-1} \left(Y^{ii} \right) \left(Y^{ii} \right)^{-1} \right]^{-1} - Y^{ii} \right\}^{-1} Y^{ie}. \quad (\text{F.16})$$

Taking the common elements, $\left(Y^A + Y^{ii} \right)^{-1}$ and $\left(Y^{ii} \right)^{-1}$, out of the square brackets leads to

$$Y^{OS} = Y^{ei} \left\{ \left(Y^{ii} \right) \left[Y^A \right]^{-1} \left(Y^A + Y^{ii} \right) - Y^{ii} \right\}^{-1} Y^{ie}, \quad (\text{F.17})$$

which is equivalent to

$$Y^{OS} = Y^{ei} \left\{ Y^{ii} + Y^{ii} \left[Y^A \right]^{-1} Y^{ii} - Y^{ii} \right\}^{-1} Y^{ie}. \quad (\text{F.18})$$

This gives a very simple relationship between Y^{OS} and Y^A ,

$$Y^{OS} = Y^{ei} \left(Y^{ii} \right)^{-1} Y^A \left(Y^{ii} \right)^{-1} Y^{ie}. \quad (\text{F.19})$$

Although (F.19) is derived for on-wafer parasitics and starts from open-short de-

embedding, the solution is general to single-step calibration as long as $Y^{A,open} = [0]_{2 \times 2}$

and $\left(Y^{A,short} \right)^{-1} = [0]_{2 \times 2}$. The only difference is that, when it is applied on the measured

raw S-parameters without ISS calibration, $Y^{DUT,open}$, $Y^{DUT,short}$, and Y^{OS} do not have

their physical meanings as what they have in two-step calibration.

APPENDIX G

SINGULARITY OF LINEAR EQUATION SET

G.1 Typical calibration standards

The most common calibration standards used for S-parameter measurement are two-port standards, through (THRU) and delay (DELAY), and one-port standards, match (M), short (S), and open (O). The one-port standards are used in pairs to build a two-port standard for two-port system calibration. For example, the LEFT standard used for four-port calibration can be viewed as a M-O standard, which means a matched load at Port 1, and an open standard at Port 2. A zero length THRU is kept for all of the combinations examined below for two reasons. First, the set of standards must include a two-port standard to measure the transmission errors. That means, a THRU or DELAY standard must be included. Secondly, the ends of the interconnects of Port 1 and Port 2 are very close for on-wafer transistor structures. Thus, a zero length THRU structure is the one of the simplest structures to be built on-wafer. The results shown below are from Cadence simulation. The parasitic network is built using ideal resistor, capacitor, and inductors with values close to the values extracted from measurement. The M standard is an ideal 50Ω resistor since the VNA system is a 50Ω system.

G.2 Singularity of on-wafer standards

An analytical proof for the singularity of the combinations of standards is complicated. However, it can be easily examined by numerical simulation examples using condition number of the coefficient matrix. The condition number is defined as the ratio of the largest singular value over the smallest singular value of the matrix. For four standards, there are 16 equations written in matrix as $A_{16 \times 15} T'_{15 \times 1} = B_{16 \times 1}$. If the coefficient matrix $A_{16 \times 15}$ has zero singular values, $A_{16 \times 15}$ is not full rank, and the number of unknowns can be solved equals to the number of non-zero singular values. If $A_{16 \times 15}$ is full rank, but has extremely small singular values, which leads to an extremely large condition number, the set of equation is ill-conditioned (singular), and the validity of the solution is questionable. Assuming THRU is taken as one of the four standards, and the other three standards are chosen from the pairs consisting O, S or M, i.e. O-O, S-S, M-M, O-S, S-O, O-M, M-O, S-M, M-S, there are 84 different combinations.

Fig. G.1 compares the condition number, the minimum singular value and the maximum singular value for four sets of standards. The coefficient matrix $A_{16 \times 15}$ are all singular since the condition numbers are extremely large for all cases. For five standards, the coefficient matrix is $A_{20 \times 15}$, and there are 126 possible combinations if THRU is chosen. 46 combinations was shown to be singular in the reference. The nonsingular combinations are listed in Table G.1.

Fig. G.2 compares condition number, minimum and maximum singular values for five sets of standards. The results indicate that these combinations are nonsingular and can provide valuable T solutions. Among the five sets of standards, the combination of

THRU, O-O, S-S, S-M, M-S gives the smallest condition number, and thus the best tolerance to measurement errors. If the five standards are nonsingular, then adding more standards will not help to improve the validity of the T solution. Fig. G.3 compares the condition number, minimum and maximum singular values for 5, 6, and 7 standards. Two sets of nonsingular five standards are compared. Both of them show that adding more standards do not reduce the condition number of the coefficient matrix.

