Characterization and Modeling of 28nm RF MOSFETs and SiGe HBTs

by

Zhen Li

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 12, 2015

Keywords: characterization, modeling, intermodulation linearity, $IP3$, $IP2$, 28nm RF CMOS, high-k/metal gate, SiGe HBT

Approved by

Guofu Niu, Chair, Professor of Electrical and Computer Engineering
Fa Foster Dai, Professor of Electrical and Computer Engineering
Stuart M. Wentworth, Associate Professor of Electrical and Computer Engineering
Bogdan M Wilamowski, Professor of Electrical and Computer Engineering

Abstract

Both the second and third order intermodulation linearities ($IP3$ and $IP2$) are characterized on 28nm RF MOSFETs. The impacts of biasing, device size, frequency, and gate length on $IP3$ are discussed. With both DC and RF tuning, the PSP core model based design kit provides reasonable fitting result, enabling comparison of $IP3$ simulation with measurement. Effective gate voltage $IP3$, a figure-of-merit for $IP3$ extracted from RF $IP3$ measurement is proposed to include important output conductance related nonlinearities and compared with traditional gate voltage $IP3$.

The real part of $Y22$ is generally fitted by adjusting body resistance, which is found to be insufficient experimentally. A new tradeoff between imaginary part of $Y22$ and real part of $Y22$ fitting is identified and used to enable fitting of both real and imaginary part of $Y22$ across a wide range of frequency and bias.

For 200 GHz HBTs, a simplified model is used to analyze characterization result of $IP2$. The impacts of biasing, device size, frequency and breakdown version on $IP2$ are discussed.

Acknowledgments

I owe my whole PhD study to my adviser, Dr. Guofu Niu, who has been encouraging, guiding and mentoring me. His spirit of excellence molds me throughout my doctoral training. His impressive knowledge and experience in many research areas enable me to search for solutions to a number of problems when I work toward this dissertation. I am grateful to Dr. Foster Dai for allowing me to use RF measurement instruments in his lab. I would like to thank Dr. Stuart M. Wentworth for his kind encouragement all the time. In particular, I appreciate the opportunity Dr. Bogdan M Wilamowski's brings to me to teach both undergraduate and graduate students on nMOSFET/pMOSFET manufacturing for such a long time. I want to thank Dr. Xiao Qin for his advice to improve this dissertation.

I would like to express my deep appreciation to group alumni, Prof. Xiaoyun Wei at HUST, as I benefit a lot from her PhD work. And it is an memorable experience to work with group mates Xiaojia Jia, Yiao Li, Rongchen Ma, Wei-Chung Shih, Jingyi Wang, Ruocan Wang, Yili Wang, Huaiyuan Zhang and Jiabi Zhang. I also would like to thank Dr. Qingqing Liang and Mr. Kimihiko Imura of MaxLinear Inc. for sharing me advanced modeling resources.

Finally, I am indebted to my parents and my wife who have provided me strong support all the time.

Table of Contents

Abstract ii

Acknowledgments iii

List of Figures vii

List of Abbreviations xiii

1 Introduction 1

1.1 Tested RF CMOS and HBT technologies 3

1.1.1 28nm HK/MG RF CMOS technology 3

1.1.2 200 GHz HBT technology 3

1.2 Intermodulation linearity 6

1.3 Outlines and contributions of this dissertation 7

2 RF measurement techniques 9

2.1 Measurement system 9

2.2 S-parameter measurement 10

2.2.1 Cabling and propbe resistances 10

2.2.2 Measurement setup 13

2.3 Intermodulation linearity measurement 15

2.3.1 Measurement setup 15

2.3.2 Extrapolation of intermodulation linearity 17

2.4 Measurement techniques 18

3 Linearity in RF CMOS 24

3.1 A new figure-of-merit $V_{GS,IP3}$ 24

3.2 V_{GS} dependence 30

3.3 V_{DS} impact 32

3.4 Impact of device size . . . 32
3.5 Impact of channel length . . . 35
3.6 Impact of frequency . . . 38
3.7 Analytic expression of $IIP3$ . . . 39
4 CMOS model calibration . . . 47
4.1 Key PSP parameters to drain current fitting . . . 47
4.2 Improved DC and Y-parameter fittings . . . 49
5 Y22 modeling . . . 54
5.1 $real(Y22)$ problem . . . 54
5.2 DNW model . . . 55
5.3 Expression of $real(Y22)$ . . . 57
5.4 V_{GS} dependence . . . 58
5.5 Improved modeling . . . 59
6 IP2 of 200 GHz SiGe HBT . . . 68
6.1 Analytical model . . . 68
6.2 SBV device . . . 70
6.3 MBV device . . . 70
6.4 HBV device . . . 72
6.5 Comparison of breakdown voltages . . . 72
6.6 Impact of device size . . . 77
6.7 Impact of frequency . . . 77
6.8 Model comparison . . . 77
7 Conclusion . . . 81
Bibliography . . . 83
A Taylor series expansion of output voltage for a two-tone input . . . 85
B Volterra Series based IP3 derivation for MOSFET . . . 87
B.1 First order kernels . . . 89

B.2 Second order kernels 90
B.3 Third order kernels 91
B.4 Expression of $IIP3$ 92
C Power calibration procedures for IP2/IP3 measurement 95
D S-parameter setup 97

List of Figures

1.1 Measured (a) I_{DS} versus V_{GS}; (b) f_T versus I_{DS} and (c) I_{DS} versus V_{DS} for the examined 28nm technology. 4

1.2 Measured output characteristics of HBTs with different breakdown voltage. I_B = 20 nA, 100 μA, 200 μA, 300 μA. 5

1.3 Measured cut-off frequency versus I_C for SBV and HBV HBTs at $V_{CE} = 1.9$ V. $A_E = 0.12\text{x}18\ \mu\text{m}^2$. 6

1.4 Illustration of spectrum output for a two-tone input. 7

2.1 Measurement system. 10

2.2 Measurement setup for cabling and contact resistances. 11

2.3 Extract of cabling and contact resistances. 12

2.4 Measured cabling and contact resistances. 13

2.5 Diagram for S-parameter measurement setup. 14

2.6 Measurement setup for $IP3$ and $IP2$. 16

2.7 Measured output spectrum with a two-tone input at 2 and 2.001 GHz. 16

2.8 Extrapolation illustration for (a) $IP3$ and (b) $IP2$. 17

2.9 Measured (a) $IIP3$ and (b) $OIP3$ from both upper and lower $IM3$ as a function of input power at 5 GHz. $V_{GS} = 0.6$ V, $V_{DS} = 0.6$ V 19

2.10 Fundamental output power $P_{out,1st}$ and second order intermodulation power $P_{out,2nd}$ versus P_{in} illustrating $IP2$ calculation for HBT. 20

2.11 Gain comparison between intermodulation linearity measurement and S-parameter measurement as a function of V_{GS}. $V_{DS} = 0.6$ V. 21

2.12 Measured (a) $IIP3$; (b) Gain and (c) $OIP3$ as a function of V_{GS} with different tone spacing . $V_{DS} = 0.8$ V. 22

2.13 Measured $IIP2$ versus I_C for different tone spacing at 5 GHz for a SBV HBT. $V_{CE} = 1.5$ V. 23

2.14 Verification of 6 dB difference between $IM2$ with $HD2$ at 5 GHz for a SBV HBT. $V_{CB} = 0.5$ V. 23

3.1 Simplified equivalent circuit used for $IP3$ derivation using Volterra series. . . . 25

3.2 Measured $IIP3$ in watt versus ω^2. $V_{GS} = 0.7$ V and $V_{DS} = 1.0$ V. 28

3.3 Measured $IIP3$ in dBm versus frequency. $V_{GS} = 0.7$ V and $V_{DS} = 1.0$ V. . . . 29

3.4 Comparison of measured (a) $IIP3$ (b) $VIP3$ (c) $V_{GS,IP3}$ and (d) $K_{3_{g_m}}$ as a function of V_{GS} with simulation at $V_{DS} = 0.6$ V. 31

3.5 Measured (a) $IIP3$; (b) $IIP2$ and (c) $V_{GS,IP3}$ and (d) $VIP3$ as a function of V_{GS} for different V_{DS}. 33

3.6 Measured $IIP3$ at 5 GHz for two 30nm devices with different total width. . . . 34

3.7 Measured $IIP3$ as a function of V_{GS} for two 30nm devices with same total width. 34

3.8 Measured (a) $IIP3$ and (b) Gain as a function of V_{GS} for three thick oxide devices with different channel length. $V_{DS} = 1.5$ V. 36

3.9 Measured $K_{2_{g_m}}$ as a function of V_{GS} for thick oxide devices with different channel length at $V_{DS} = 1.5$ V. Devices have the same total width. 36

3.10 Measured $IIP3$ as a function of I_{DS} for three thick oxide devices with different channel length. $V_{DS} = 1.5$ V. 37

3.11 Measured $IIP3$ as a function of V_{GS} for thick oxide devices with different channel length at $V_{DS} = 0.8$ V. Devices have the same total width. 38

3.12 Measured $IIP3$ as a function of (a) V_{GS} and (b) I_{DS} for two thin oxide devices with different channel length. $V_{DS} = 1.0$ V. 39

3.13 Measured (a) $IIP3$ (b) Gain and (c) $OIP3$ as a function of I_{DS} for different frequency. 40

3.14 Comparison of measured $IIP3$ with simulation and calculation by (3.11) as functions of (a) V_{GS} and (b) I_{DS}. $V_{DS} = 0.6$ V. 41

3.15 V_{GS} dependence of $\frac{K_{3_{g_m}}}{g_m} + \Delta$, $\frac{K_{3_{g_m}}}{g_m}$, $\Delta 1$, $\Delta 2$, $\Delta 3$ and $\Delta 4$. 42

3.16 V_{GS} dependence of (a) IIP3 by (3.11) and (b) $\frac{K_{3_{g_m}}}{g_m} + \Delta$ with different Δ contributions. 44

3.17 (a) $IIP3$ by (3.11) (b)$\frac{K_{3_{g_m}}}{g_m} + \Delta$ and (c) $\frac{K_{3_{g_m}}}{g_m}$ as a function of V_{GS} at $V_{DS} = 0.6$ and 1.0 V. 45

3.18 V_{GS} dependence of (a) $\frac{K_{3_{g_m}}}{g_m} + \Delta$, (b) $\frac{K_{3_{g_m}}}{g_m}$, (c) $\Delta 1$, (d) $\Delta 2$, (e) $\Delta 3$ and (f) $\Delta 4$ at $V_{DS} = 0.6$ and 1.0 V. 46

4.1 Comparison of simulated I_{DS} versus V_{GS} with measurement at $V_{DS} = 0.6$ and 1.0 V. 48

4.2 Effective regions for model parameters in affecting drain current. 50

4.3 Comparison of simulated I_{DS} versus V_{DS} with measurement at V_{GS} = 0.4, 0.6 and 0.8 V. 50

4.4 Comparison of simulated I_{DS} versus V_{GS} with measurement on (a) linear and (b) log I_{DS} scales at V_{DS} = 0.6, 0.8 and 1.0 V. 51

4.5 Comparison of simulated (a) Y11 (b) Y21 (c) Y22 and (d) Y12 versus frequency at V_{GS} = 0.6 V and V_{DS} = 0.6 V. 53

5.1 Comparison of simulated $real(Y22)$ as a function of frequency with measurement for a 30nm MOSFET at V_{GS} = 0, 0.2, 0.4, 0.6, 0.8 and 1.0 V. V_{DS} = 0.6 V. . . . 55

5.2 RF nMOSFET structure. 56

5.3 Implementation illustration of DNW model. 57

5.4 Equivalent circuit for $(Y22)$. 58

5.5 Simplified equivalent circuit used for $real(Y22)$ discussion. 59

5.6 (a) Comparison of simulated V_{GS} dependence of $\frac{real(Y22)-g_{ds}}{\omega^2}$ with measurement result at 5, 10, 20 and 30 GHz. (b)Simulated g_{mb} as a function V_{GS} using the original kit. 60

5.7 Comparisons of simulated V_{GS} dependences of $\frac{real(Y22)-g_{ds}}{\omega^2}$ and $real(Y22)$ using both the original and modified kits with measurement. 61

5.8 Comparison of simulated frequency dependence of Y-parameters using both modified and original kits with measurement. V_{GS} = 0 V. V_{DS} = 0.6 V. 62

5.9 Comparison of simulated Y-parameters using both modified and original kits with measurement. V_{GS} = 0.2 V. V_{DS} = 0.6 V. 63

5.10 Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 0.4$ V. $V_{DS} = 0.6$ V. 64

5.11 Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 0.6$ V. $V_{DS} = 0.6$ V. 65

5.12 Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 0.8$ V. $V_{DS} = 0.6$ V. 66

5.13 Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 1.0$ V. $V_{DS} = 0.6$ V. 67

6.1 Simplified small signal equivalent circuit for BJT. 69

6.2 (a) $IIP2 - I_C$ (b) $f_T - I_C$ (c) $OIP2 - I_C$ (d) $IIP3 - I_C$. Standard breakdown voltage HBT. 71

6.3 (a) $IIP2$-I_C (b) f_T-I_C (c) $IIP3$-I_C for $V_{CE} = 1.3$, 1.6, 1.9, 2.5 V. MBV HBT. 73

6.4 (a) $IIP2$-I_C (b) f_T-I_C (c) $IIP3$-I_C for $V_{CE} = 1.3$, 1.9, 2.5, 3.1 V. HBV HBT. . 74

6.5 (a) $IIP2$-I_C (b) f_T-I_C for different breakdown versions. $V_{CE} = 1.9$ V. 75

6.6 (a) $IIP2$-I_C (b) $OIP2$-I_C and (c) f_T-I_C for different breakdown versions. . . . 76

6.7 (a) $IIP2$-I_C (b) G-I_C and (c) $OIP2$-I_C. SBV HBT at $V_{CB} = 0.5$ V. 78

6.8 Measured (a) $OIP2$ and (b) Gain as a function of I_C at different frequency for a SBV HBT at $V_{CE} = 1.5$ V. 79

6.9 Comparison of measured $IIP2$ with simulation using three compact models at 5 GHz for a SBV HBT. $V_{CE} = 1.3$ V. 80

B.1 Simplified equivalent circuit used for $IP3$ derivation using Volterra series. . . . 87

B.2 Equivalent circuit to solve the first order kernels. 89

B.3 Equivalent circuit to solve the second order kernels. 90

B.4 Equivalent circuit to solve the third order kernels. 92

C.1 Front panel of HP 437B power meter. 96

D.1 Check and configure HP-IB interface. 97

D.2 Check and configure HP-IB interface. 98

D.3 Check power meter's communication with PNA. 98

D.4 Verify if a chosen input signal is appropriate. 99

D.5 Use PNA calibration settings in IC-CAP. 100

D.6 Verify SOLT calibration. 100

List of Abbreviations

RFIC radio frequency integrated circuit

CMOS complementary metal-oxide semiconductor

FET field-effect transistor

SiGe silicon-germanium

HBT heterojunction bipolar transistor

HKMG high-k/metal gate

SBV standard breakdown voltage

MBV medium breakdown voltage

HBV high breakdown voltage

IM3 third order intermodulation

IP3 third order intercept point

IIP3 input third order intercept point

OIP3 output order intercept point

IP2 second order intercept point

IIP2 input second order intercept point

OIP2 output second order intercept point

BSIM Berkeley Short-channel IGFET Model

ISS impedance standard substrate

SOLT short-open-load-thru

GSG ground-signal-ground

Chapter 1

Introduction

Modern CMOS technology scaling is no longer just a matter of shrinking physical dimensions. A key to down scale the equivalent oxide thickness (EOT) in recent technologies is the replacement of classic poly-Si gate/SiO_2 gate stack with a high-k dielectric/metal gate stack. Given the tremendous interest in scaled RF CMOS and RF system-on-chip that integrates digital and RF functions, it is necessary to examine the RF performance of the core transistors in these scaled technologies.

Intermodulation linearity is an important consideration for RF technologies. In this dissertation, both third and second order intermodulation linearities are investigated on a 28nm high-k/metal gate (HKMG) RF CMOS technology by using both measurement and simulation. We focus on $IP3$ as it is more relevant. Third order intermodulation products are close to the fundamental frequencies of interest and cannot be filtered out. Mixing of adjacent channel interferers produces undesired output in the frequency band of interest. Third order nonlinearities are also responsible for desensitization and cross-modulation.

From a gate capacitance perspective, poly depletion effect is no longer present with the use of metal gate, the change of gate-to-source capacitance C_{gs} with gate voltage is less in strong inversion, and linearity should improve compared to poly-gate transistors according to [1]. That analysis, however, assumed velocity saturation at the source, which is not the case in today's advanced CMOS. Scaling, and the associated changes in doping, effective oxide thickness, strain are all expected to change device $I-V$ characteristics as well as the various transconductance nonlinearities, output conductance nonlinearities, and cross nonlinearities.

Typical compact model parameters are extracted by fitting DC I-V curves and sometimes first order derivatives. A good fitting does not necessarily guarantee good accuracy of

higher order derivatives, which are difficult to evaluate experimentally due to the increase of numerical and experimental error in differentiation. Direct RF intermodulation measurements are therefore necessary, as shown in this dissertation together with simulations using a compact model with DC I-V and Y-parameter calibration. Details for both DC and RF tuning are included in Chapter 4.

High frequency (HF) modeling is challenging. The body resistance is generally raised to increase real part of $Y22$ ($real(Y22)$) in order to minimize the difference between simulation and measurement. In fact, the main contribution for $real(Y22)$ comes from the capacitance between drain and body in series with a parallel connection of the capacitance between drain and source and the body resistance. Therefore, the body resistance should be increased together with the capacitance between drain and body in order to improve frequency dependence of $real(Y22)$. For V_{GS} dependence of $real(Y22)$, a technique to normalize bias dependence at all frequency is needed. First, output conductance should be excluded from $real(Y22)$ as it is relatively large and makes other V_{GS}-dependent contributions to $real(Y22)$ unidentifiable. Second, frequency itself should be taken out to expose the real V_{GS} dependence.

For HBTs, there is a number of literature available to discuss third order intermodulation linearity as third order output is close to the fundamentals [2]. Second order intermodulation linearity has received much less attention, in part because second order intermodulation products are far away from the fundamentals, and may be filtered [3]. While differential circuits can decrease second order nonlinearity, cancellation is not complete, and is worsened by transistor mismatch. The same second order nonlinearity also produces dc offset, which presents a significant problem for direct-conversion receivers [4]. Therefore, RF characterization of second order intermodulation linearity is important to provide $IP2$ design guidelines collector current, collector-emitter voltage, transistor size and breakdown voltage version.

In this chapter, two tested technologies are introduced with DC and cut-off frequency characteristics. $I_{DS}-V_{GS}$, $I_{DS}-V_{DS}$ and cut-off frequency versus I_{DS} are presented for 28nm

HK/MG RF CMOS technology. For 200 GHz HBT technology, maximum allowable collector-emitter voltage is obtained from output characteristics for different breakdown versions. Cut-off frequency for different breakdown versions is compared as a function of I_C. With a simple power series, the second and third intermodulation products are introduced. At last, outlines and contributions of this dissertation are summarized.

1.1 Tested RF CMOS and HBT technologies

1.1.1 28nm HK/MG RF CMOS technology

Figure 1.1(a) shows typical $I_{DS}-V_{GS}$ characteristics of a 30nm device from the examined 28nm technology. Figure 1.1(b) shows measured cut-off frequency f_T as a function of I_{DS}. A 304 GHz peak f_T is reached at 0.45 mA/ μm at $V_{DS} = 1.05$ V. Figure 1.1(c) shows typical $I_{DS} - V_{DS}$ characteristics.

1.1.2 200 GHz HBT technology

Devices of three breakdown voltages are measured, including standard breakdown voltage (SBV) HBT, medium breakdown voltage (MBV) HBT, and high breakdown voltage (HBV) HBT. Higher breakdown voltage usually means lower doping concentration. Figure 1.2 shows the output characteristics in an effort to avoid making devices operate in breakdown region. All three HBTs have the same emitter area of 0.12x18 μm^2 and are driven by the same base current. Maximum collector-emitter voltage (V_{CE}) is chosen as 1.9, 2.5 and 3.1 V for SBV, MBV and HBV devices.

Figure 1.3 shows measured current dependence of cut-off frequency for SBV and HBV HBTs with $A_E = 0.12\text{x}18$ μm^2. $V_{CE} = 1.9$ V. For simplicity, cut-off frequency is related to I_C by

$$\frac{1}{2\pi f_T} \approx \tau_F + \frac{\phi_T A_E}{I_C} C_t \tag{1.1}$$

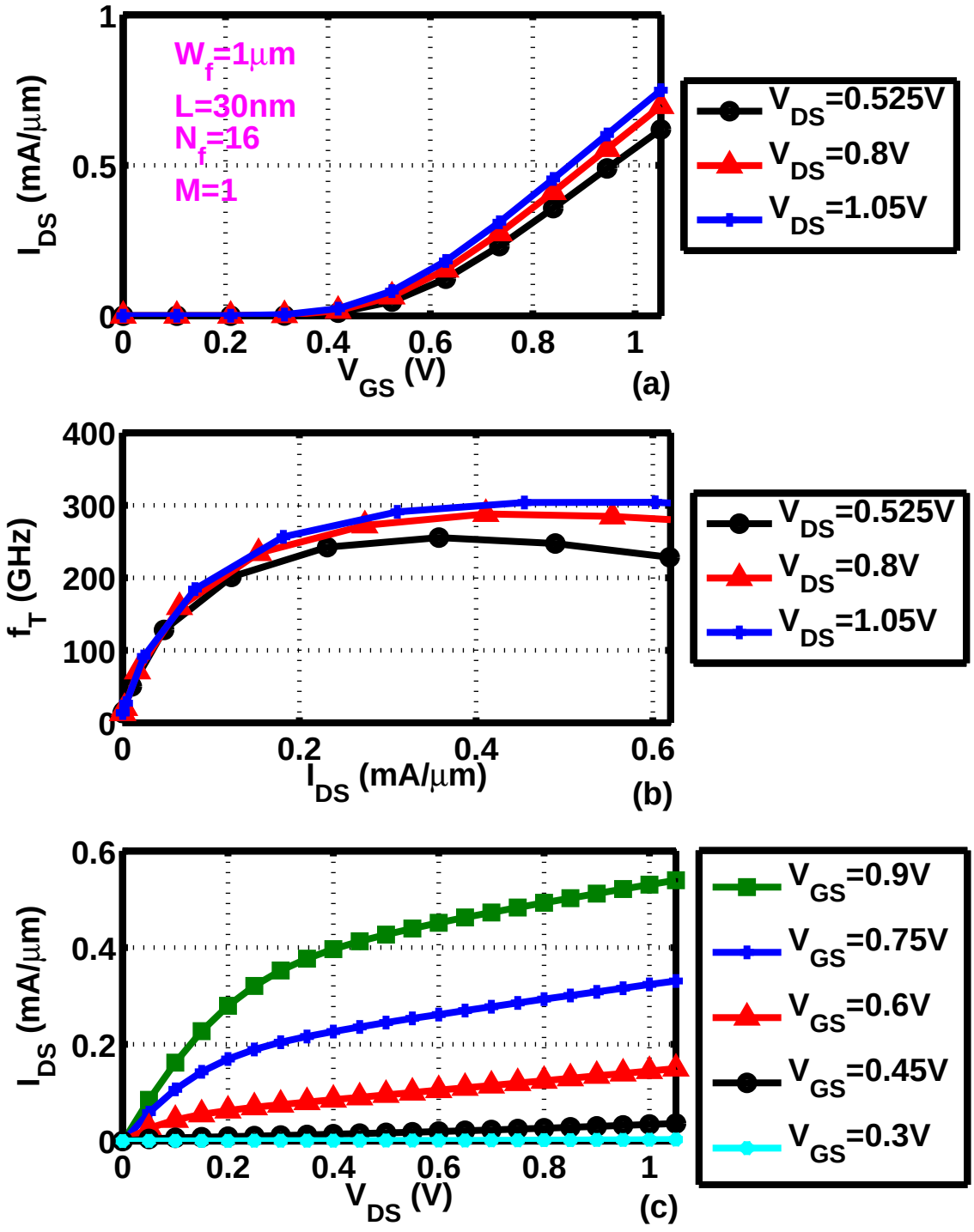

Figure 1.1: Measured (a) I_{DS} versus V_{GS}; (b) f_T versus I_{DS} and (c) I_{DS} versus V_{DS} for the examined 28nm technology.

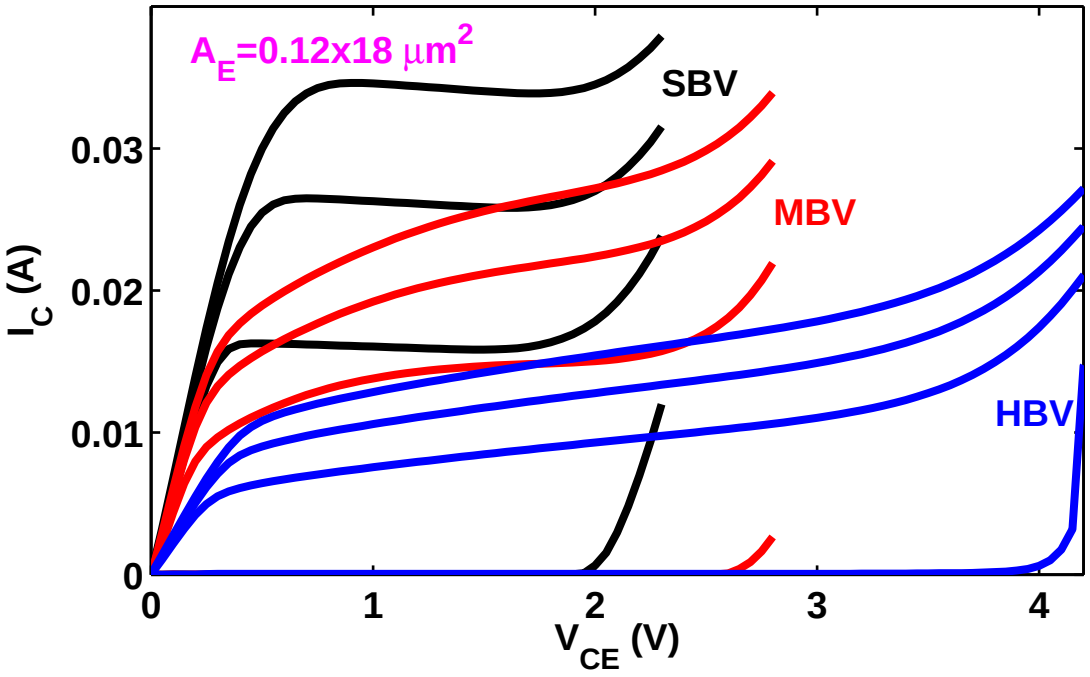

Figure 1.2: Measured output characteristics of HBTs with different breakdown voltage. I_B = 20 nA, 100 μA, 200 μA, 300 μA.

where τ_F is the forwarding transit time, $\phi_T = \frac{kT}{q}$ is the thermal voltage and C_t is the total junction depletion capacitance [5].

Prior to high injection, τ_F and C_t are constant for typical RF applications. So f_T increases with I_C both SBV and HBT HBTs. This explains why high bias is favored by circuit designers to make use of high performance (speed). However, at sufficiently large bias, base push-out (or Kirk effect) happens due to partial replacement of the collector base depletion region by neutral base region. This results in the increase of effective width of the base region, which in turn increases the forward transit time and decreases f_T substantially. The current density for the onset of Kirk effect ($J_{C,Kirk}$) can be expressed as [6]

$$J_{C,Kirk} = qv_{sat}N_{DC}[1 + \frac{2\epsilon(V_{CB} + \phi_{bi})}{qN_{DC}W_C^2}] \tag{1.2}$$

where v_{sat} is the electron saturation velocity, W_C is collector thickness, N_{DC} is uniform collector doping concentration, and ϕ_{bi} represents the built-in potential of the collector base

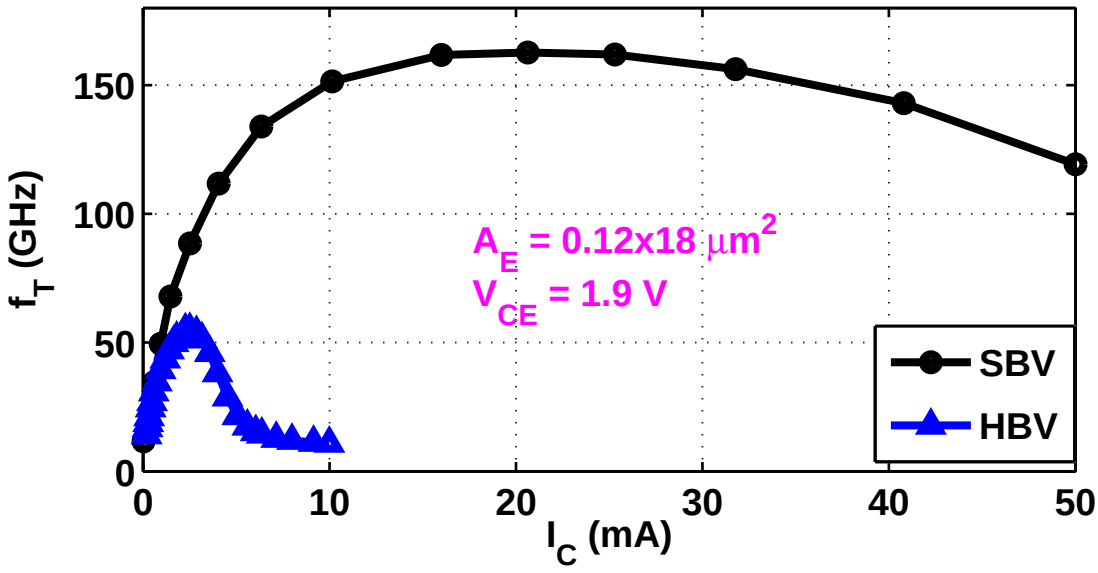

Figure 1.3: Measured cut-off frequency versus I_C for SBV and HBV HBTs at $V_{CE} = 1.9$ V. $A_E = 0.12\text{x}18\ \mu\text{m}^2$.

depletion region. SBV HBT has a high collector doping concentration and therefore allows high current operation.

1.2 Intermodulation linearity

For a two-tone input with the same amplitude (A), $v_{in}(t) = Acos(\omega_1 t) + Acos(\omega_2 t)$, output v_{out} can be expressed as:

$$
\begin{aligned}
v_{out} \;=\; & a_2A^2 + && DC\ component \\
& +(a_1A + \frac{9}{4}a_3A^3)cos(\omega_1 t) + (a_1A + \frac{9}{4}a_3A^3)cos(\omega_2 t) && Fundamental \\
& +\frac{1}{2}a_2A^2cos(2\omega_1 t) + \frac{1}{2}a_2A^2cos(2\omega_2 t) && HD2 \\
& +a_2A^2cos((\omega_2 - \omega_1)t) + a_2A^2cos((\omega_2 + \omega_1)t) && IM2 \\
& +\frac{1}{4}a_3A^3cos(3\omega_1 t) + \frac{1}{4}a_3A^3cos(3\omega_2 t) && HD3 \\
& +\frac{3}{4}a_3A^3cos((2\omega_1 - \omega_2)t) + \frac{3}{4}a_3A^3cos((2\omega_1 + \omega_2)t) && IM3 \\
& +\frac{3}{4}a_3A^3cos((2\omega_2 - \omega_1)t) + \frac{3}{4}a_3A^3cos((2\omega_2 + \omega_1)t) && IM3 \qquad (1.3)
\end{aligned}
$$

where second and third harmonic distortion products ($HD2$ and $HD3$) and second and third intermodulation products ($IM2$ and $IM3$) are produced at $2\omega_1/2\omega_2$, $3\omega_1/3\omega_2$, $(\omega_2-\omega_1)/(\omega_2+\omega_1)$, and $(2\omega_1-\omega_2)/(2\omega_2-\omega_1)$, correspondingly. More details are shown in Appendix A. As shown in Figure 1.4, third order intermodulation linearity is typically considered, as third order intermodulation products ($2\omega_1-\omega_2$ and $2\omega_2-\omega_1$) are close to the fundamentals while second order intermodulation products ($\omega_2-\omega_1$ and $\omega_2+\omega_1$) are far away from the fundamentals.

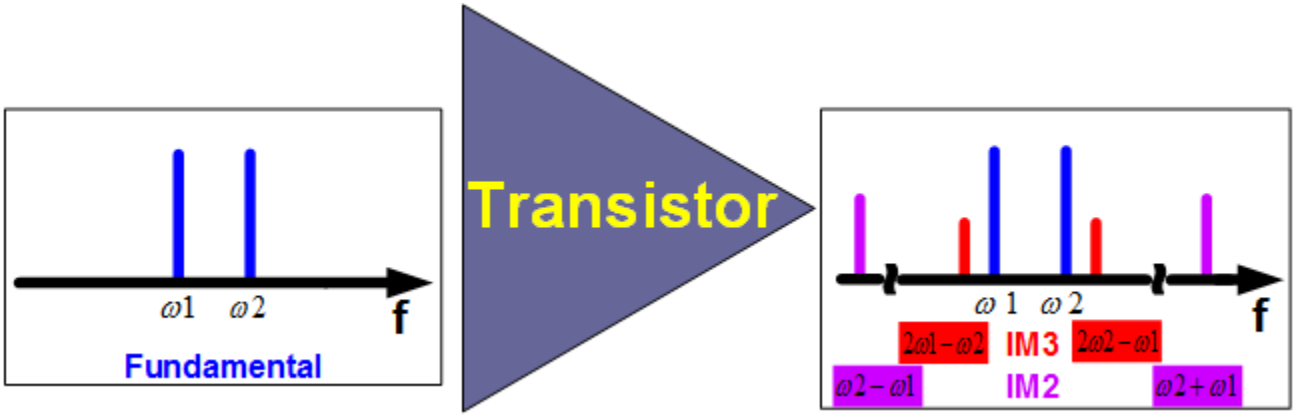

Figure 1.4: Illustration of spectrum output for a two-tone input.

1.3 Outlines and contributions of this dissertation

This dissertation focuses on the characterization and modeling of intermodulation linearity for both RF MOSFETs and SiGe HBTs. In addition, design kit calibration and modeling of $real(Y22)$ are investigated for RF MOSFETs.

In Chapter 1, 28nm high-k/metal gate (HKMG) RF CMOS and 200 GHz HBT technologies tested are introduced.

In Chapter 2, S-parameter and intermodulation linearity measurements are explained with helpful measurement techniques. Extraction of the important DC loss inside S-parameter measurement system is illustrated with measurement data.

In Chapter 3, both the second and third order intermodulation nonlinearities ($IP3$ and $IP2$) are characterized on 28nm RF MOSFETs. With a focus on $IP3$, the impacts of drain current, drain voltage, device size, frequency, and gate length on $IP3$ are discussed. Volterra series is used to derive an analytical expression of $IIP3$, which is proved close to simulation and measurement result. Based on this analytical expression, the effective gate voltage $IP3$, a figure-of-merit for $IP3$ extracted from RF $IP3$ measurement is proposed to include important output conductance related nonlinearities and compared with traditional gate voltage $IP3$.

In Chapter 4, The original design kit for $IP3$ simulation is not customized for tested devices. Five parameters in the core PSP model are identified to be effective to model drain current at different V_{GS}. With both DC and RF tuning, the kit is improved and able to predict measured $IP3$.

In Chapter 5, modeling of $real(Y22)$ is discussed. $real(Y22)$ is generally fitted by adjusting body resistance, which is found to be insufficient experimentally. A new tradeoff between imaginary part of $Y22$ and real part of $Y22$ fitting is identified and used to enable fitting of both real and imaginary part of $Y22$ across a wide range of frequency and bias.

In Chapter 6, characterization result of second order intermodulation linearity on SiGe HBTs is presented. An $IP2$ model is proposed to explain measured impacts of collector current, device size, frequency and breakdown version on $IP2$. Design guidelines for high $IP2$ are provided.

In Chapter 7, major conclusions of this dissertation are summarized.

Chapter 2

RF measurement techniques

RF modeling heavily relies on the accuracy of measurement data as extrinsic components play an important roles at high frequency and high bias applications. In addition, special de-embedding and calibration procedures are involved in DC, S-parameter and intermodulation linearity measurement.

In this chapter, on-wafer S-parameter and intermodulation linearity measurements are discussed. Extraction of cabling and contact resistances is introduced with an example on 40 GHz RF cables and probes. In addition, measurement setup and procedures of S-parameter are discussed. Measurement system for intermodulation linearity is explained with useful techniques included.

2.1 Measurement system

The complete measurement system is controlled by Matlab programs using GPIB bus. As shown in Figure 2.1, the system is mainly consisted of a pair of Agilent E8257D PSG analog signal generators, an Alessi REL-4300 thermal probe station, an HP 437B power meter (PM), an Agilent 8563EC PSA series spectrum analyzer, an Agilent E8364B PNA series microwave network analyzer, an Agilent 6625A as DC power supply for intermodulation linearity measurement, and HP 4155C as DC power supply for S-parameter measurement.