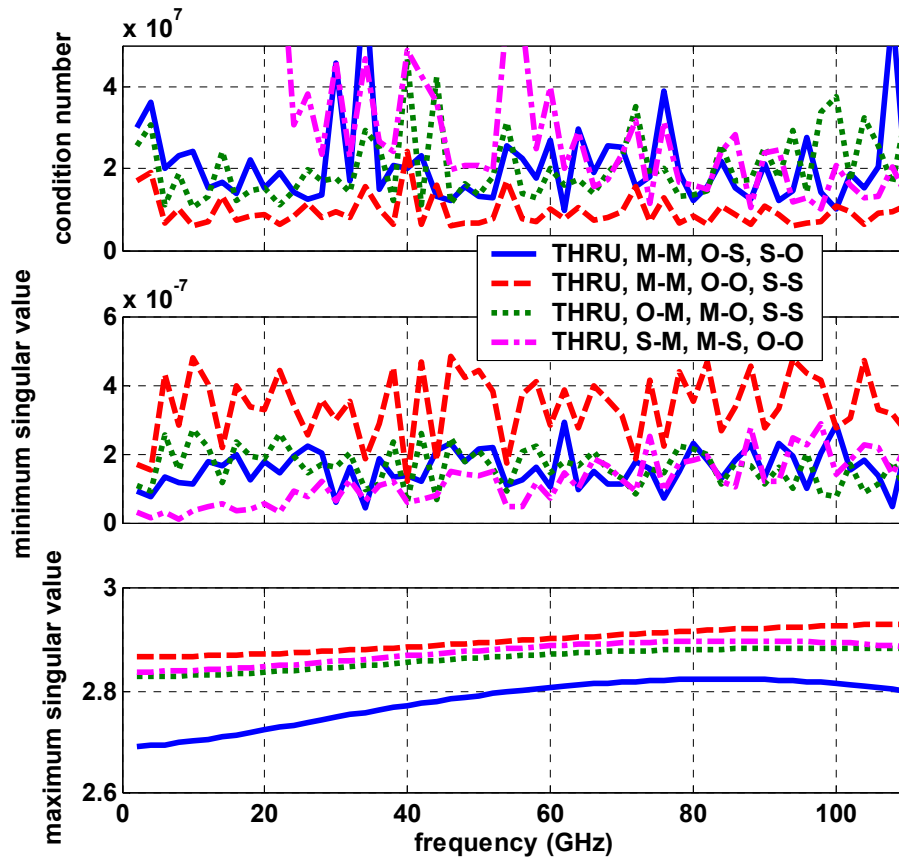


Fig. G.1 Condition number, minimum and maximum singular value for four standards.

Table G.1 Nonsingular combinations of five two-port calibration standards for 16 term error model. Assuming one standard is a zero length THRU.

THRU	A-M	M-A	A-A	B-B	THRU	M-M	A-A	B-B	A-B
THRU	A-M	B-A	A-A	B-B	THRU	M-M	A-A	B-B	A-M
THRU	A-M	B-M	A-A	B-B	THRU	M-M	A-A	A-B	B-A
THRU	A-M	M-A	B-M	A-A	THRU	M-M	A-A	A-B	M-A
THRU	A-M	M-B	B-M	A-A	THRU	M-M	A-A	A-B	M-B
THRU	A-M	B-A	B-M	A-A	THRU	M-M	A-A	A-M	M-A
THRU	A-M	B-A	M-B	A-A	THRU	M-M	A-A	A-M	M-B
THRU	A-M	B-A	A-B	B-B	THRU	M-M	A-A	B-M	M-B
THRU	A-M	M-A	A-B	B-B	THRU	M-M	A-B	A-M	B-A
THRU	A-M	B-M	A-B	B-B	THRU	M-M	A-B	A-M	M-A
THRU	A-M	M-B	B-A	B-B	THRU	M-M	A-B	A-M	M-B
THRU	A-M	M-B	B-M	B-B	THRU	M-M	A-B	B-M	M-A
THRU	A-M	M-A	B-M	B-B					
THRU	A-M	M-A	B-M	A-B					
THRU	A-M	M-A	B-M	B-A					
THRU	A-M	M-B	A-B	B-A					
THRU	A-M	M-B	B-M	A-B					
THRU	A-M	M-B	B-M	B-A					
THRU	A-M	B-M	A-B	B-A					

A = open, B = short, or A = short, B = open.
Reference: K. J. Silvonen, "Calibration of 16-term error model," Electronics Lett., vol. 29, no. 17, pp. 1544-1545, 1993.

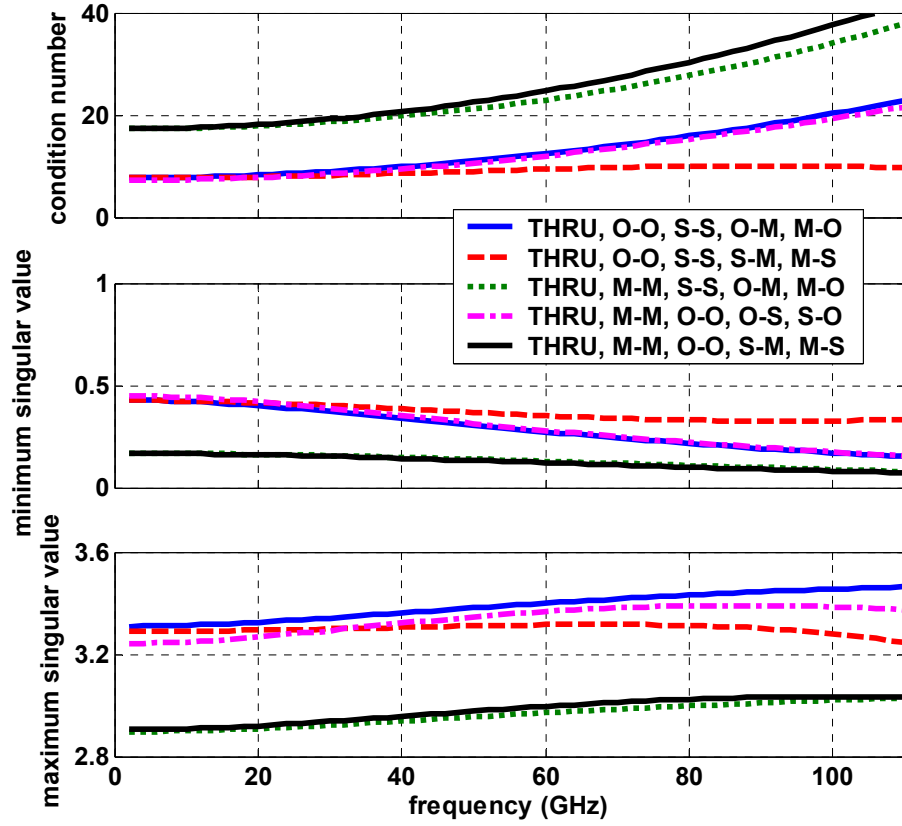


Fig. G.2 Condition number, minimum and maximum singular value for five standards.