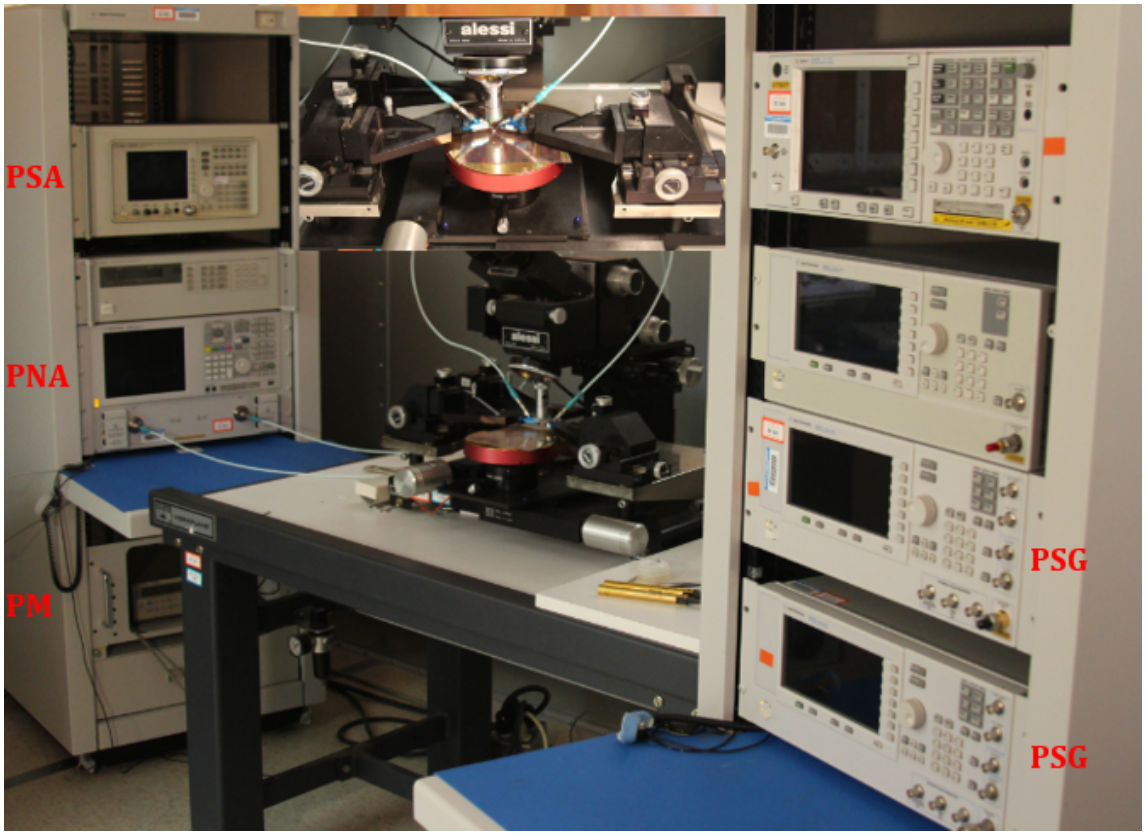

Figure 2.1: Measurement system.

2.2 S-parameter measurement

2.2.1 Cabling and propbe resistances

RF losses due to cabling and connectors are calibrated on network analyzer. However, the ohmic loss can not be excluded and stays within the measurement system. This leads to an inaccurate larger DC voltage than actual value DC bias in S-parameter simulation, resulting in incorrect RF modeling result. So overall ohmic loss on cables, bias tee, connector and contact of the S-parameter measurement system is suggested to be accurately measured and recorded for device modeling [7].

As shown in Figure 2.2, coaxial cables, connectors, BNC cables, bias tee sets in network analyzer, and RF cables all result in cabling resistance. In addition, contact resistance exists

due to non-ideal contact between RF probe and metal pad. In following discussion, cabling and contact resistances are denoted as $(R_{cable} + R_{biasT})$ and $R_{contact}$, correspondingly.

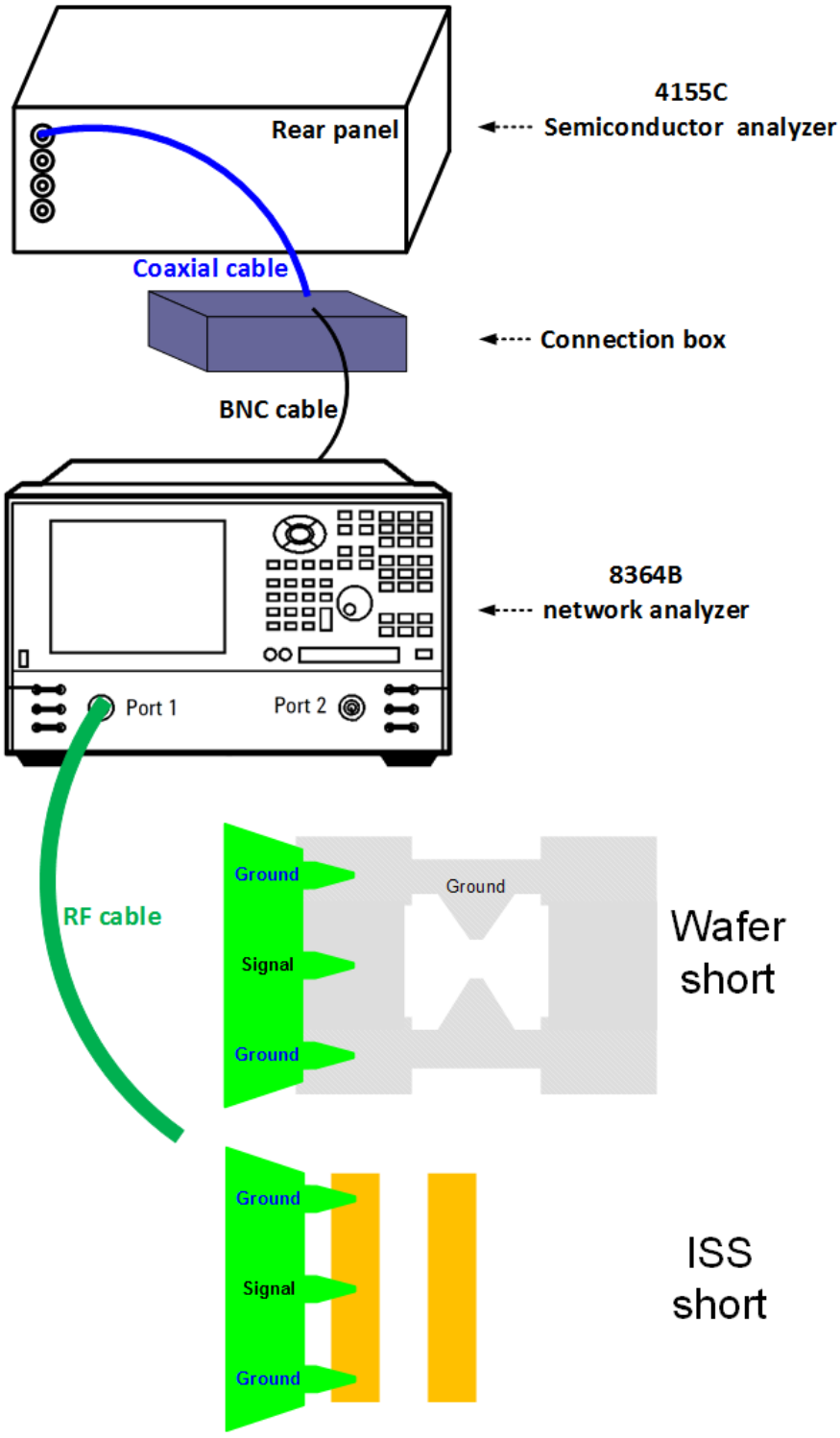

Figure 2.2: Measurement setup for cabling and contact resistances.

In order to extract the total ohmic loss, a short de-embedding structure or any on-wafer metal line should be probed by using a ground-signal-ground (GSG) RF probe. In this case, total resistance is $(R_{cable} + R_{biasT} + 1.5R_{contact})$ as $R_{contact}$ on signal path (S) is in series with two $R_{contact}$ in parallel on two ground (G) paths. If one needs to distinguish contact resistance from total resistance, the short standard on impedance standard substrate (ISS) can be probed. Assuming an ideal contact on the gold short standard, the resistance network can be simplified to $(R_{cable}+R_{biasT})$ as shown in Figure 2.3. The resistance difference between two times of probing is $1.5R_{contact}$.

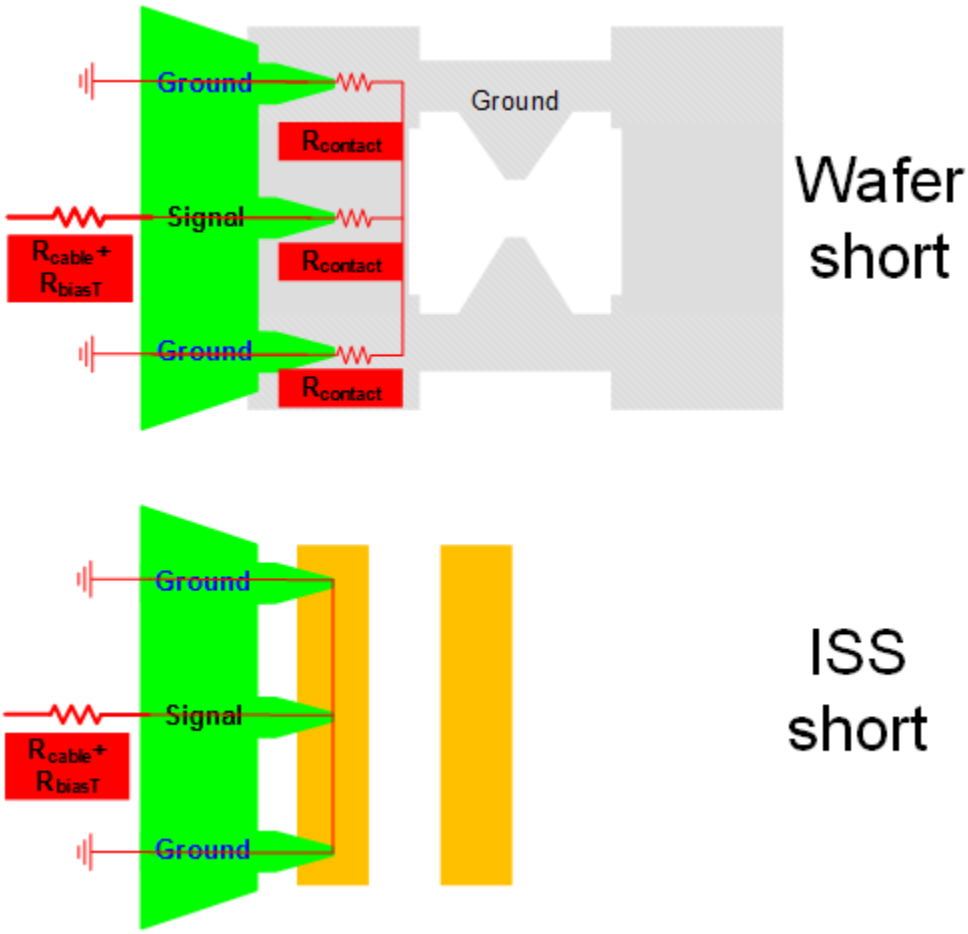

Figure 2.3: Extract of cabling and contact resistances.

Figure 2.4 shows an example of measured cabling and contact resistances by using an Infinity RF probe from Cascade Microtech. With measuring the output voltage for a sweeping current input from 5 to 25 mA, $R_{cable} + R_{biasT}$ is extracted when probing an

ISS short standard. On average, $R_{cable} + R_{biasT}$ is 1.38 Ω. Replacing probing with short de-embedding structure, average $R_{cable} + R_{biasT} + R_{contact}$ is obtained as 1.47 Ω. Therefore, $R_{contact}$ is about 80 mΩ.

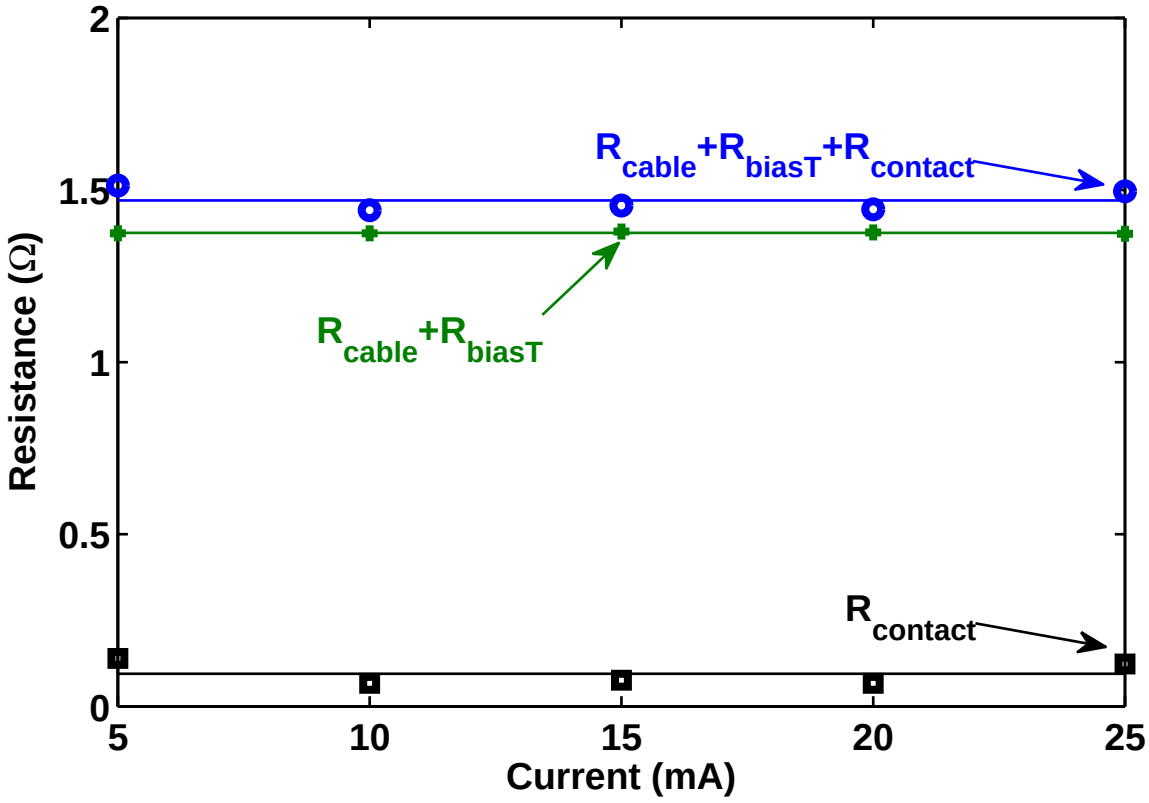

Figure 2.4: Measured cabling and contact resistances.

Total cabling and contact resistances ($R_{cable} + R_{biasT} + R_{contact}$) can be modeled by a test circuit including these resistances parallel with an ideal capacitor should be created to make sure these resistances do not affect S-parameter simulation result [8]. This makes simulation setup is the same as the actual measurement environment.

2.2.2 Measurement setup

Figure 2.5 shows the diagram for S-parameter measurement system using internal bias tee sets of the Agilent E8364B PNA Series microwave network analyzer, which is valid to measure up to 50GHz with 2.4 mm input. The internal bias tee inputs using female BNC

connectors with maximum voltage of 40 V and maximum current of 200 mA. So DC biasing for S-parameter measurement is limited below 200 mA. Detailed procedures for S-paremeter measurement are given in Appendix D.

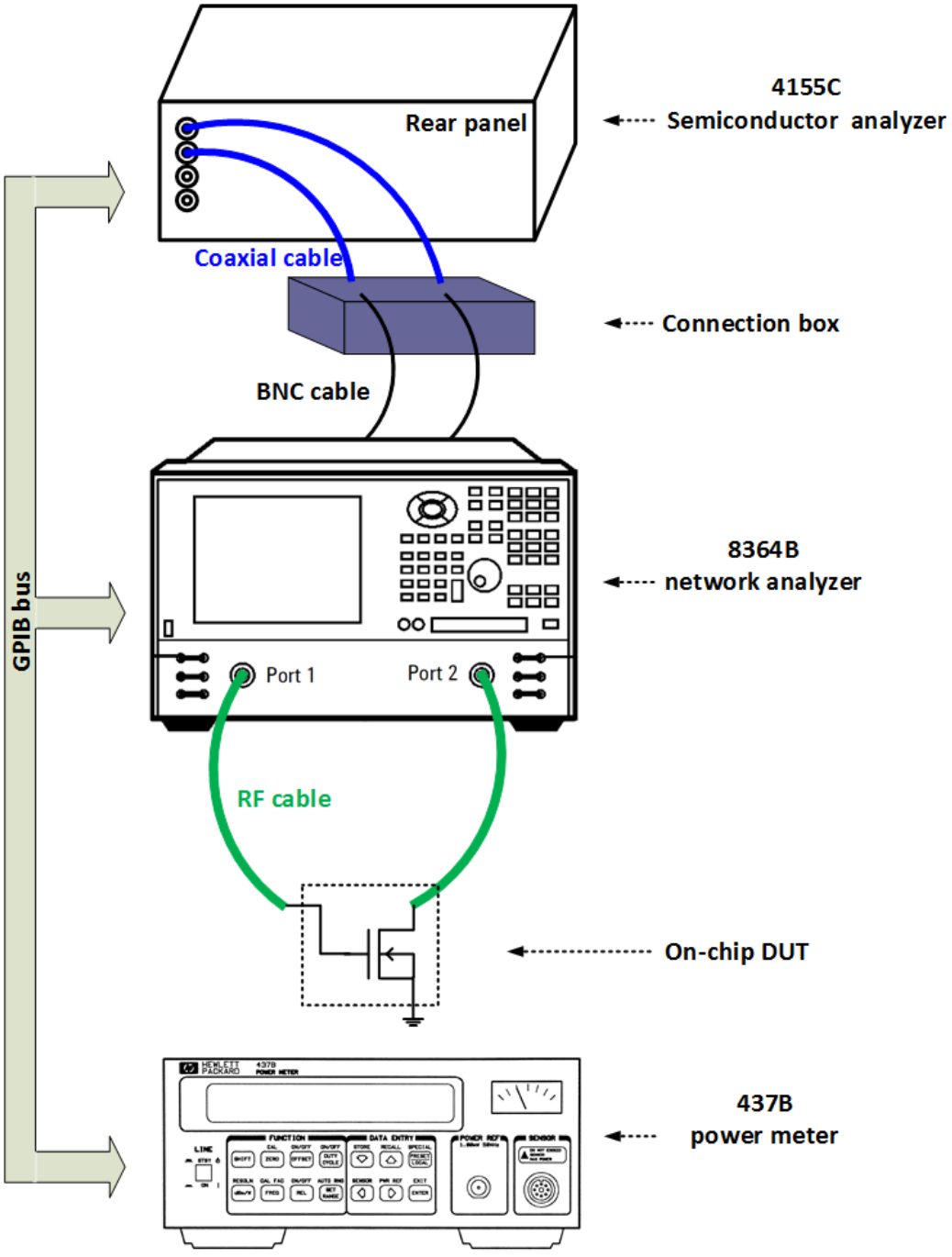

Figure 2.5: Diagram for S-parameter measurement setup.