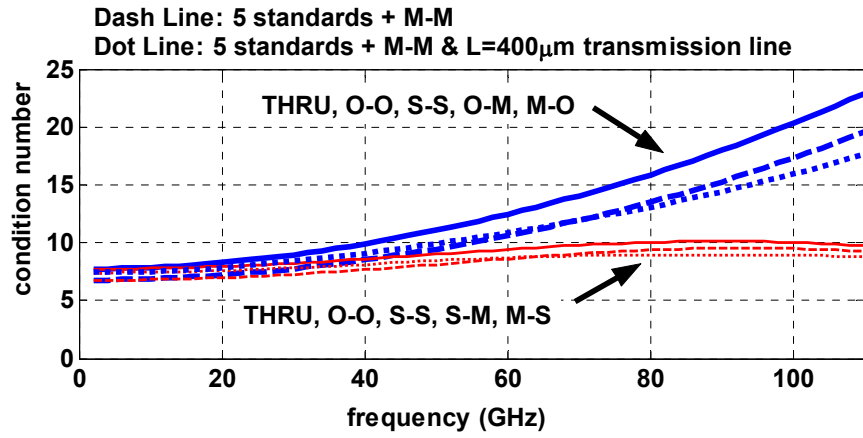


Fig. G.3 Condition number for multiple number of standards.

APPENDIX H

ONE-PORT ERROR CORRECTION

Fig. H.1 shows the block diagram for a one-port system. The system consists of a sweep oscillator, a dual-refractometer consisting of two couplers connected back-to-back, and the unknown one-port DUT, Γ^{DUT} . The direction of power flow through the system is indicated using arrows. a_0 and b_0 are the incident and reflected waves measured by the VNA. The measured reflection coefficient of the unknown one-port is defined as $\Gamma^M = b_0 / a_0$. The linear errors introduced by the imperfect reflectometer can be modeled by a fictitious two-port error adapter between the reflectometer and the unknown one-port. This results in a perfect reflectometer with no loss, no mismatch, and no frequency response errors.

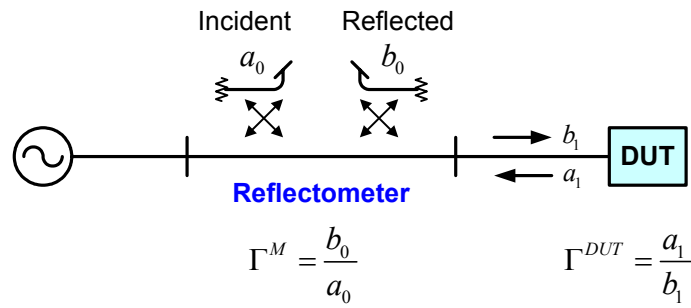


Fig. H.1 The block diagram for a one-port measurement.

H.1 Error adaptor for one-port system

Fig. H.2 shows the fictitious two-port error adaptor for a one-port system. The error adaptor has four error terms. Defining incident waves to the error adaptor as a_0 and a_1 , the reflected waves to the error adaptor as b_0 and b_1 . a means incident wave, b means reflected wave. The subscript is the port number. The measured and the actual reflection coefficients of the unknown one-port are $\Gamma^M = b_0 / a_0$ and $\Gamma^{DUT} = a_1 / b_1$. Written in matrix, the S-parameters of the two-port error adaptor can be defined using the waves as

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} e_{00} & e_{01} \\ e_{10} & e_{11} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}, \quad E = \begin{bmatrix} e_{00} & e_{01} \\ e_{10} & e_{11} \end{bmatrix} \quad (\text{H.1})$$

The 2×2 matrix E is the S-parameters of the two-port error adaptor.

The same relation can be equivalently represented using the signal flow graph in Fig. H.3. The system directivity e_{00} can be best understood when an ideal match load is under test. Part of the incident a_0 is reflected back to b_0 through the branch labeled e_{00} , independent of the Γ^{DUT} . Thus, when measuring Γ^M , there must be some residual signals measured.

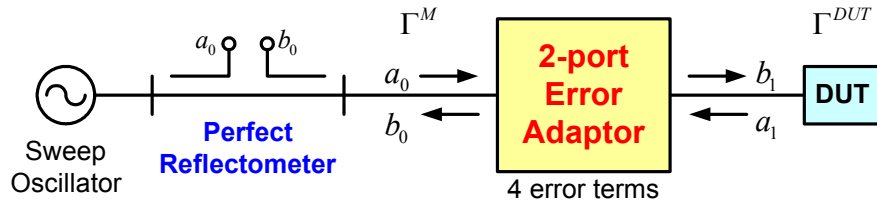


Fig. H.2 The combined two-port error adaptor for one-port S-parameter measurement.

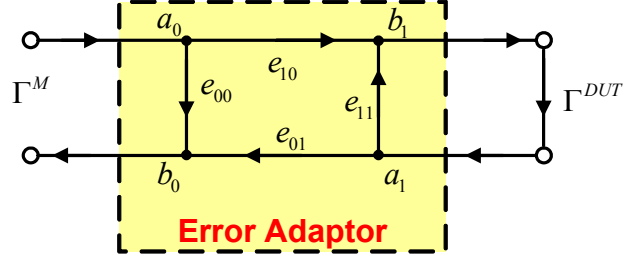


Fig. H.3 Signal flow graph of the two-port error adaptor in one-port measurement.

H.2 Relationship between Γ^M and Γ^{DUT}

Denoting $\Gamma^M = b_0 / a_0$ and $\Gamma^{DUT} = a_1 / b_1$ in (H.1), Γ^M and Γ^{DUT} can be related through

$$(e_{10}e_{01} - e_{00}e_{11})\Gamma^{DUT} + \Gamma^M e_{11}\Gamma^{DUT} + e_{00} = \Gamma^M. \quad (\text{H.2})$$

By measuring three standards with known Γ^{DUT} , three equations containing the unknown error terms are built. Then the error terms e_{00} , $(e_{10}e_{01})$, and e_{11} can be solved.

After that, Γ^{DUT} for any measured Γ^M can be obtained using

$$\Gamma^{DUT} = \frac{\Gamma^M - e_{00}}{(\Gamma^M - e_{00})e_{11} + e_{10}e_{01}}. \quad (\text{H.3})$$

Note that, only three error terms, e_{00} , $(e_{10}e_{01})$, and e_{11} , need to be solved for error correction purpose. This is because of the ratio nature of S-parameter measurement. The most widely used standards are OPEN, SHORT, and LOAD. Without specification, LOAD standard in this work means matched Z_0 load. Fig. H.4 show the magnitude of the solved error terms, $|e_{00}|$, $|e_{10}e_{01}|$, and $|e_{11}|$.

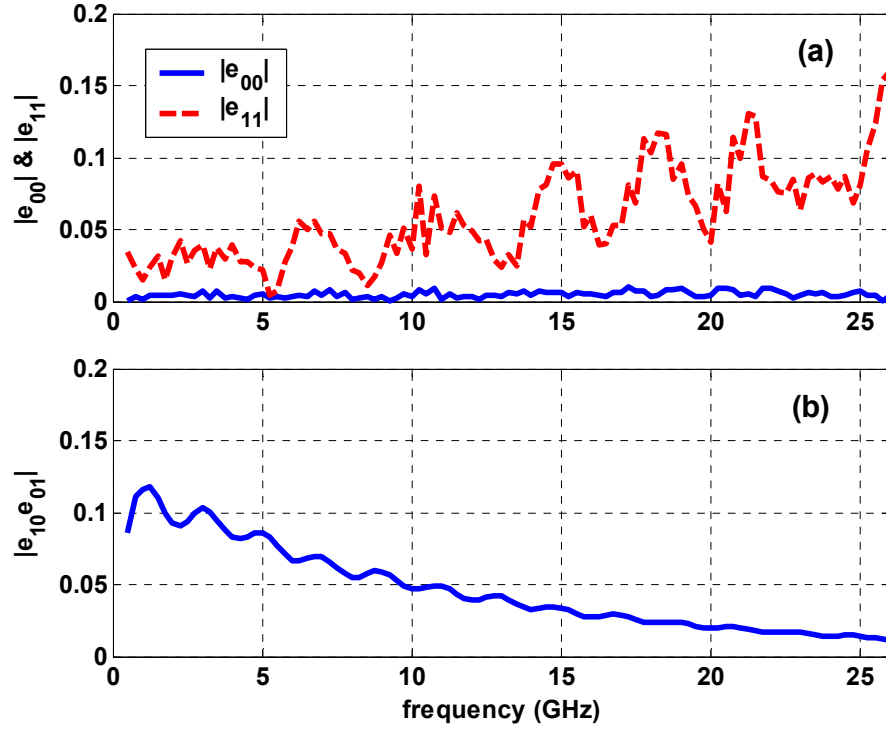


Fig. H.4 The three error terms solved using OPEN, SHORT and LOAD standards.

The relationship between Γ^M and Γ^{DUT} in (H.2) is a nonlinear function in terms of the error terms, e_{00} , $(e_{10}e_{01})$, and e_{11} . Due to the difficulty in solving nonlinear equations, a linear equation in terms of the error terms is developed next as a generalized interpretation which can be easily extended to two-port system.

H.3 A generalized interpretation

The linear equation is derived from the transmission parameters (T-parameters) of the two-port error adapter. In matrix, the T-parameters of the error adapter is written as

$$\begin{bmatrix} b_0 \\ a_0 \end{bmatrix} = \begin{bmatrix} t_1 & t_3 \\ t_2 & t_4 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}. \quad (\text{H.4})$$

Similarly, denoting $\Gamma^M = b_0 / a_0$ and $\Gamma^{DUT} = a_1 / b_1$, Γ^M and Γ^{DUT} are related through

$$t_1 \Gamma^{DUT} - \Gamma^M t_2 \Gamma^{DUT} + t_3 - \Gamma^M t_4 = 0. \quad (\text{H.5})$$

This is a linear equation in terms of the elements in T . Since T-parameters represent the S-parameters of the same error adapter, (H.5) can be rewritten in a similar format as (H.2),

$$\frac{t_1}{t_4} \Gamma^{DUT} - \Gamma^M \frac{t_2}{t_4} \Gamma^{DUT} + \frac{t_3}{t_4} = \Gamma^M \quad (\text{H.6})$$

Comparing (H.2) and (H.6), the elements in T can be related to the elements in E as

$$\frac{t_1}{t_4} = e_{10}e_{01} - e_{00}e_{11}, \quad \frac{t_2}{t_4} = -e_{11}, \quad \frac{t_3}{t_4} = e_{00}. \quad (\text{H.7})$$

Note that all of the unknown terms are normalized to t_4 and the equation is still a linear equation of the unknown terms. After normalization, only three unknowns need to be solved. Three standards, e.g. OPEN, SHORT, and LOAD, can be used to solve the three equations as below

$$\begin{bmatrix} \Gamma^{DUT1} & -\Gamma^{M1}\Gamma^{DUT1} & 1 \\ \Gamma^{DUT2} & -\Gamma^{M2}\Gamma^{DUT2} & 1 \\ \Gamma^{DUT3} & -\Gamma^{M3}\Gamma^{DUT3} & 1 \end{bmatrix} \begin{bmatrix} t_1 / t_4 \\ t_2 / t_4 \\ t_3 / t_4 \end{bmatrix} = \begin{bmatrix} \Gamma^{M1} \\ \Gamma^{M2} \\ \Gamma^{M3} \end{bmatrix}. \quad (\text{H.8})$$