2.3 Intermodulation linearity measurement

2.3.1 Measurement setup

Figure 2.6 shows the experimental setup used, which is similar to the setup in [9]. Broadband 50 Ω terminations are used in this work so that they do not filter out the second order harmonics which may remix with the fundamental output to produce third order intermodulation ($IM3$). Devices are probed on-wafer using Cascade Infinity GSG probes. Two Agilent signal sources are synchronized and combined using a power combiner to produce a two tone input. Attenuators are used to reduce the intermodulation within the sources. The automatic level control (ALC) function in the sources is turned off to minimize intermodulation generated by the sources. An HP-6625 power supply is used to provide precision DC biases. A spectrum analyzer is used to measure the output spectrum. Power meters are used for calibration of power loss on cables and probes.

As the intermodulation products are as small as -90 dBm, it is critical to optimize spectrum analyzer settings such as input attenuation and filter bandwidth to increase sensitivity of the analyzer, and at the same time to minimize the intermodulation distortion of the mixers inside the analyzer, which are conflicting goals [10]. The settings need to be different or optimized for different input power, and vary with device and bias. Analyzer setting is optimized for each measurement to minimize analyzer $IM3$ and maximize signal to noise ratio. For each bias point and frequency, the input power is swept and the third order intercept is obtained by extrapolation. The analyzer setting is optimized dynamically for each input power level. The measurement system intermodulation is verified to be well below the intermodulation from the device under test. The upper $(2f2 - f1)$ and lower $(2f1 - f2)$ $IM3$ are the same in our measurements.

Fig. 2.7 shows intermodulation measurement result at 2 GHz with a 1 MHz tone spacing for a 30nm MOSFET at $V_{GS} = 0.46$ V and $V_{DS} = 0.6$ V. Device total width W_{total} is 256μm. Gate finger width W_f is 1μm, number of finger N_f is 16, and multiplicity $M = 16$. Output

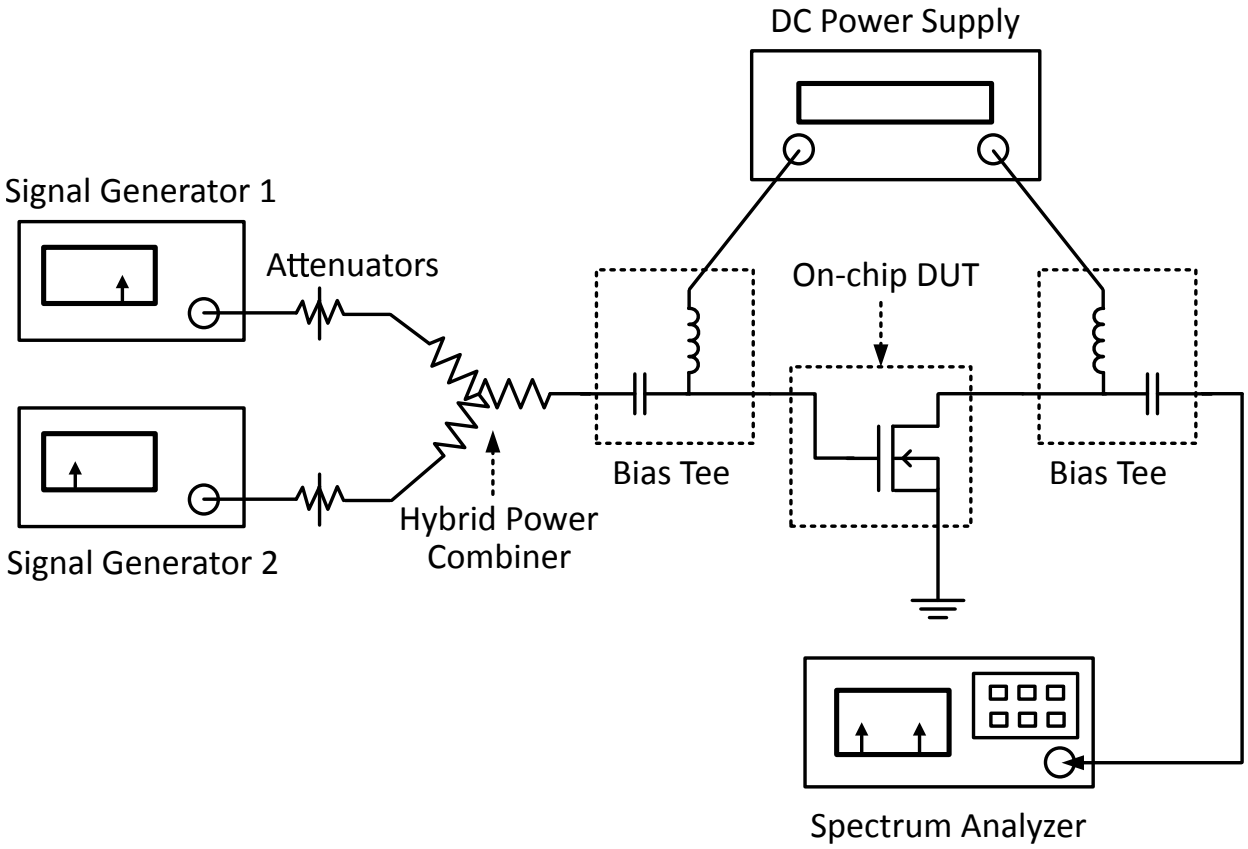

Figure 2.6: Measurement setup for $IP3$ and $IP2$.

power is measured at 2.000, 2.001, 1.999, and 2.002 GHz, which are corresponding to first order ($f1$, $f2$) output and third order intermodulation ($2f1$-$f2$, $2f2$-$f1$) output levels.

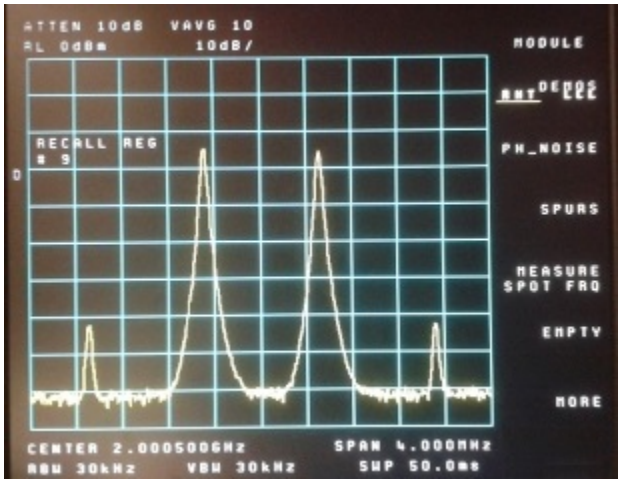

Figure 2.7: Measured output spectrum with a two-tone input at 2 and 2.001 GHz.

2.3.2 Extrapolation of intermodulation linearity

Figure 2.8(a) and (b) illustrates how $IP3$ and $IP2$ are determined for a 30nm MOSFET biased at $V_{GS} = 0.44$ V, $V_{DS} = 0.6$ V. Device total width is 256μm. Gate finger width W_f is 1μm, number of finger N_f is 16, and multiplicity $M = 16$. At low P_{in}, first order output $P_{out,1st}$ increases linearly with P_{in} at a slope of 1:1, while the third and the second order intermodulation output ($P_{out,3rd}$ and $P_{out,2nd}$) increase at slopes of 3:1 and 2:1, respectively. $IP3$ is obtained as the extrapolated intercept of $P_{out,1st}$ and $P_{out,3rd}$ in a region of P_{in} where the ideal slopes are observed. The input and output powers at $IP3$ are denoted as $IIP3$ and $OIP3$.

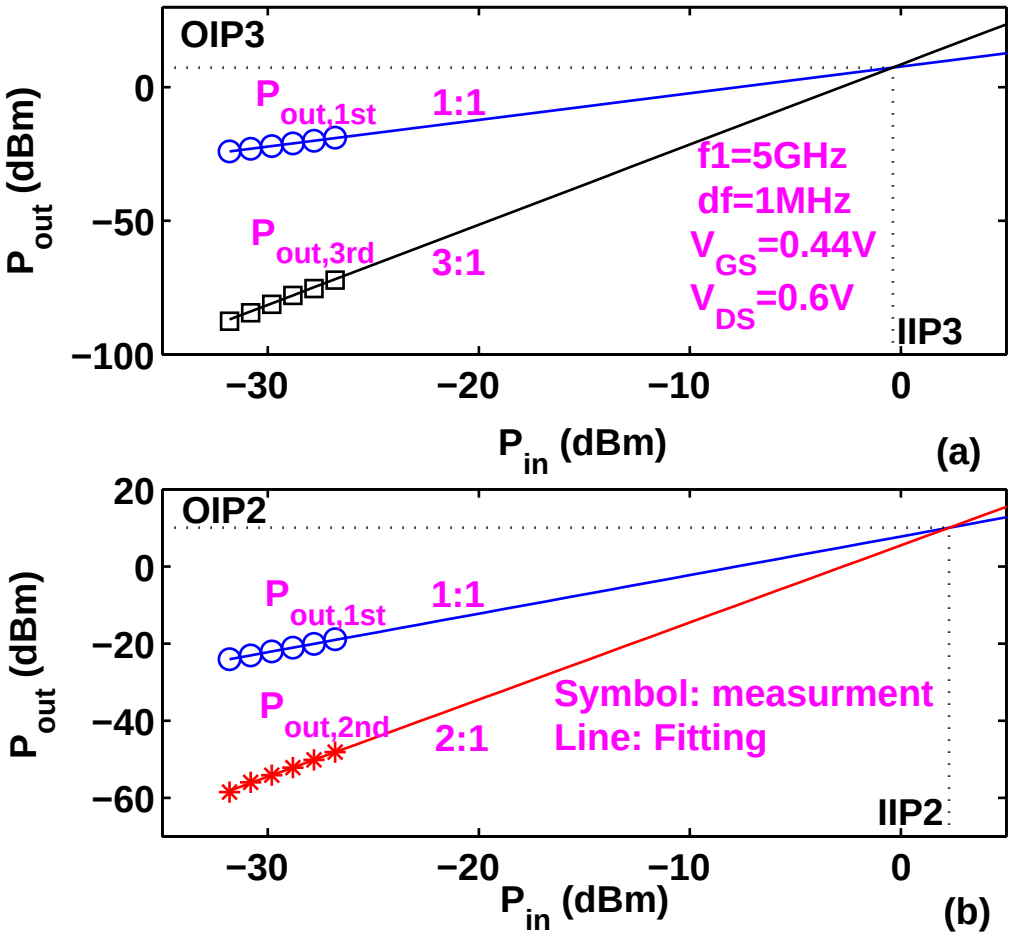

Figure 2.8: Extrapolation illustration for (a) $IP3$ and (b) $IP2$.

(2.1) and (2.2) show how $IIP3$ and $OIP3$ are obtained from measured $P_{out,1st}$, $P_{out,3rd}$ and P_{in} in dBm [11].

$$IIP3 = \frac{P_{out,1st} - P_{out,3rd}}{2} + P_{in} \tag{2.1}$$

and

$$OIP3 = \frac{3P_{out,1st} - P_{out,3rd}}{2} \tag{2.2}$$

Figure 2.9(a) and (b) show measured $IIP3$ and $OIP3$ determined by both upper and lower $IM3$ as a function of input power. For both $IIP3$ and $OIP3$, the difference between upper and lower $IM3$ methods is less than 1 dB, which is our measurement tolerance.

Similarly, we can obtain $IIP2$ and $OIP2$ from the extrapolation intercept of $P_{out,1st}$ and $P_{out,2nd}$. Figure 2.10 is an example of measuring $IIP2$ for an HBT of standard breakdown voltage (SBV) with an emitter area (A_E) of 0.12x18μm^2. P_{in} is swept from -37 dBm to -5 dBm.

To simulate $IP3$, quasi periodic steady state (QPSS) analysis is used in Cadence SpectreRF to calculate two-tone large signal behavior [12]. For each bias point, a series of input power level is swept. The output is plotted using ipnVRI function to ensure the extrapolation point for $IP3$ is within the linear range, in the same manner $IP3$ is determined in measurement illustrated earlier in Figure 2.8.

2.4 Measurement techniques

Due to the extrapolation nature for $IP3/IP2$ measurement, small error of output power makes a big difference for ultimate result. In addition, spectrum analyzer itself needs to be calibrated and optimized. To improve linearity measurement accuracy, several strategies are implemented in this work. Some examples are given below.

The difference between $IIP3$ and $OIP3$ is gain. Forward voltage gain $S21$ from S-parameter measurement should be close to the gain from intermodulation linearity measurement, which is used in this work as a reference to check power calibration and therefor

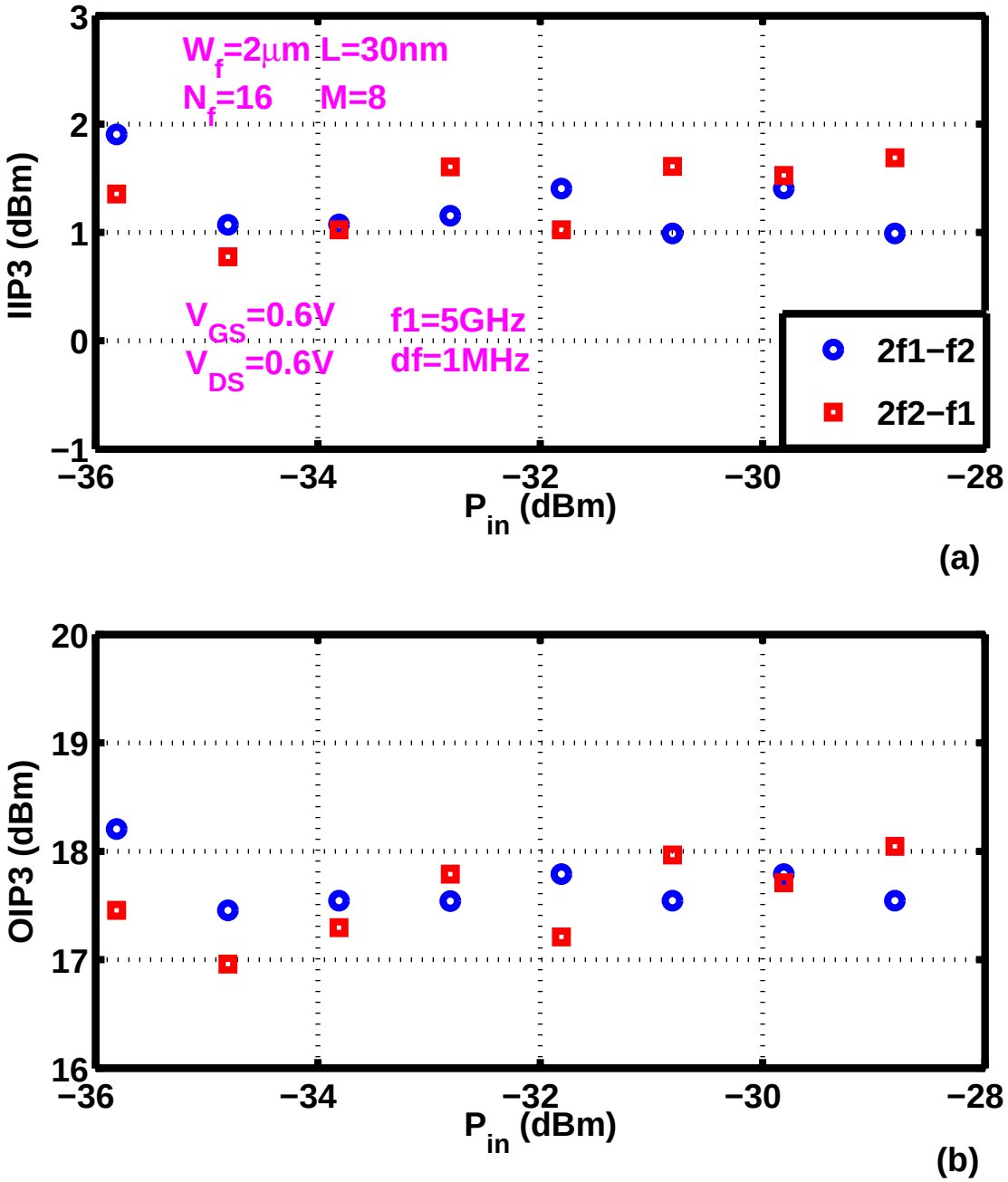

Figure 2.9: Measured (a) $IIP3$ and (b) $OIP3$ from both upper and lower $IM3$ as a function of input power at 5 GHz. $V_{GS} = 0.6$ V, $V_{DS} = 0.6$ V

measurement accuracy. Figure 2.11 compares measured gain with $S21$ for a 30nm MOSFET

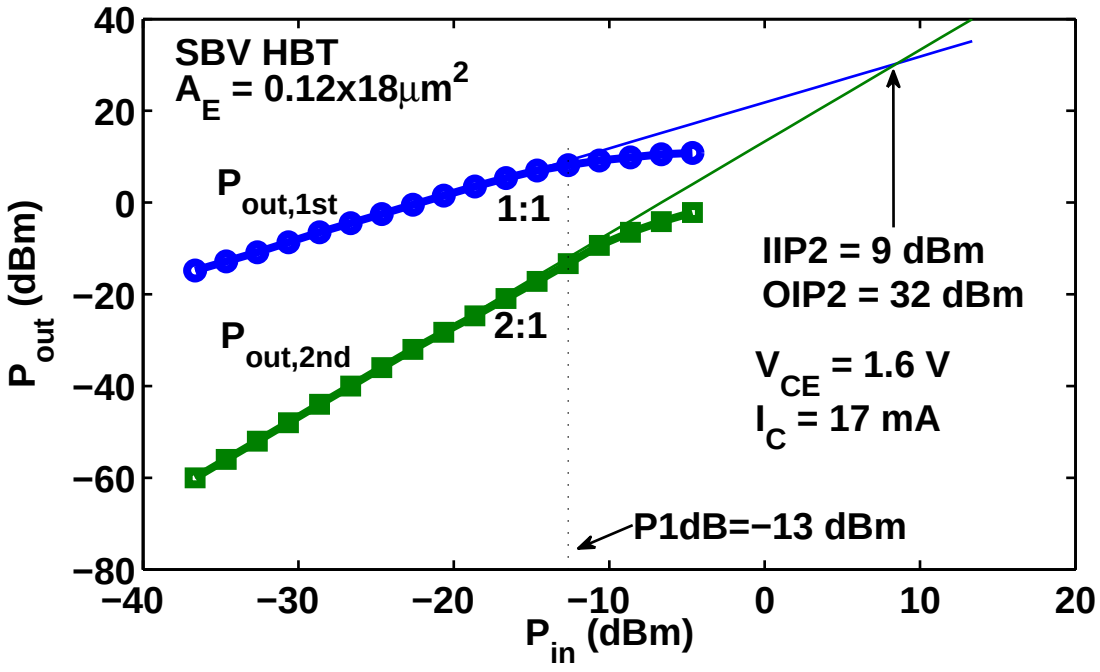

Figure 2.10: Fundamental output power $P_{out,1st}$ and second order intermodulation power $P_{out,2nd}$ versus P_{in} illustrating $IP2$ calculation for HBT.

at V_{DS} = 0.6 V. Device total width is 256μm. Gate finger width W_f is 4μm, number of finger N_f is 16, and multiplicity $M = 4$.