Once the three error terms are solved, the system errors of any measured Γ^M can then be calibrated using an alternative of (H.6) as

$$\Gamma^{DUT} = \frac{\Gamma^M - \frac{t_3}{t_4}}{\frac{t_1}{t_4} - \Gamma^M \frac{t_2}{t_4}}. \quad (\text{H.9})$$

Considering the linear equation in (H.5), at first glance, one may think with four measurements, t_1 , t_2 , t_3 , and t_4 can be completely solved without normalization. However the resulting linear matrix problem is homogenous. For four measurements, the four linear equations written in matrix are

$$\begin{bmatrix} \Gamma^{DUT1} & -\Gamma^{M1}\Gamma^{DUT1} & 1 & -\Gamma^{M1} \\ \Gamma^{DUT2} & -\Gamma^{M2}\Gamma^{DUT2} & 1 & -\Gamma^{M2} \\ \Gamma^{DUT3} & -\Gamma^{M3}\Gamma^{DUT3} & 1 & -\Gamma^{M3} \\ \Gamma^{DUT4} & -\Gamma^{M4}\Gamma^{DUT4} & 1 & -\Gamma^{M4} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (\text{H.10})$$

If the four unknowns can all be solved, the coefficient matrix must be full rank. This leads to an all zero solution of T . So the rank of the coefficient matrix must be smaller than 4, which means in maximum, only three of the unknowns can be solved. This is theoretically attributed to the ratio nature of S-parameters and the inability to solve e_{10} and e_{01} independently. The normalization of T elements will not affect error calibration at all.

APPENDIX I

DERIVATION OF FIRST ORDER INPUT IP3

Fig. 8.1 shows the small signal equivalent circuit used for analytical IP3 analysis. $v_s = V_s (\cos \omega_1 t + \cos \omega_2 t)$ is the two tone input signal. $\omega_1 = 2\pi f_1$ and $\omega_2 = 2\pi f_2$. R_s is the source resistance, while R_L is the load resistance. C_{gs} and C_d are small signal gate to source capacitance and drain to substrate capacitance. First order IP3 theory considers the small-signal nonlinear current source i_{ds} as a function of v_{gs} only. With small-signal input, it can be approximated by the first three order Taylor expansion as

$$i_{ds} = g_m v_{gs} + K_{2g_m} v_{gs}^2 + K_{3g_m} v_{gs}^3. \quad (I.1)$$

g_m , K_{2g_m} , and K_{3g_m} are the first three order nonlinearity coefficients of i_{ds} , which can be calculated as

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}}, \quad K_{2g_m} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2}, \quad K_{3g_m} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3}. \quad (I.2)$$

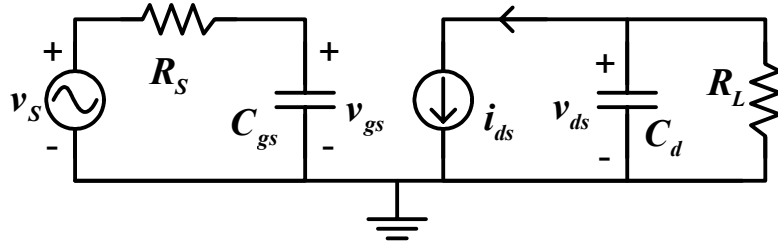


Fig. I.1 The small signal equivalent circuit used for IP3 analysis.

For a two-tone input signal, $v_{gs} = A(\cos \omega_1 t + \cos \omega_2 t)$. The amplitudes of v_s and v_{gs} are related by $V_s = A|1 + j\omega C_{gs} R_s|$. The two frequencies are f_1 and f_2 . $\omega_1 = 2\pi f_1$, and $\omega_2 = 2\pi f_2$. Therefore, the output drain current in (I.1) contains components at frequencies $m\omega_1 + n\omega_2$, m and n are integers. The magnitude of the fundamental components at ω_1 and ω_2 are $g_m A + 9/4 K_3 g_m A^3$, and the 3rd order intermodulation components at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are $3/4 K_3 g_m A^3$. Under small signal excitation, the magnitude of the fundamental components are approximately $g_m A$, since the second term can be ignored when compared with the $g_m A$ term.

The 3rd order intermodulation distortion (IM3) is defined as the ratio of the 3rd order intermodulation components and the fundamental components,

$$IM_3 = \frac{\frac{3}{4} K_3 g_m A^3}{g_m A}. \quad (I.3)$$

The 3rd order intercept point is the point where the fundamental and the 3rd order intermodulation components are equal, which is $IM_3 = 1$ in (I.3). The amplitude of v_{gs} at the 3rd order intercept point is calculated as

$$A^2 = \frac{4}{3} \left| \frac{g_m}{K 3g_m} \right|. \quad (\text{I.4})$$

Therefore, V_s at the 3rd order intercept point is

$$V_s^2 = \frac{4}{3} \left| \frac{g_m}{K 3g_m} \right| \left[1 + \left(\omega C_{gs} R_s \right)^2 \right]. \quad (\text{I.5})$$

The corresponding maximum available power at the power source v_s is defined as input referred IP3 (IIP3) as

$$IIP3 = \frac{V_s^2}{8R_s} = \frac{1}{6R_s} \frac{1 + \left(\omega C_{gs} R_s \right)^2}{\left| \frac{K 3g_m}{g_m} \right|}. \quad (\text{I.6})$$

APPENDIX J

DERIVATION OF INPUT IP3 BASED ON VOLTERRA SERIES

Volterra Series approximates the output of a nonlinear system in a manner similar to Taylor series approximation. For sufficiently small inputs, the output of a nonlinear system can be described as the sum of the transfer functions below order three. The first order transfer function $H_1(s)$ is essentially the transfer function of the linearized circuit. The 2nd and 3rd order transfer functions, $H_2(s_1, s_2)$ and $H_3(s_1, s_2, s_3)$, can be solved in increasing order by repeatedly solving the same linear circuit using different order excitations.

Fig. J.1 shows the small signal equivalent circuit for a MOS transistor excited by a voltage source with source resistance R_s and loaded with a resistance R_L . C_{gs} and C_d are the gate-source and drain-bulk capacitance. The nonlinear current i_{ds} is controlled by gate-source and drain-source voltages, which can be approximately calculated as the sum of a series containing powers of the control voltages. The i_{ds} expression limited to first-, second-, and third-order nonlinear behavior is

$$\begin{aligned}
 i_{ds} = & g_m v_{gs} + g_o v_{ds} \quad \dots\dots\dots \text{first order(linear)} \\
 & + K_2 g_m v_{gs}^2 + K_2 g_o v_{ds}^2 + K_2 g_m g_o v_{gs} v_{ds} \quad \dots\dots\dots \text{second order} \\
 & + K_3 g_m v_{gs}^3 + K_3 g_o v_{ds}^3 + K_3 2g_m g_o v_{gs}^2 v_{ds} + K_3 g_m 2g_o v_{gs} v_{ds}^2 \quad \dots\dots \text{third order}
 \end{aligned} \quad . \text{ (J.1)}$$

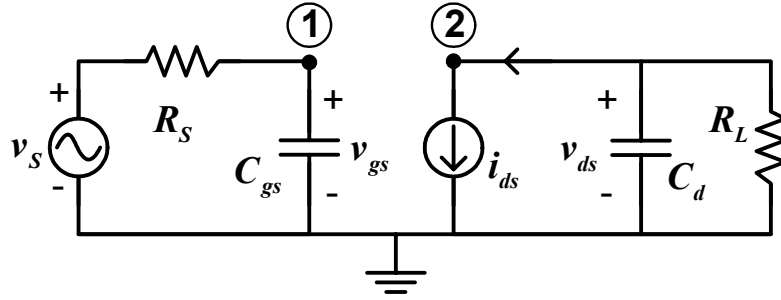


Fig. J.1 The small signal equivalent circuit used for IP3 analysis.

Applying Kirchoff's current law at node 1 and 2 in Fig. J.1 yield

$$\begin{bmatrix} \frac{1}{R_s} + sC_{gs} & 0 \\ g_m & \frac{1}{R_L} + g_o + sC_d \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{R_s} V_s \\ 0 \end{bmatrix} \quad (\text{J.2})$$

The voltages above are Laplace transforms. Denoting $Y_L(s) = g_o + sC_d + 1/R_L$ and

$Y_s(s) = 1/R_s + sC_{gs}$, (J.2) can be rewritten as

$$\begin{bmatrix} Y_s(s) & 0 \\ g_m & Y_L(s) \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{R_s} V_s \\ 0 \end{bmatrix}. \quad (\text{J.3})$$

The 2×2 matrix in the left-hand side is the admittance matrix of the circuit. V_1 , V_2 and

V_s are Laplace transforms.

J.1 First order kernels

The first order kernels are calculated from the response of the linearized circuit to external input V_s . Fig. J.2 gives the linearized equivalent circuit. The voltage source is converted to a current source, which is the only excitation of the circuit when

calculating the first order kernels. V_1 and V_2 reduce to the first order transfer functions of the voltages at node 1 and 2 when V_s is set to one. The transfer functions at node 1 and 2 are denoted as $H_{1_1}(s)$ and $H_{1_2}(s)$. The first subscript indicates the order of the transfer functions, while the second subscript corresponds to the number of the node. Hence the transfer functions can be solved from the matrix equation below,

$$\begin{bmatrix} Y_s(s) & 0 \\ g_m & Y_L(s) \end{bmatrix} \begin{bmatrix} H_{1_1}(s) \\ H_{1_2}(s) \end{bmatrix} = \begin{bmatrix} \frac{1}{R_s} \\ 0 \end{bmatrix} \quad (\text{J.4})$$

Solving (J.4) gives the first order transfer functions at node 1 and 2 as

$$H_{1_1}(s) = \frac{1}{Y_s(s)} \frac{1}{R_s} \quad (\text{J.5})$$

$$H_{1_2}(s) = \frac{-g_m}{Y_s(s)Y_L(s)} \frac{1}{R_s} \quad (\text{J.6})$$

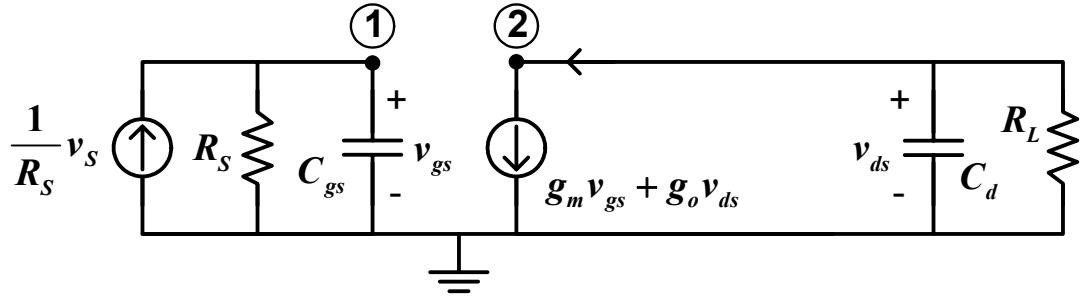


Fig. J.2 The linearized equivalent circuit for solving first order kernels.