Figure 2.12(a), (b) and (c) show measured $IIP3$, gain and $OIP3$ at 5 GHz using different tone spacing ($df = f2 - f1$) for a 30nm MOSFET at V_{DS} = 0.8 V. Device total width is 256μm. Gate finger width W_f is 1μm, number of finger N_f is 16, and multiplicity $M = 16$. As expected, tone spacing does not affect $IIP3$ measurement result.

Figure 2.13 shows measured $IIP2$ for a SBV HBT with different tone spacing at 5 GHz. V_{CE} = 1.5 V. Emitter area (A_E) is 0.12x3μm^2. Different tone spacing gives the same $IIP2$. Therefore it is valid to use 1 MHz as tone spacing to discuss intermodulation linearity in next few chapters.

As illustrated in Figure 1.4, second order intermodulation products are far away from fundamental output power. Take 5 GHz with a 1 MHz tone spacing as an example, after searching for peak amplitude at 5 ($f1$) and 5.001 GHz ($f2$), spectrum analyzer needs to

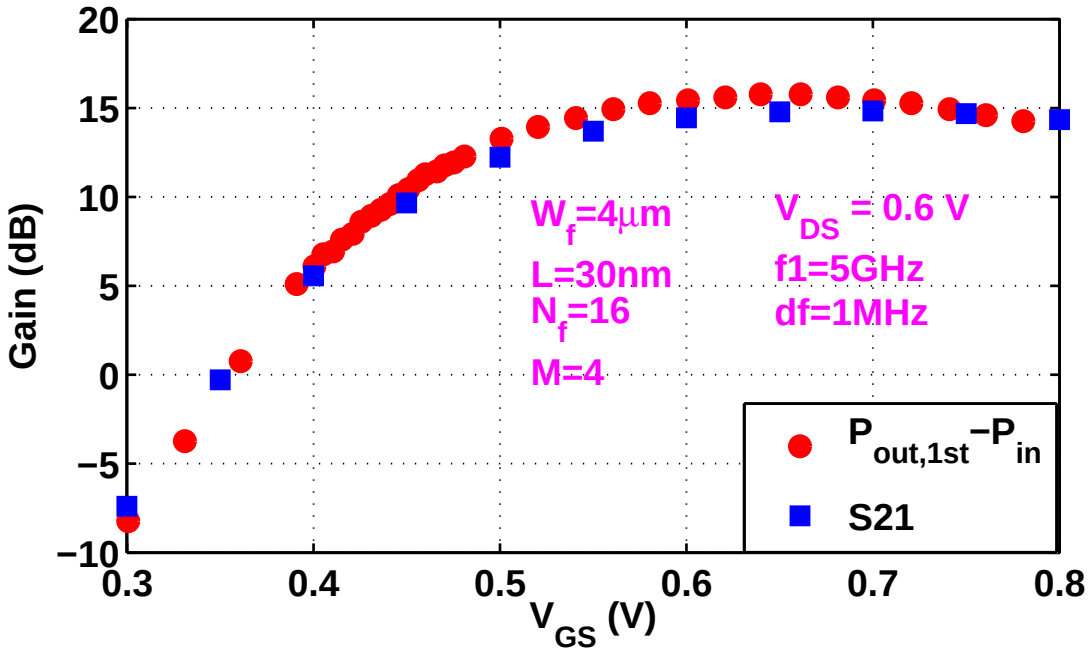

Figure 2.11: Gain comparison between intermodulation linearity measurement and S-parameter measurement as a function of V_{GS}. $V_{DS} = 0.6$ V.

change "frequency" and "span" setup twice to measure output power levels at 1 MHz $(f2 - f1)$ and 10.1 GHz $(f2 + f1)$. In addition, "reference level" is optimized by adjusting internal attenuator for each output power to minimize resolution error. So second order intermodulation measurement takes much longer time than third order intermodulation. In this work, even with known bias information for peak $IP2$, it takes three hours or so to finish one second order intermodulation measurement for one V_{CE}. So it is critical to implement real-time monitoring technique to terminate measurement when noticeable errors show up. Here 6 dB is used to verify the difference between second order intermodulation product $(2f2 - f1)$ and second order harmonics $(2f1$ and $2f2)$ [13]. The derivation is included in Appendix A following [14]. Figure 2.14 shows measured difference between second order harmonics and second order intermodulation products plus 6 dB. Device is the same in Figure 2.10. V_{CB} is fixed as 0.5 V.

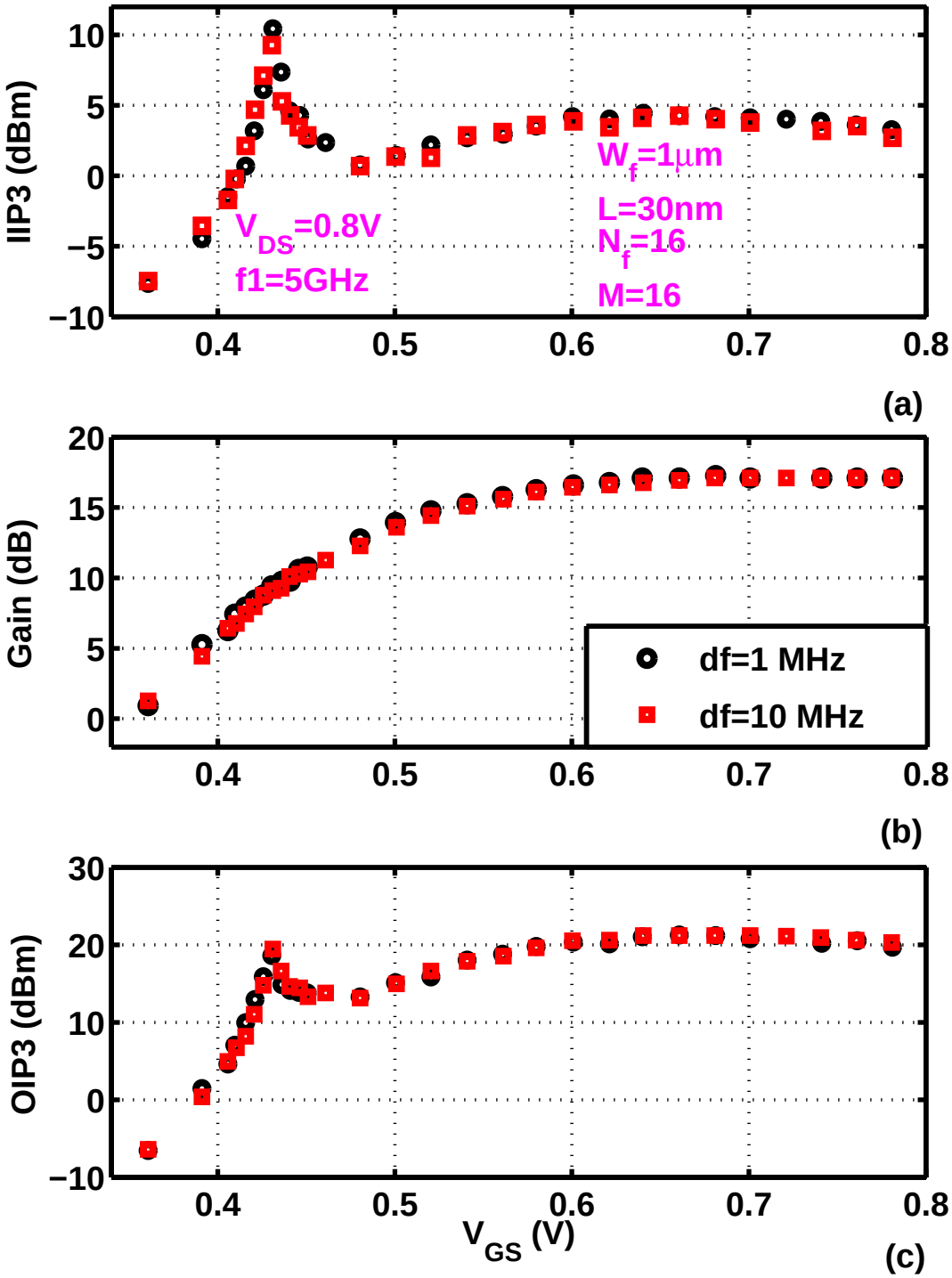

Figure 2.12: Measured (a) $IIP3$; (b) Gain and (c) $OIP3$ as a function of V_{GS} with different tone spacing . $V_{DS} = 0.8$ V.

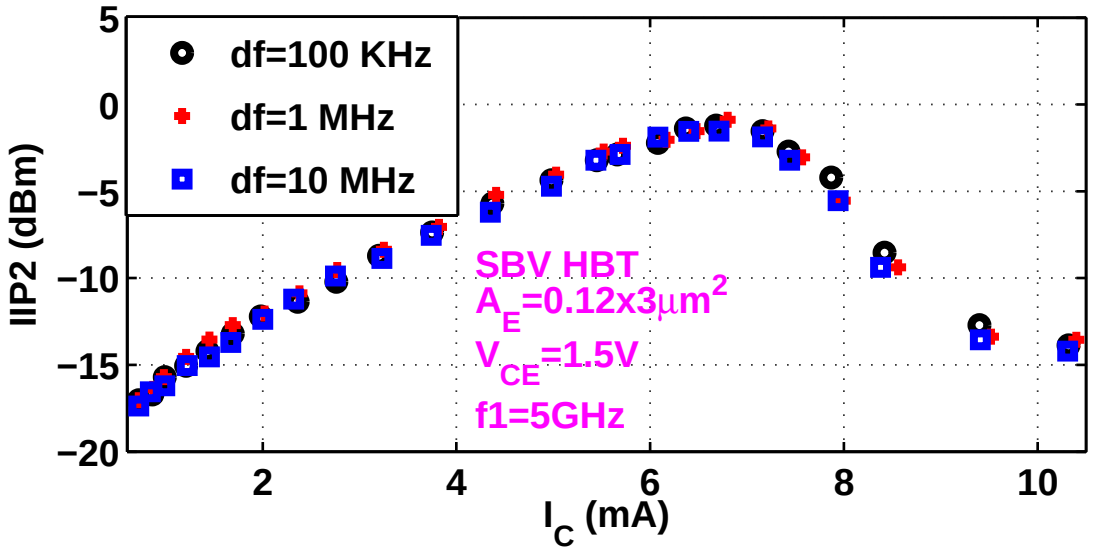

Figure 2.13: Measured $IIP2$ versus I_C for different tone spacing at 5 GHz for a SBV HBT. $V_{CE} = 1.5$ V.

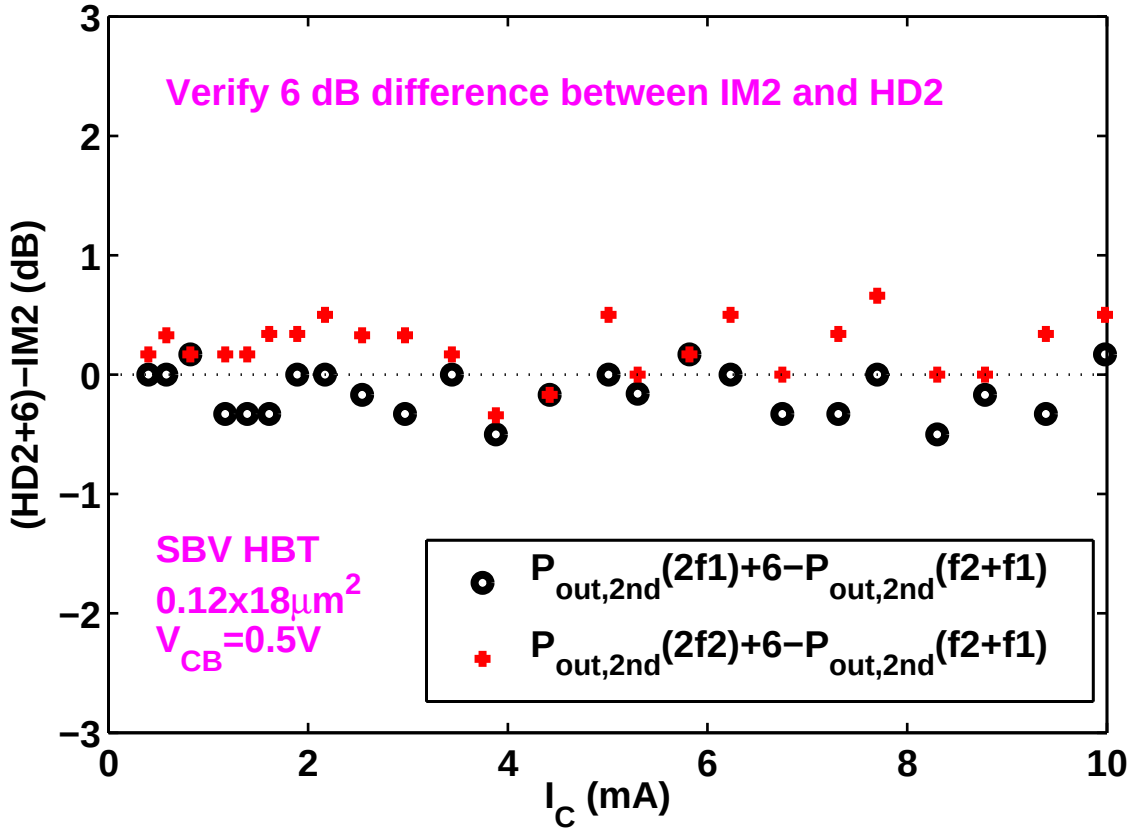

Figure 2.14: Verification of 6 dB difference between $IM2$ with $HD2$ at 5 GHz for a SBV HBT. $V_{CB} = 0.5$ V.

Chapter 3

Linearity in RF CMOS

This chapter presents experimental characterization, simulation, and Volterra series based analysis of intermodulation linearity on a high-k/metal gate 28nm RF CMOS technology. A figure-of-merit is proposed to account for both V_{GS} and V_{DS} nonlinearity, and extracted from frequency dependence of measured $IIP3$ [15]. Implications to biasing current and voltage optimization for linearity are discussed.

3.1 A new figure-of-merit $V_{GS,IP3}$

Harmonic gate voltage $IP3$ of 28nm RF CMOS devices has been recently examined using third-order derivative of $I_{DS} - V_{GS}$ data [16]. However, no experimental RF measurement of $IP3$ has been reported. Previous investigations using Volterra series analysis [17] showed that such estimation using third-order transconductance nonlinearity alone is not sufficient in characterizing transistor $IP3$. Drain conductance nonlinearity as well as cross terms involving partial derivatives of I_{DS} with respect to both V_{GS} and V_{DS} are also important [18].

As $IP3$ in RF measurements is determined using available RF power of the voltage source, the result in general depends on frequency, and cannot be directly compared with traditional gate voltage $IP3$ that is defined using the gate voltage. We propose below a new figure-of-merit that can be extracted from RF measurements so that meaningful comparison with traditional intermodulation gate voltage $IP3$ can be made with ease. The new figure-of-merit accounts for both V_{GS} and V_{DS} related nonlinearities, and reduces to traditional intermodulation gate voltage $IP3$ when all of the V_{DS} related intermodulation products are neglected.

As mentioned earlier, in RF measurement, the intercept point is defined using RF input power. The input third order intermodulation intercept point, $IIP3$, is thus dependent on frequency, because of finite source impedance, which for our case, is a 50 Ω resistance. For a given RF input power, the RF gate voltage varies with frequency, as transistor input impedance varies with frequency. For analysis as well as estimation of $IIP3$ at another design frequency from measurement at one frequency, it is desirable to find a figure-of-merit that does not depend on frequency. Such figure-of-merit is more useful if it can relate to the traditional figure-of-merit, gate voltage $VIP3$, but also include effects of drain voltage related nonlinearities. We derive such a figure-of-merit below using Volterra series analysis.

A simplified equivalent circuit as shown in Figure 3.1 is used. Gate-drain capacitance (C_{gd}) is omitted, as the result is much simpler and sufficient for most purposes [17]. $R_S = 50\ \Omega$. C_{gs} is gate-to-source capacitance. C_d is drain capacitance. $R_L = 50\ \Omega$ is load resistance.

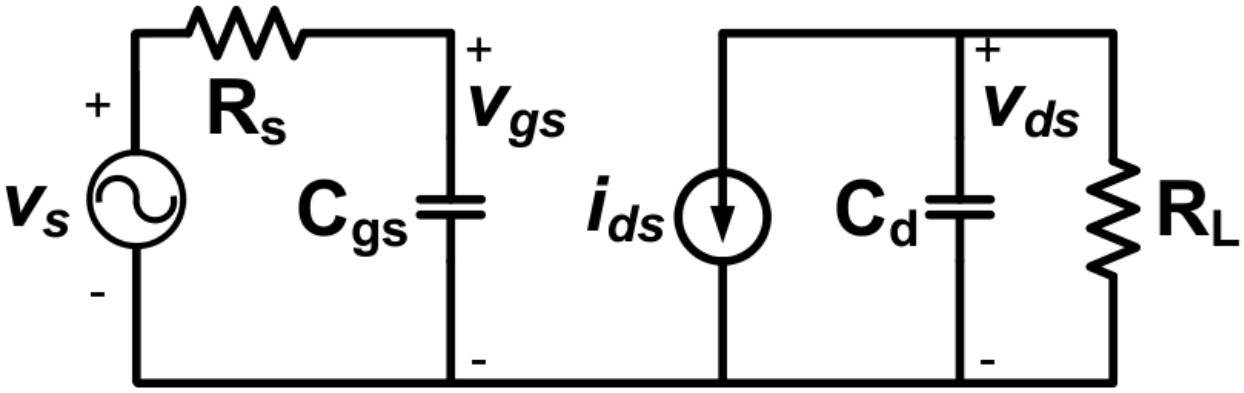

Figure 3.1: Simplified equivalent circuit used for $IP3$ derivation using Volterra series.