J.2 Second order kernels

The second order kernels are calculated from the response of the linearized circuit to the second order virtual nonlinear current source, i_{NL2} as shown in Fig. J.3. The

virtual excitation $iNL2$ is placed in parallel with the corresponding linearized element, and is the only excitation applied to the circuit when calculating second order kernels. The external excitation V_s is grounded. Denoting the second order kernels at node 1 and 2 as $H_{2_1}(s_1, s_2)$ and $H_{2_2}(s_1, s_2)$, these transfer functions can be solved from the matrix equation as

$$\begin{bmatrix} Y_s(s_1 + s_2) & 0 \\ g_m & Y_L(s_1 + s_2) \end{bmatrix} \begin{bmatrix} H_{2_1}(s_1, s_2) \\ H_{2_2}(s_1, s_2) \end{bmatrix} = \begin{bmatrix} 0 \\ -iNL2 \end{bmatrix}. \quad (J.7)$$

$iNL2$ is determined by the second order coefficients in (J.1) and the first order kernels of their corresponding controlling voltages,

$$\begin{aligned} iNL2 = & K_{2g_m} H_{1_1}(s_1) H_{1_1}(s_2) \\ & + K_{2g_o} H_{1_2}(s_1) H_{1_2}(s_2) \\ & + \frac{1}{2} K_{2g_m g_o} [H_{1_1}(s_1) H_{1_2}(s_2) + H_{1_2}(s_1) H_{1_1}(s_2)] \end{aligned} \quad (J.8)$$

Solving (J.7) gives the second order kernels at node 1 and 2 as

$$H_{2_1}(s_1, s_2) = 0 \quad (J.9)$$

$$H_{2_2}(s_1, s_2) = \frac{-iNL2}{Y_L(s_1 + s_2)} \quad (J.10)$$

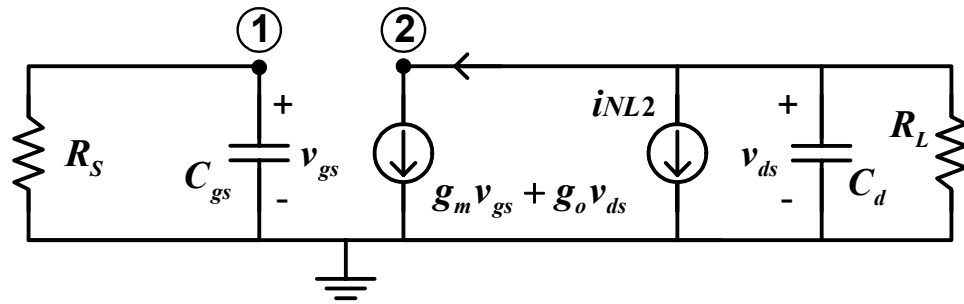


Fig. J.3 The equivalent circuit for solving the second order kernels.

J.3 Third order kernels

Similarly, the third order kernels are calculated using the equivalent circuit shown in Fig. J.4. i_{NL3} is the third order virtual nonlinear current source. Denoting the third order kernels at node 1 and 2 as $H_{3_1}(s_1, s_2, s_3)$ and $H_{3_2}(s_1, s_2, s_3)$, these transfer functions can be solved from

$$\begin{bmatrix} Y_S(s_1 + s_2 + s_3) & 0 \\ g_m & Y_L(s_1 + s_2 + s_3) \end{bmatrix} \begin{bmatrix} H_{3_1}(s_1, s_2, s_3) \\ H_{3_2}(s_1, s_2, s_3) \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL3} \end{bmatrix} \quad (J.11)$$

i_{NL3} is determined by the third order coefficients in (J.1) and the first- and second-order kernels of their corresponding controlling voltages,

$$\begin{aligned} i_{NL3} = & K_{3g_m} H_{1_1}(s_1) H_{1_1}(s_2) H_{1_1}(s_3) \\ & + \frac{2}{3} K_{2g_m} [H_{1_1}(s_1) H_{2_1}(s_2, s_3) + H_{1_1}(s_2) H_{2_1}(s_1, s_3) + H_{1_1}(s_3) H_{2_1}(s_1, s_2)] \\ & + K_{3g_o} H_{1_2}(s_1) H_{1_2}(s_2) H_{1_2}(s_3) \\ & + \frac{2}{3} K_{2g_o} [H_{1_2}(s_1) H_{2_2}(s_2, s_3) + H_{1_2}(s_2) H_{2_2}(s_1, s_3) + H_{1_2}(s_3) H_{2_2}(s_1, s_2)] \\ & + \frac{1}{3} K_{2g_m g_o} \left[\begin{aligned} & H_{1_1}(s_1) H_{2_2}(s_2, s_3) + H_{1_1}(s_2) H_{2_2}(s_1, s_3) + H_{1_1}(s_3) H_{2_2}(s_1, s_2) \\ & + H_{1_2}(s_1) H_{2_1}(s_2, s_3) + H_{1_2}(s_2) H_{2_1}(s_1, s_3) + H_{1_2}(s_3) H_{2_1}(s_1, s_2) \end{aligned} \right] \\ & + \frac{1}{3} K_{32g_m g_o} \left[\begin{aligned} & H_{1_1}(s_1) H_{1_1}(s_2) H_{1_2}(s_3) \\ & + H_{1_1}(s_1) H_{1_1}(s_3) H_{1_2}(s_2) \\ & + H_{1_1}(s_2) H_{1_1}(s_3) H_{1_2}(s_1) \end{aligned} \right] \\ & + \frac{1}{3} K_{3g_m 2g_o} \left[\begin{aligned} & H_{1_1}(s_1) H_{1_2}(s_2) H_{1_2}(s_3) \\ & + H_{1_1}(s_2) H_{1_2}(s_1) H_{1_2}(s_3) \\ & + H_{1_1}(s_3) H_{1_2}(s_1) H_{1_2}(s_2) \end{aligned} \right] \end{aligned} \quad (J.12)$$

Solving (J.11) gives the third-order kernels at node 1 and 2 as

$$H_{3_1}(s_1, s_2) = 0 \quad (\text{J.13})$$

$$H_{3_2}(s_1, s_2, s_3) = \frac{-iNL3}{Y_L(s_1 + s_2 + s_3)} \quad (\text{J.14})$$