Up to third order, small signal nonlinear drain current i_{ds} can be expressed as:

$$\begin{aligned} i_{ds} &= g_m v_{gs} + K_{2_{g_m}} v_{gs}^2 + K_{3_{g_m}} v_{gs}^3 + g_0 v_{ds} + K_{2_{g_0}} v_{ds}^2 + K_{3_{g_0}} v_{ds}^3 + K_{2_{g_m g_0}} v_{gs} v_{ds} \\ &\quad + K_{3_{2g_m g_0}} v_{gs}^2 v_{ds} + K_{3_{g_m 2g_0}} v_{gs} v_{ds}^2 \end{aligned} \tag{3.1}$$

g_m and g_0 are transconductance and output conductance:

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} \tag{3.2}$$

$$g_0 = \frac{\partial I_{DS}}{\partial V_{DS}} \tag{3.3}$$

$K_{2_{g_m}}, K_{3_{g_m}}, K_{2_{g_0}}, K_{3_{g_0}}, K_{2_{g_m g_0}}, K_{3_{2g_m g_0}}$ and $K_{3_{g_m 2g_0}}$ are nonlinearity coefficients that relate to higher order partial derivatives as defined in [3]:

$$K_{2_{g_m}} = \frac{1}{2}\frac{\partial^2 I_{DS}}{\partial V_{GS}^2} \tag{3.4}$$

$$K_{3_{g_m}} = \frac{1}{6}\frac{\partial^3 I_{DS}}{\partial V_{GS}^3} \tag{3.5}$$

$$K_{2_{g_0}} = \frac{1}{2}\frac{\partial^2 I_{DS}}{\partial V_{DS}^2} \tag{3.6}$$

$$K_{3_{g_0}} = \frac{1}{6}\frac{\partial^3 I_{DS}}{\partial V_{DS}^3} \tag{3.7}$$

$$K_{2_{g_m g_0}} = \frac{1}{2}\frac{\partial I_{DS}}{\partial V_{GS}}\frac{\partial I_{DS}}{\partial V_{DS}} \tag{3.8}$$

$$K_{3_{2g_m g_0}} = \frac{1}{6}\frac{\partial^2 I_{DS}}{\partial V_{GS}^2}\frac{\partial I_{DS}}{\partial V_{DS}} \tag{3.9}$$

$$K_{3_{g_m 2g_0}} = \frac{1}{6}\frac{\partial I_{DS}}{\partial V_{GS}}\frac{\partial^2 I_{DS}}{\partial V_{DS}^2} \tag{3.10}$$

Using the nonlinear current source method of Volterra series, $IIP3$ in Walt can be derived [17]:

$$IIP3 = \frac{1+\omega^2 C_{gs}^2 R_S^2}{6R_S}\frac{1}{|\frac{K_{3_{g_m}}}{g_m} + \Delta|} \tag{3.11}$$

where $\Delta = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$. Δ_1 through Δ_4 are functions of nonlinear output conductance, its high order terms and cross terms with transconductance nonlinearity as follows:

$$\Delta_1 = -\frac{1}{3} K_{2_{g_m} g_0} \frac{K_{2_{g_m}}}{g_m} Z_1 - \frac{1}{3} K_{3_{2_{g_m} g_0}} Z_2 \tag{3.12}$$

$$\Delta_2 = \frac{2}{3} K_{2_{g_m}} K_{2_{g_0}} Z_3 + \frac{1}{3} K_{3_{g_m 2_{g_0}}} g_m Z_4 + \frac{1}{3} K^2_{2_{g_m g_0}} Z_5 \tag{3.13}$$

$$\Delta_3 = -K_{3_{g_0}} g_m^2 Z_6 - \frac{1}{3} K_{2_{g_m g_0}} K_{2_{g_0 g_m}} Z_7 \tag{3.14}$$

$$\Delta_4 = \frac{2}{3} K^2_{2_{g_0}} g_m^2 Z_8 \tag{3.15}$$

Z_1 through Z_8 are given by:

$$Z_1 = Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2) \tag{3.16}$$

$$Z_2 = Z_L(\omega_1) + [Y_S(-\omega_2) Y_S^{-1}(\omega_1) + 2] \tag{3.17}$$

$$Z_3 = 2Z_L(\omega_1 - \omega_2) Z_L(\omega_1) + Z_L(2\omega_1) Z_L(-\omega_2) \tag{3.18}$$

$$Z_4 = Z_L^2(\omega_1)[2Y_S(-\omega_2) Y_S^{-1}(\omega_1) + 1] \tag{3.19}$$

$$Z_5 = 2Z_L(\omega_1 - \omega_2) Z_L(-\omega_2) + Z_L(2\omega_1) Z_L(-\omega_1) \tag{3.20}$$

$$Z_6 = Z_L^2(\omega_1) Z_L(-\omega_2) \tag{3.21}$$

$$Z_7 = Z_L^2(\omega_1)[Z_L(2\omega_1) + 2Z_L(2\omega_1) + 6Z_L(\omega_1 - \omega_2)] \tag{3.22}$$

$$Z_8 = Z_L^2(\omega_1) Z_L(-\omega_2)[Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2)] \tag{3.23}$$

with $Z_L(\omega) = \frac{1}{\frac{1}{R_L} + g_0 + j\omega C_d}$ and $Y_S(\omega) = \frac{1}{R_S} + j\omega C_{gs}$. Detailed derivation of $IIP3$ in [19] is reproduced in Appendix B.

A close inspection of the Volterra series based derivation details shows that at the intermodulation $IP3$ point, the first order v_{gs} has an amplitude of:

$$V_{GS,IP3} = \sqrt{\frac{1}{|\frac{3}{4}\frac{K_{3_{g_m}}}{g_m} + \Delta|}} \tag{3.24}$$

For typical transistor sizes of interest, the Δ term is found to have a negligibly weak frequency dependence, making $V_{GS,IP3}$ nearly frequency independent in practice. We thus propose to use $V_{GS,IP3}$ as a figure-of-merit as it includes output conductance effect, and is more general than the traditional $VIP3$ defined solely using $K_{3_{g_m}}$ and g_m. The designation GS in the subscript refers to the fact that this is the V_{GS} amplitude at the intercept. The value of $V_{GS,IP3}$, however, is clearly a function of the V_{DS} dependence of I_{DS}, through the Δ term.

Using $V_{GS,IP3}$, Equation (3.11) can then be rewritten as

$$IIP3 = \frac{C_{gs}^2 R_S V_{GS,IP3}^2}{8}\omega^2 + \frac{V_{GS,IP3}^2}{8R_S} \tag{3.25}$$

(3.25) indicates that $IIP3$ increases linearly with ω^2 and $V_{GS,IP3}$ can be obtained experimentally by plotting measured $IIP3$ as a function of ω^2. An example is shown in Figure 3.2.

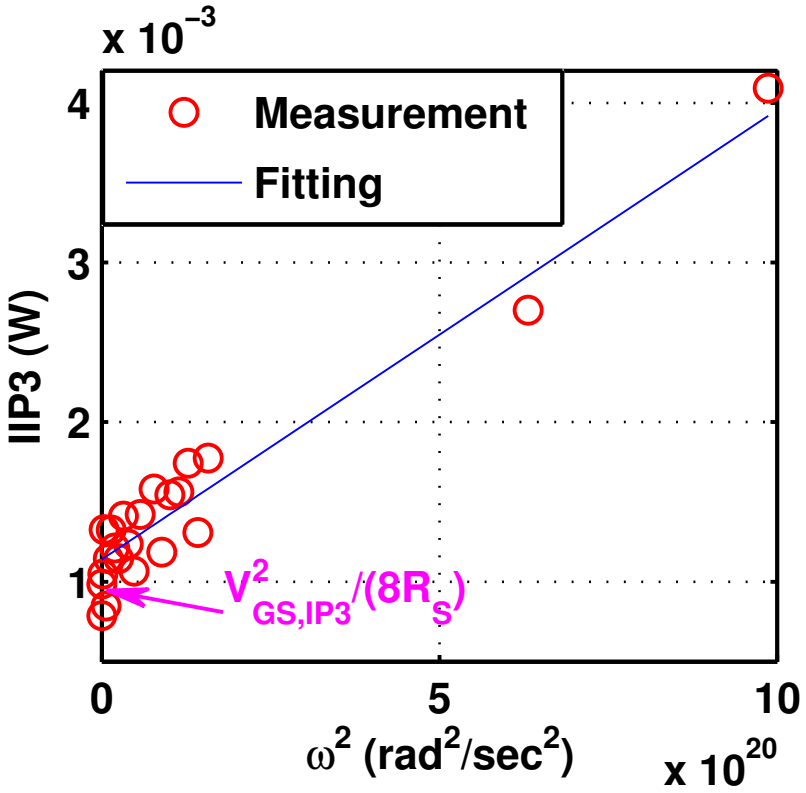

Figure 3.2: Measured $IIP3$ in watt versus ω^2. $V_{GS} = 0.7$ V and $V_{DS} = 1.0$ V.

A linear fitting is made. The intercept with the $IIP3$ axis gives $V_{GS,IP3}^2/8R_S$. Note that the unit used for $IIP3$ is watt instead of dBm. As measured $IIP3$ in dBm is shown in Figure 3.3. The device has a drawn gate length of 30nm. $W_f = 4$ μm. $N_f = 16$. Multiplicity $M = 4$. The total width $W_{total} = 256\mu$m. $V_{GS} = 0.7$ V and $V_{DS} = 1.0$ V. Measurement frequency ranges from 100 MHz to 10 GHz. Within measurement uncertainty, the data shows an expected linear dependence on the square of fundamental angular frequency. This linear dependence of $IIP3$ on ω^2 is found to be valid for other bias points as well. The slope is given by $\frac{C_{gs}^2 R_S V_{GS,IP3}^2}{8}$ from which C_{gs} can be extracted. The C_{gs} calculated is fairly close to that extracted from S-parameter measurements, thus supporting the validity of the proposed technique.

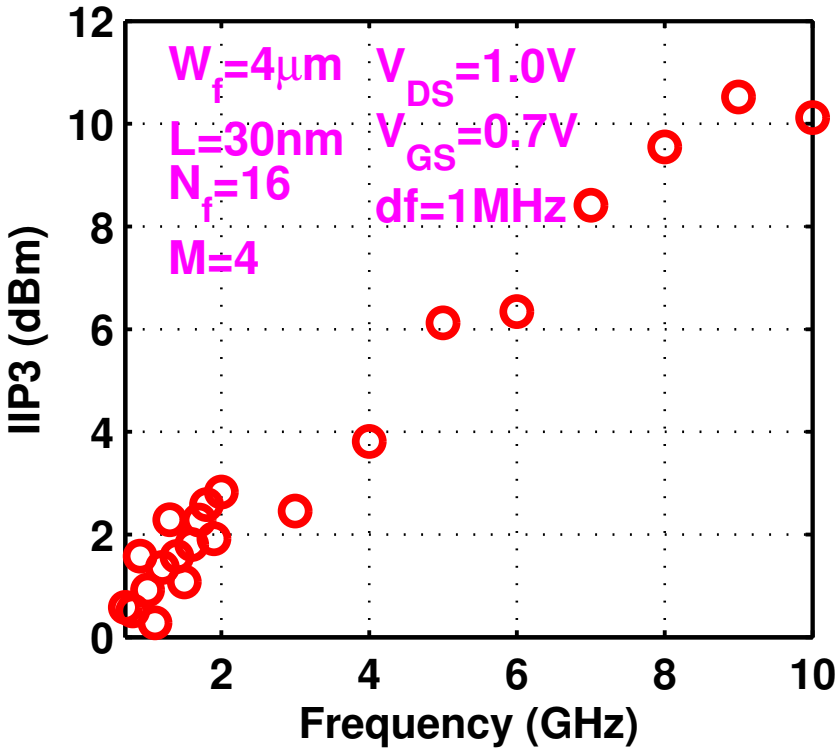

Figure 3.3: Measured $IIP3$ in dBm versus frequency. $V_{GS} = 0.7$ V and $V_{DS} = 1.0$ V.

If we ignore the Δ term that originates from the v_{ds} dependence of i_{ds}, $V_{GS,IP3}$ reduces to

$$VIP3 = \sqrt{\frac{4}{3}|\frac{g_m}{K_{3_{g_m}}}|} \tag{3.26}$$

This is essentially the $V_{GS,IP3}$ one would get if transistor drain current depends on V_{GS} only.

In [16, 20], third order harmonic distortion $VIP3$ is expressed as

$$VIP3_{hd} = \sqrt{24|\frac{g_m}{K_{3_{g_m}}}|} \tag{3.27}$$

This $VIP3$ differs from the third order intermodulation $VIP3$ in (3.26) by a constant.

3.2 V_{GS} dependence

Figure 3.4(a) shows both measured and simulated $IIP3$ at 5 GHz as a function of V_{GS} at $V_{DS} = 0.6$ V for a 30nm MOSFET. The total width $W_{total} = 256\mu$m. $W_f = 1$ μm. $N_f = 16$. Multiplicity $M = 16$. $V_{GS} = 0.7$ V and $V_{DS} = 1.0$ V. Measurements and simulations are also made at 2 and 10 GHz. At each V_{GS}, from frequency dependence of $IIP3$, a $V_{GS,IP3}$ is extracted. From 0.5 to 0.7 V, simulated $IIP3$ is higher than measured $IIP3$ by as much as 3.8 dB. This indicates that simulated $IIP3$ for such technologies may be optimistic. In future work, model parameters can be further optimized to see if $IIP3$ can be better fitted. To our knowledge, there are no direct knobs to turn to tune higher order derivatives in compact models. Improvement of $IIP3$ simulation may require new improvements of the model formulation itself in addition to better parameter extraction and optimization.

Figure 3.4(b) shows the $VIP3$ calculated from $K_{3_{g_m}}$ and g_m using Equation (3.26). Fitting of $VIP3$, which is determined by the first and third order derivatives of I_{DS}-V_{GS}, is clearly worse than the fitting of I_{DS}-V_{GS} itself shown earlier in Figure 4.4.

Figure 3.4(c) and (d) show $V_{GS,IP3}$ and $K_{3_{g_m}}$ as a function of V_{GS}. The $K_{3_{g_m}} = 0$ point is clearly different from the measured $IIP3$ and $V_{GS,IP3}$ peak positions. The peak $IIP3$ V_{GS} is 55 mV lower than the peak $VIP3$ V_{GS}. As was observed in 90 nm technology [17], $VIP3$ does not correctly predict the linearity sweet spot, due to omission of the Δ term. Around $V_{GS} = 0.6$ V, $V_{GS,IP3}$ and the traditional $VIP3$ are close to each other, as the Δ term is small. Beyond its peak, $IIP3$ drops to a valley and starts rising slowly. However,

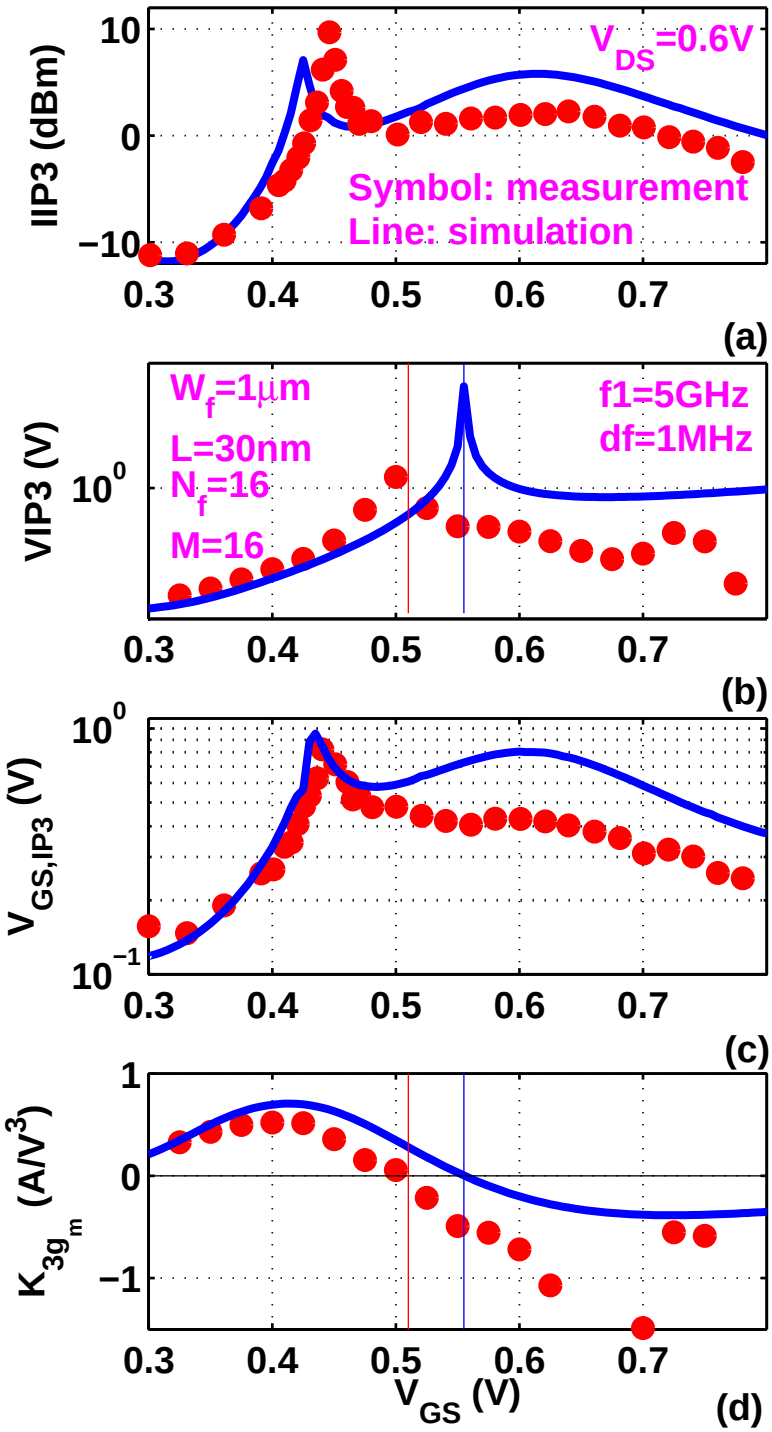

Figure 3.4: Comparison of measured (a) $IIP3$ (b) $VIP3$ (c) $V_{GS,IP3}$ and (d) $K_{3_{g_m}}$ as a function of V_{GS} with simulation at $V_{DS} = 0.6$ V.

when $V_{GS} > 0.65$ V, as the device gets closer to linear operation region, $IIP3$ shows a slight decrease.

3.3 V_{DS} impact

Figure 3.5(a), (b), (c) and (d) show measured $IIP3$, $IIP2$, $V_{GS,IP3}$, and $VIP3$ as a function of V_{GS} at $V_{DS} = 0.6$ and 1.0 V. The same device as in Figure 3.4 is used. As can be seen from Figure 3.5(a), $IIP3$ curves at high V_{DS} are shifted towards low V_{GS} direction due to decreased threshold voltage, a consequence of drain induced barrier lowering. In strong inversion region, at the same V_{GS}, a higher V_{DS} results in a higher $IIP3$. For instance, at $V_{GS} = 0.8$ V, $IIP3$ increases by 7.7 dB when V_{DS} increases from 0.6 to 1.0 V. As shown in Figure 3.5(b), $IIP2$ has a clear peak, though not as sharp as $IIP3$, around $V_{GS} = 0.6$ V, in strong inversion. If both high $IIP3$ and high $IIP2$ are desired, the transistor should be biased around $V_{GS} = 0.6$ V, which is approximately 200 mV above threshold voltage. A comparison of Figure 3.5(c) and (d) shows that the V_{DS} dependence of $V_{GS,IP3}$ and hence $IIP3$ is insufficiently captured by $VIP3$, due to lack of v_{ds} related terms, as expected.

3.4 Impact of device size

Figure 3.6 shows measured $IIP3$ at 5 GHz for devices with $W_{total} = 153.6$ and 256μm. Note that the device finger widths are 0.3 and 1μm respectively. At both very low and high I_{DS}, a large device gives a large $IIP3$. Both peak $IIP3$ value and peak $IIP3$ I_{DS} decrease with device width. Narrow width effect clearly plays a role in affecting the position of the linearity peak.

Figure 3.7 shows measured $IIP3$ as a function of V_{GS} for two 30nm MOSFETs with the same total width of 256μm. As the device finger widths are both large, 2 and 4μm respectively, no narrow width effect is observed, and $IIP3$ is largely the same for the two devices as expected.

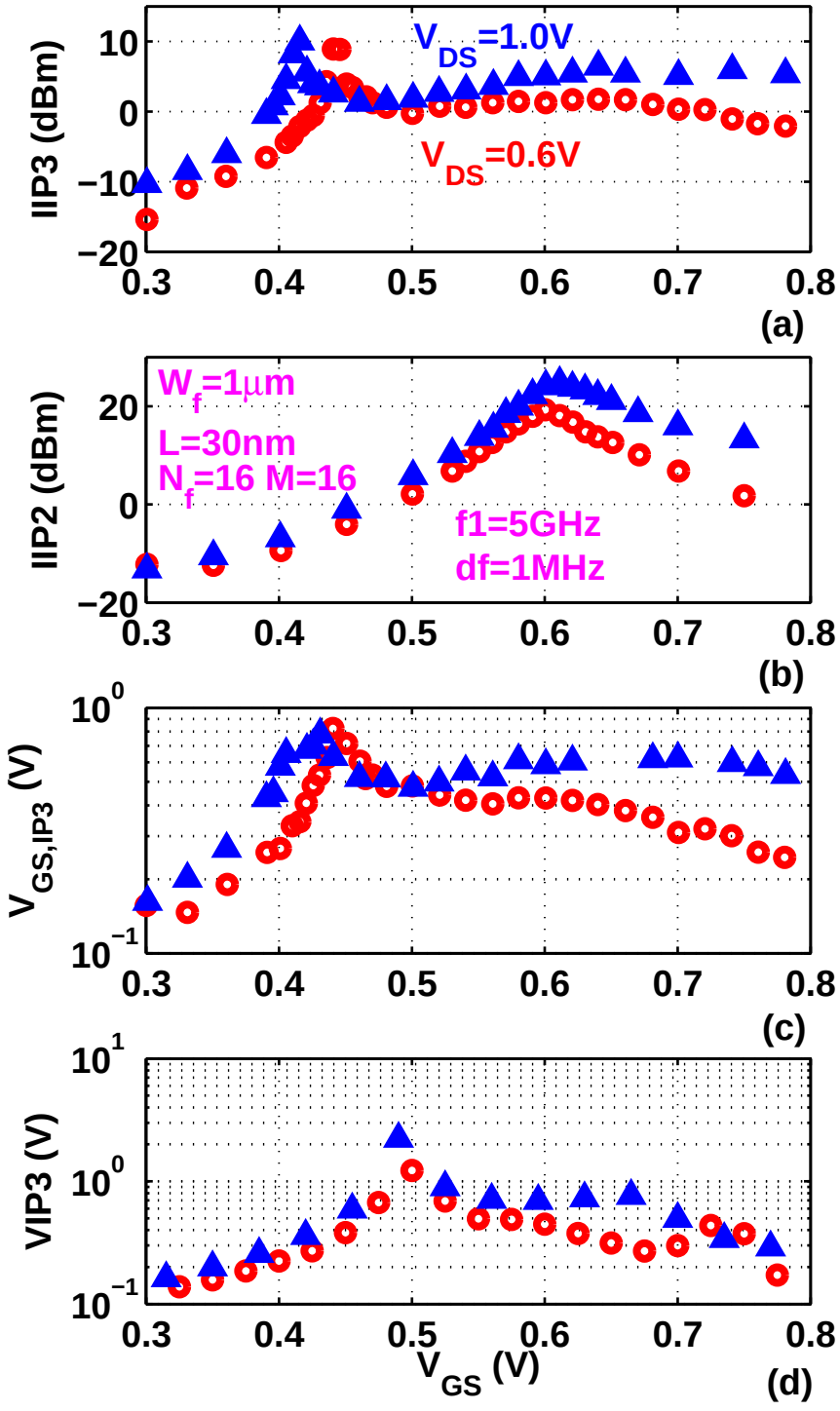

Figure 3.5: Measured (a) $IIP3$; (b) $IIP2$ and (c) $V_{GS,IP3}$ and (d) $VIP3$ as a function of V_{GS} for different V_{DS}.

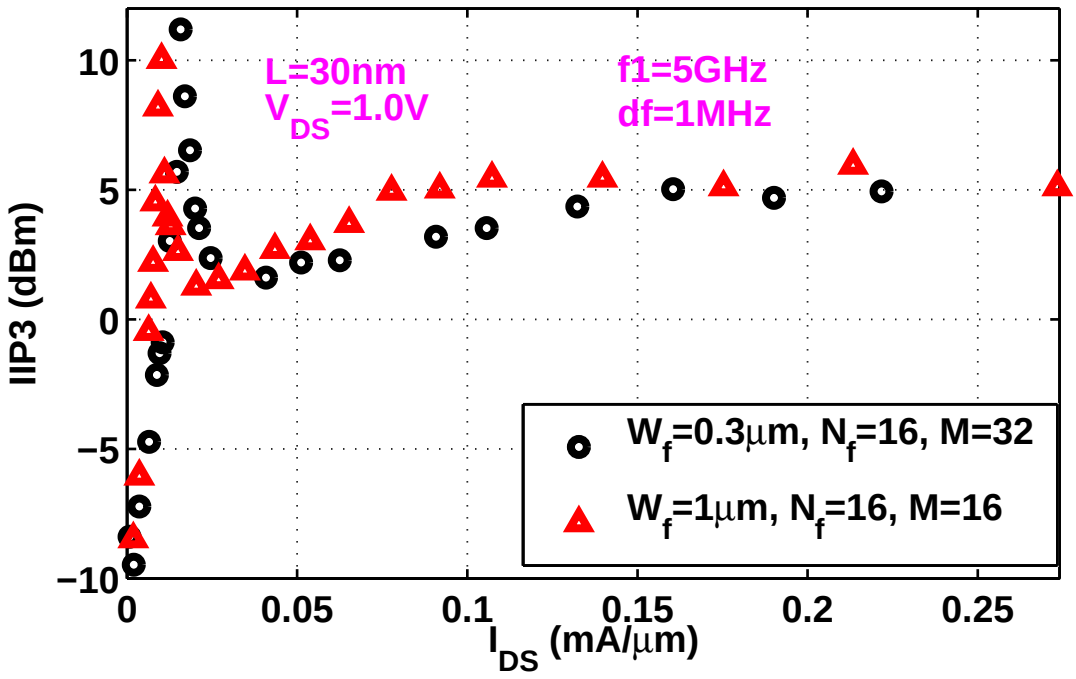

Figure 3.6: Measured $IIP3$ at 5 GHz for two 30nm devices with different total width.

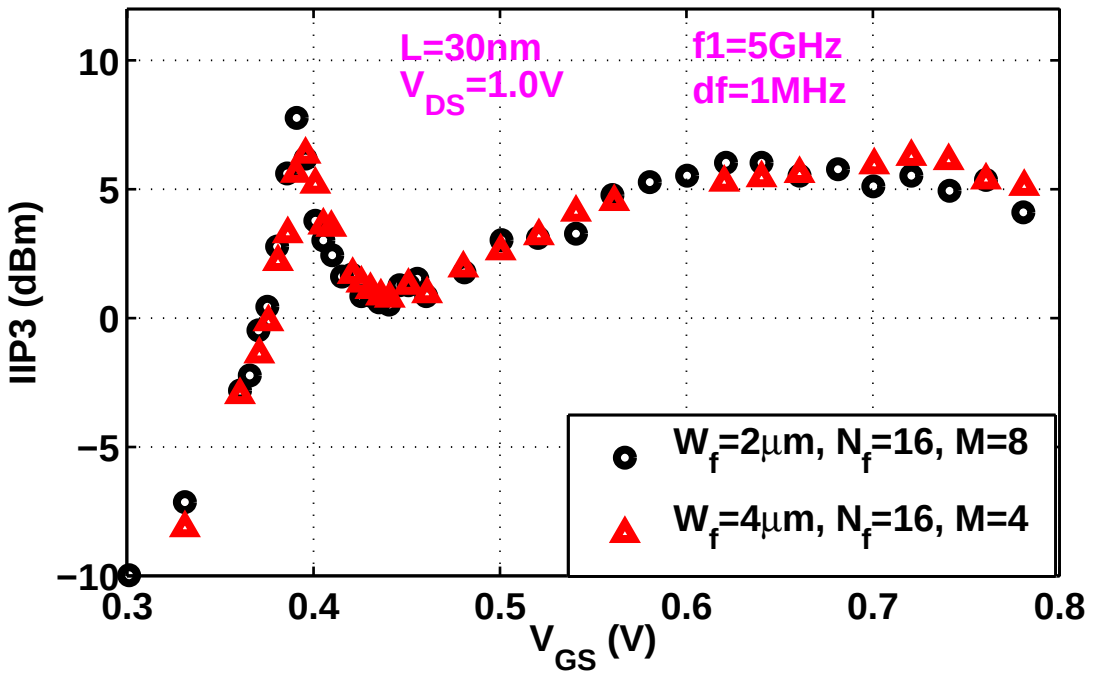

Figure 3.7: Measured $IIP3$ as a function of V_{GS} for two 30nm devices with same total width.

3.5 Impact of channel length

Channel length is one of important design variables for RFIC design. In fact, the scaled channel length in latest RF CMOS technologies provides sufficient cut-off frequency to meet speed requirement so that the channel length can be sacrificed to trade for other performance specifications.

As reviewed by $V_{GS,IP3}$ in Figure 3.4(b), both V_{GS} and V_{DS} nonlinearities need to considered. Otherwise one will end up with inaccurate result as shown in Figure 3.4(b). However, these two groups of nonlinearities both affect the impact of channel length on $IP3$. Considering other device parameters are the same, a longer channel length weakens both drain induced barrier lowering and channel length modulation effects, resulting in reduced output conductance g_0 and its nonlinearities. However, a shorter has a stronger carrier velocity which in turn gives a more linear transconductance g_m [21]. Due to well known difficulty to fine tune the design kit to model higher order nonlinearity terms for both g_m and g_0, it is hard to conclude which effect dominates over the other.

Like previous RF CMOS technologies, the 28nm HKMG RF CMOS technology we examined provides multiple oxide thickness devices for flexible circuit designs. According to [21], thicker oxide devices provide better linearity compared to thin oxide devices. for lack of the same device configuration (W_f, L, N_f, and W_{total}), in this work,we cannot conclude that oxide devices dominates thin oxide ones. Below both V_{GS} and $IIP3$ dependence of $IIP3$ are examined separately for both cases.

Figure 3.8(a) and (b) show measured $IIP3$ and Gain as a function of V_{GS} for thick oxide devices with different channel length at 5 GHz. V_{DS} = 1.5 V. Each device has same total width ($W_{total} = 128\mu$m), finger width ($W_f = 2$ μm), number of fingers ($N_f = 8$), and multiplicity ($M = 8$).

For thick oxide devices, scaling down channel length shifts peak $IIP3$ V_{GS} toward lower V_{GS} values due to decreased threshold voltage by short channel effect (SCE), which can be explained by shift of peak $K_{2_{g_m}}$ V_{GS} in Figure 3.9. In strong inversion, channel length has a

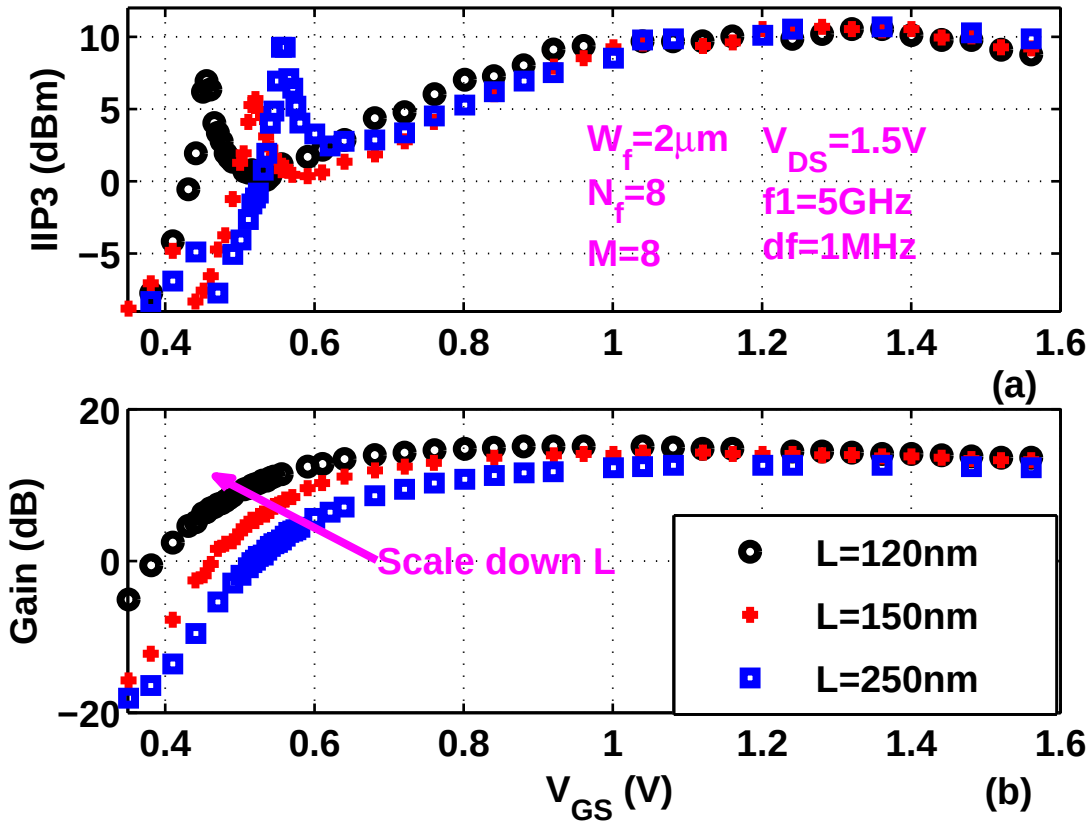

Figure 3.8: Measured (a) $IIP3$ and (b) Gain as a function of V_{GS} for three thick oxide devices with different channel length. $V_{DS} = 1.5$ V.

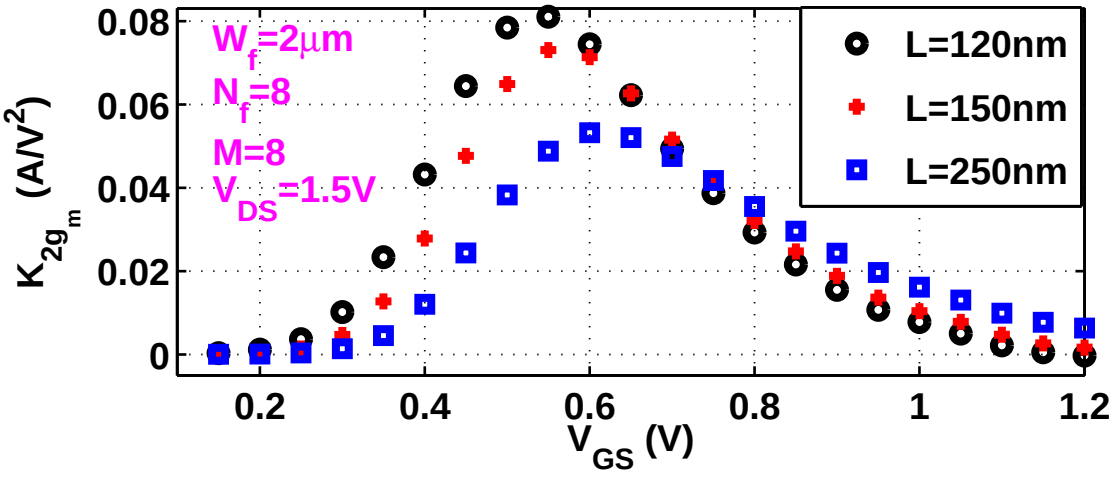

Figure 3.9: Measured $K_{2_{g_m}}$ as a function of V_{GS} for thick oxide devices with different channel length at $V_{DS} = 1.5$ V. Devices have the same total width.

small impact on $IP3$. For example, scaling down channel length from 250nm to 120nm only increases $IIP3$ by 1.9 dB at $V_{GS} = 0.76$ V. In subthreshold region, decreasing L effectively increases power gain. At $V_{GS} = 0.41$ V, Gain is increased by 18.0 dB with L decreased from 250nm to 120nm.

Figure 3.10 shows measured $IIP3$ as a function of I_{DS} for devices in Figure 3.8. $V_{DS} = 1.5$ V. For whole I_{DS} range, $IIP3$ is similar to 120nm and 150nm devices. The $IIP3$ peak occurs at a lower I_{DS} in longer channel devices. Well past $IIP3$ peak, when I_{DS} is below 0.2 mA/μm, $IIP3$ is increased by 3 dB or so by scaling L up from 150nm to 250nm for same I_{DS}. This suggests that a longer than minimum channel length device is beneficial to improve linearity when gain requirement is satisfied.