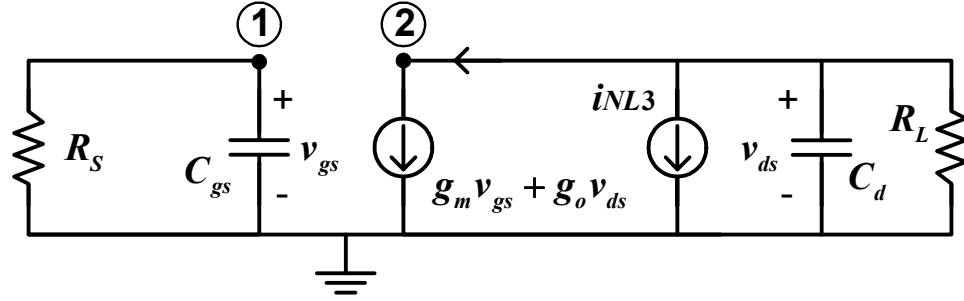


Fig. J.4 The equivalent circuit for solving the third order kernels

J.4 Input IP3

For a nonlinear system described using Volterra kernels, the amplitude of the fundamental output product is $V|H_{1_2}(j\omega_1)|$ (or $V|H_{1_2}(j\omega_2)|$), and the amplitude of the 3rd order intermodulation product is $\frac{3}{4}V^3|H_{3_2}(j\omega_1, j\omega_1, -j\omega_2)|$ (or $\frac{3}{4}V^3|H_{3_2}(-j\omega_1, j\omega_2, j\omega_2)|$), where V is the amplitude of the two-tone input signal at v_s . Then, the input IP3 (IIP3) is calculated as

$$IIP3 = \frac{1}{6R_s} \left| \frac{H_{1_2}(j\omega_1)}{H_{3_2}(j\omega_1, j\omega_1, -j\omega_2)} \right| \quad (\text{J.15})$$

where

$$H_{1_2}(j\omega_1) = \frac{-g_m}{Y_S(j\omega_1)Y_L(j\omega_1)R_s} \frac{1}{R_s} \quad (\text{J.16})$$

$$H_{3_2}(j\omega_1, j\omega_1, -j\omega_2) = \frac{-iNL3}{Y_L(2j\omega_1 - j\omega_2)} \quad (\text{J.17})$$

Substituting (J.16) and (J.17) into (J.15), we have

$$IIP3 = \frac{1}{6R_s} \left| \frac{1/R_s}{Y_S(j\omega_1)} \right| \left| \frac{g_m}{iNL3} \right|, \quad (\text{J.18})$$

since $Y_L(2j\omega_1 - j\omega_2) \approx Y_L(j\omega_1)$ for $\Delta\omega = \omega_2 - \omega_1 \ll \omega_1$.

Denoting $s_1 = j\omega_1$, $s_2 = j\omega_1$, and $s_3 = -j\omega_2$, $iNL3$ can be solved from (J.4)-(J.12).

The complete IIP3 expression for (J.18) is

$$IIP3 = \frac{1}{6R_s} \frac{1 + (\omega C_{gs} R_s)^2}{\left| \frac{K_{3g_m} + \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4}{g_m} \right|}. \quad (\text{J.19})$$

where

$$\Delta_1 = -\frac{1}{3} K_{2g_m g_o} \frac{K_{2g_m}}{g_m} Z_1 - \frac{1}{3} (K_{3g_m g_o}) Z_2,$$

$$\Delta_2 = \frac{2}{3} K_{2g_m} K_{2g_o} Z_3 + \frac{1}{3} (K_{3g_m^2 g_o}) g_m Z_4 + \frac{1}{3} (K_{2g_m g_o})^2 Z_5,$$

$$\Delta_3 = -K_{3g_o} g_m^2 Z_6 - \frac{1}{3} K_{2g_m g_o} K_{2g_o} g_m Z_7,$$

$$\Delta_4 = \frac{2}{3} (K_{2g_o})^2 g_m^2 Z_8.$$

The impedance elements (Z-elements) above are calculated as

$$Z_1 = Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2),$$

$$Z_2 = Z_L(\omega_1) \left[Y_S(-\omega_2) Y_S^{-1}(\omega_1) + 2 \right],$$

$$Z_3 = 2Z_L(\omega_1 - \omega_2) Z_L(\omega_1) + Z_L(2\omega_1) Z_L(-\omega_2),$$

$$Z_4 = Z_L^2(\omega_1) \left[2Y_S(-\omega_2) Y_S^{-1}(\omega_1) + 1 \right]$$

$$Z_5 = 2Z_L(\omega_1 - \omega_2) Z_L(-\omega_2) + Z_L(2\omega_1) Z_L(\omega_1),$$

$$Z_6 = Z_L^2(\omega_1) Z_L(-\omega_2)$$

$$Z_7 = Z_L^2(\omega_1) Z_L(2\omega_1) + 2Z_L(\omega_1) Z_L(\omega_2) Z_L(2\omega_1) + 6Z_L(\omega_1) Z_L(\omega_2) Z_L(\omega_1 - \omega_2),$$

$$Z_7 = Z_L^2(\omega_1) \left[Z_L(2\omega_1) + 2Z_L(2\omega_1) + 6Z_L(\omega_1 - \omega_2) \right],$$

$$Z_8 = Z_L^2(\omega_1) Z_L(-\omega_2) \left[Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2) \right],$$

$$\text{with } Y_L(j\omega) = \frac{1}{R_L} + g_o + j\omega C_d, \quad Z_L(\omega) = \frac{1}{Y_L(j\omega)}, \text{ and } Y_S(j\omega) = \frac{1}{R_S} + j\omega C_{gs}.$$