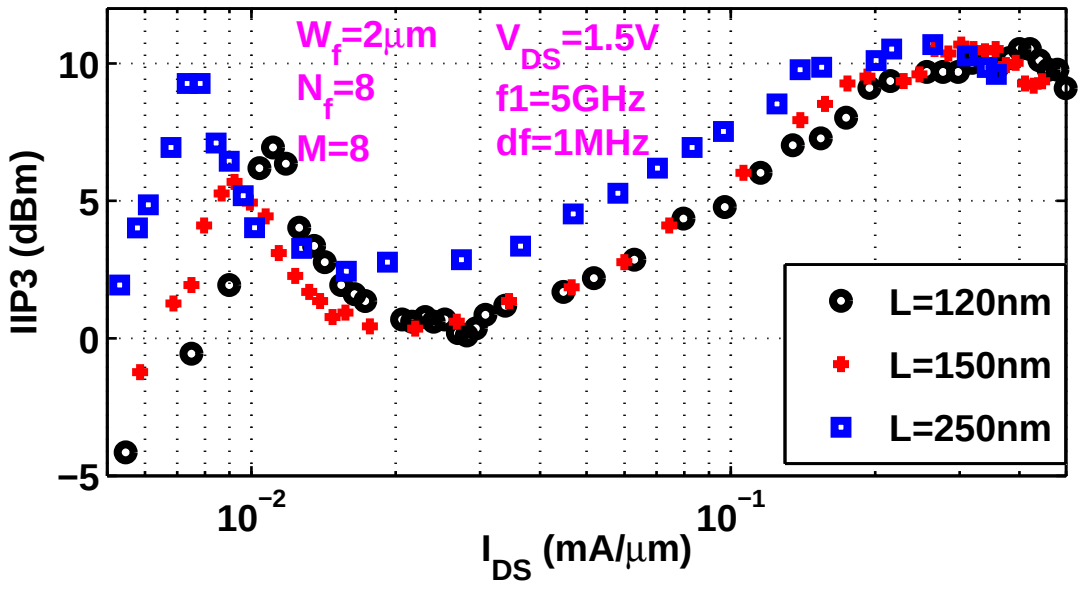

Figure 3.10: Measured $IIP3$ as a function of I_{DS} for three thick oxide devices with different channel length. $V_{DS} = 1.5$ V.

Figure 3.11 shows measured $IIP3$ as a function of V_{GS} for devices in Figure 3.8 at $V_{DS} = 0.8$ V. Interestingly, $IIP3$ does not have a as smooth plateau with V_{GS} as in Figure 3.8, but a much higher $IIP3$ shows up at $V_{GS} = 1.5$ V. Between $V_{GS} = 1.4$ and $V_{GS} = 1.6$ V, devices with different length have similar $IIP3$ for the same V_{GS}. This indicates, in this range, channel length does not need to be sacrificed much to achieve high $IIP3$.

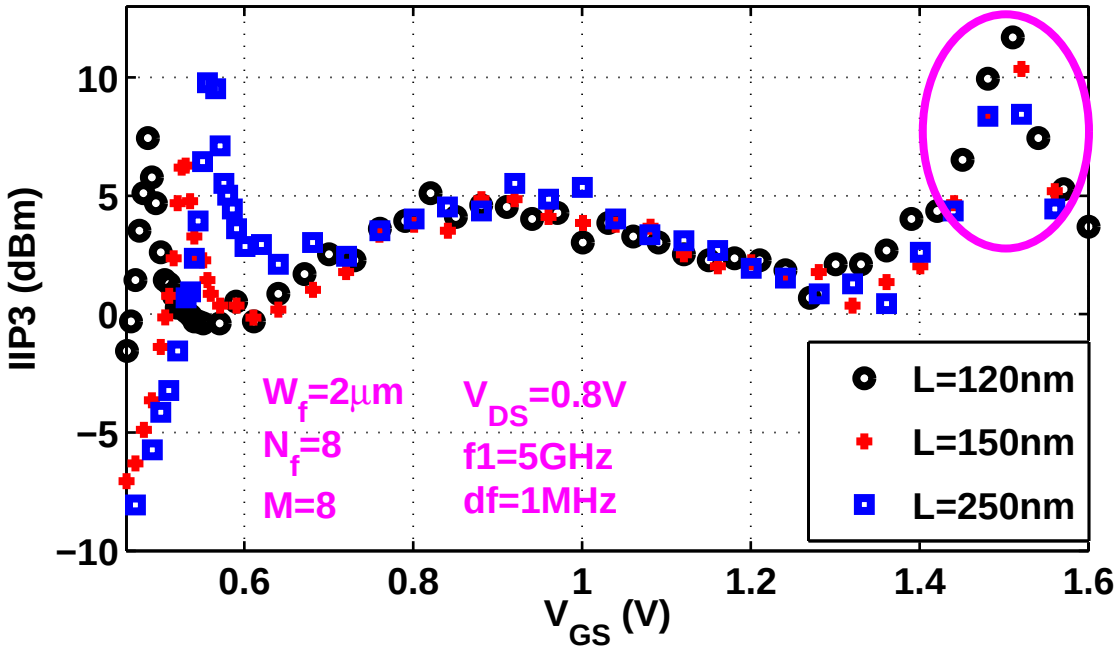

Figure 3.11: Measured $IIP3$ as a function of V_{GS} for thick oxide devices with different channel length at $V_{DS} = 0.8$ V. Devices have the same total width.

Figure 3.12(a) and (b) show measured $IIP3$ as a function of V_{GS} and I_{DS} for thin oxide devices with different channel length at 5 GHz. $V_{DS} = 1.0$ V. Two devices same total width ($W_{total} = 128\mu$m), finger width ($W_f = 1$ μm), number of fingers ($N_f = 16$), and multiplicity ($M = 8$). Similar to oxide devices, scaling down channel length results in a shift of peak $IIP3$ V_{GS} toward a lower value, which makes a shorter channel device have a much larger $IIP3$. For example, at $V_{GS} = 0.49$ V, $IIP3$ for 40nm device 9.1 dB larger than 60nm device. This indicates, in weak inversion, both higher gain and higher linearity can be obtained by choosing shorter channel length for thin oxide devices. Figure 3.12(b) shows that, for the same current, 40nm and 60nm devices have the same $IIP3$.

3.6 Impact of frequency

As indicated by (3.25), as R_S and $V_{GS,IP3}$ are frequency independent, the frequency impact on $IIP3$ is mainly affected by C_{gs} as well as the square of omega frequency. Figure

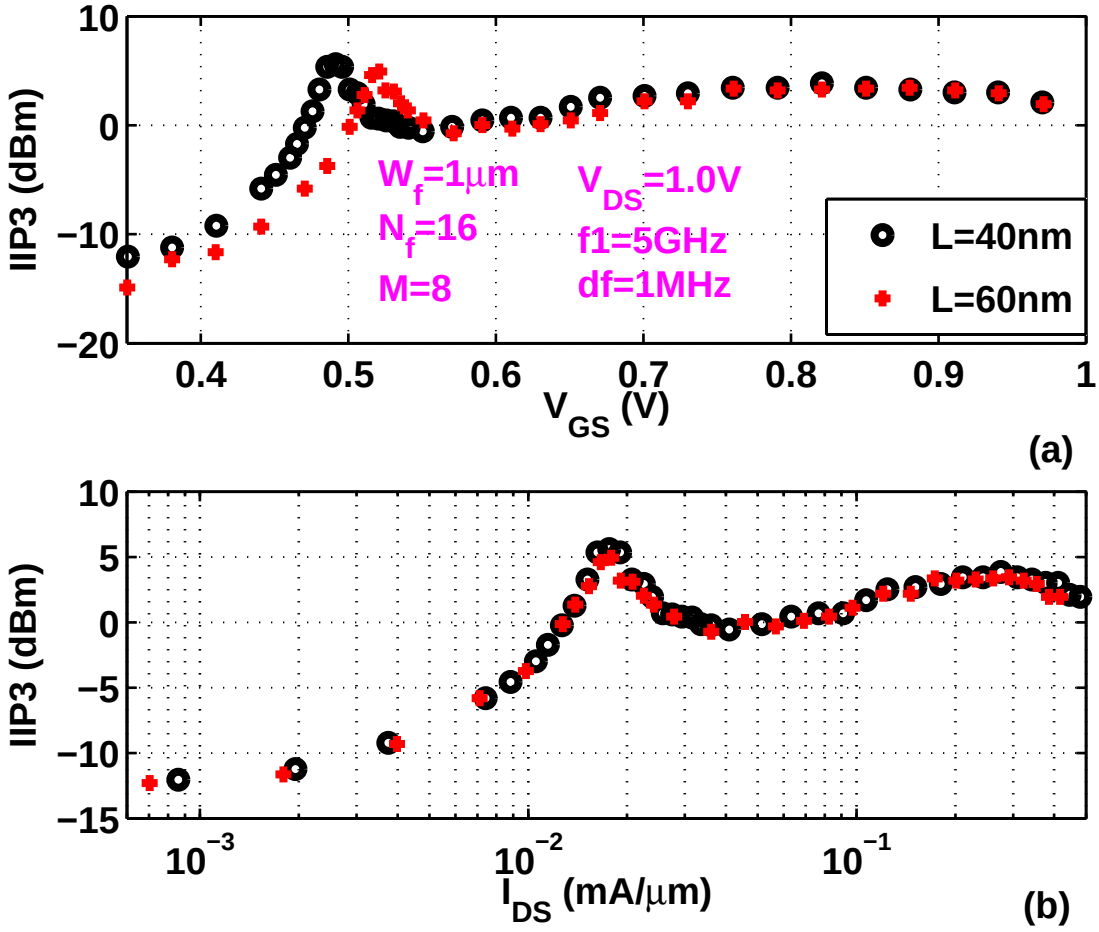

Figure 3.12: Measured $IIP3$ as a function of (a) V_{GS} and (b) I_{DS} for two thin oxide devices with different channel length. $V_{DS} = 1.0$ V.

3.13(a), (b) and (c) show measured $IIP3$, Gain and $OIP3$ at different frequency ($f1$) for the same device in Figure 3.4. $V_{DS} = 1.0$ V. High frequency gives a high $IIP3$ and a low gain at same I_{DS}. $OIP3$ is same same for different frequency.

3.7 Analytic expression of $IIP3$

Figure 3.14(a) compares measured $IIP3$ with simulation and calculation by (3.11) as a function of V_{GS} for the device in Figure 4.3. $V_{DS} = 0.6$ V. Comparison of I_{DS} dependence is shown in Figure 3.14(b). For both V_{GS} and I_{DS} dependence, Volterra series based analytic expression (3.11) accurately captures $IIP3$ variation in both weak and strong inversion

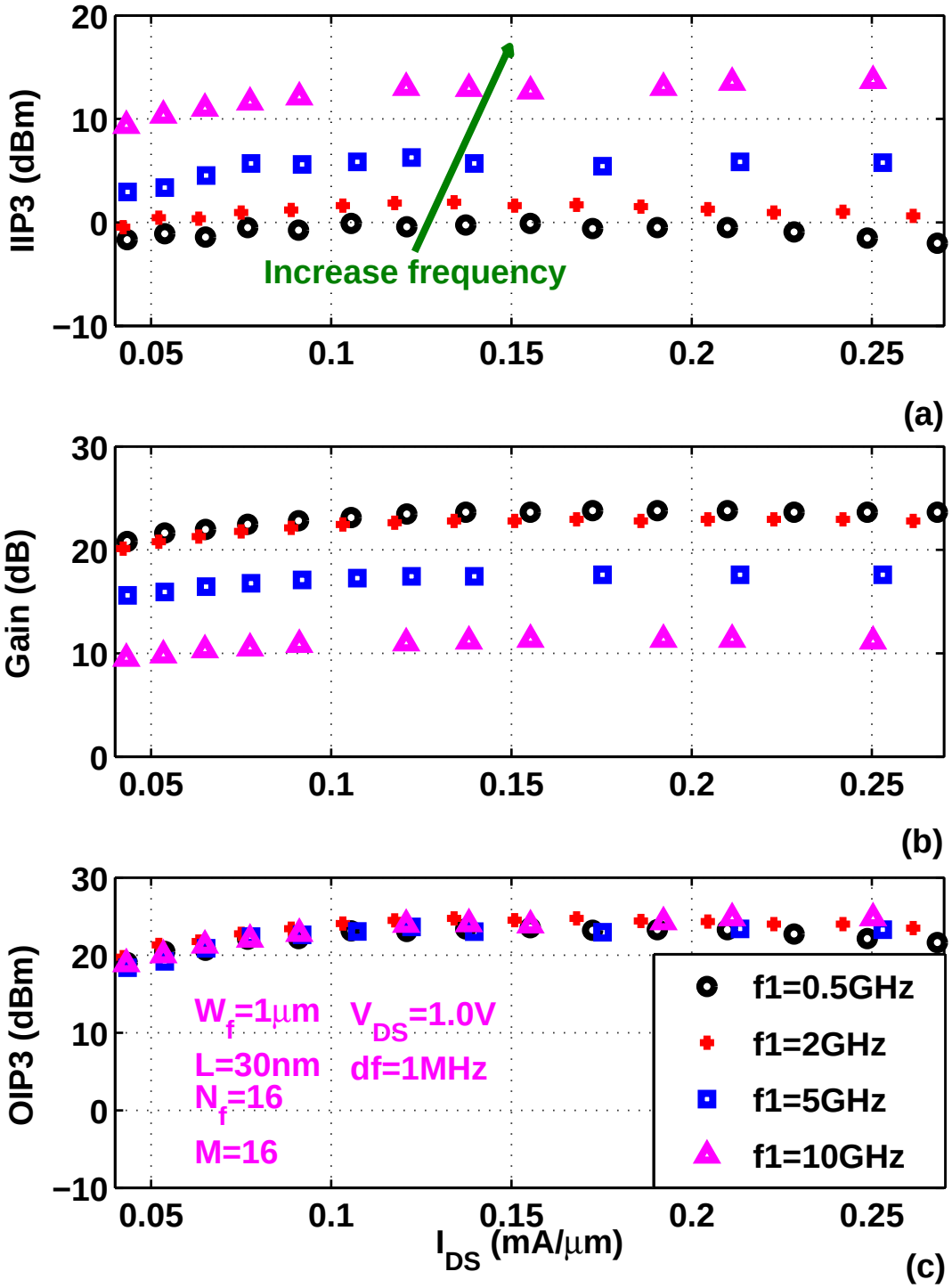

Figure 3.13: Measured (a) $IIP3$ (b) Gain and (c) $OIP3$ as a function of I_{DS} for different frequency.

regions. This provides an easy way to distinguish nonlinearity contributions without using full Volterra series analysis.

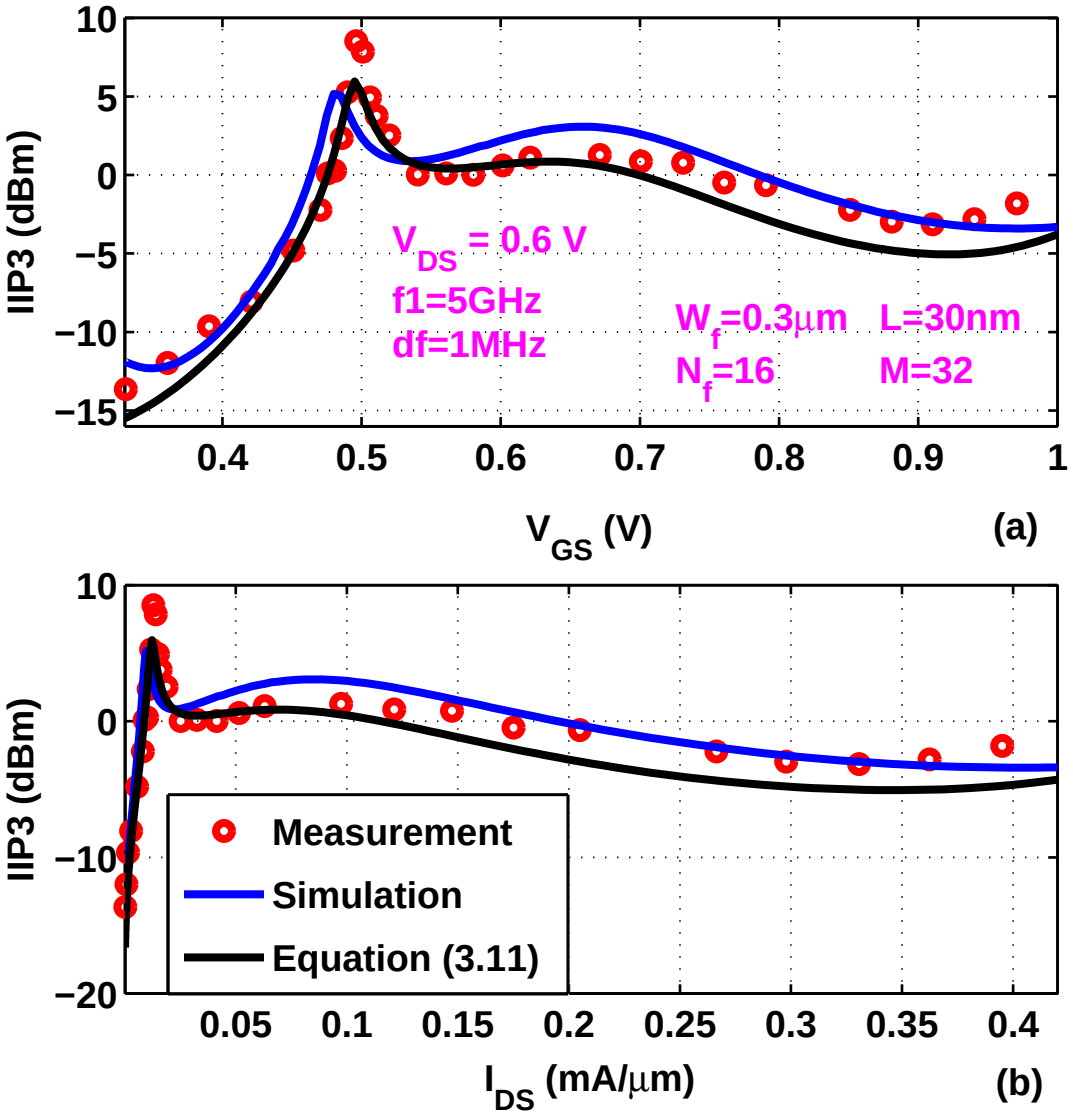

Figure 3.14: Comparison of measured $IIP3$ with simulation and calculation by (3.11) as functions of (a) V_{GS} and (b) I_{DS}. $V_{DS} = 0.6$ V.

To understand how contributions of Δ term affects V_{GS} dependence of $IIP3$ in analytic expression (3.11), $\frac{K_{3gm}}{g_m} + \Delta$, $\frac{K_{3gm}}{g_m}$, $\Delta 1$, $\Delta 2$, $\Delta 3$ and $\Delta 4$ are plotted as a function of V_{GS} in Figure 3.15. $V_{DS} = 0.6$ V.

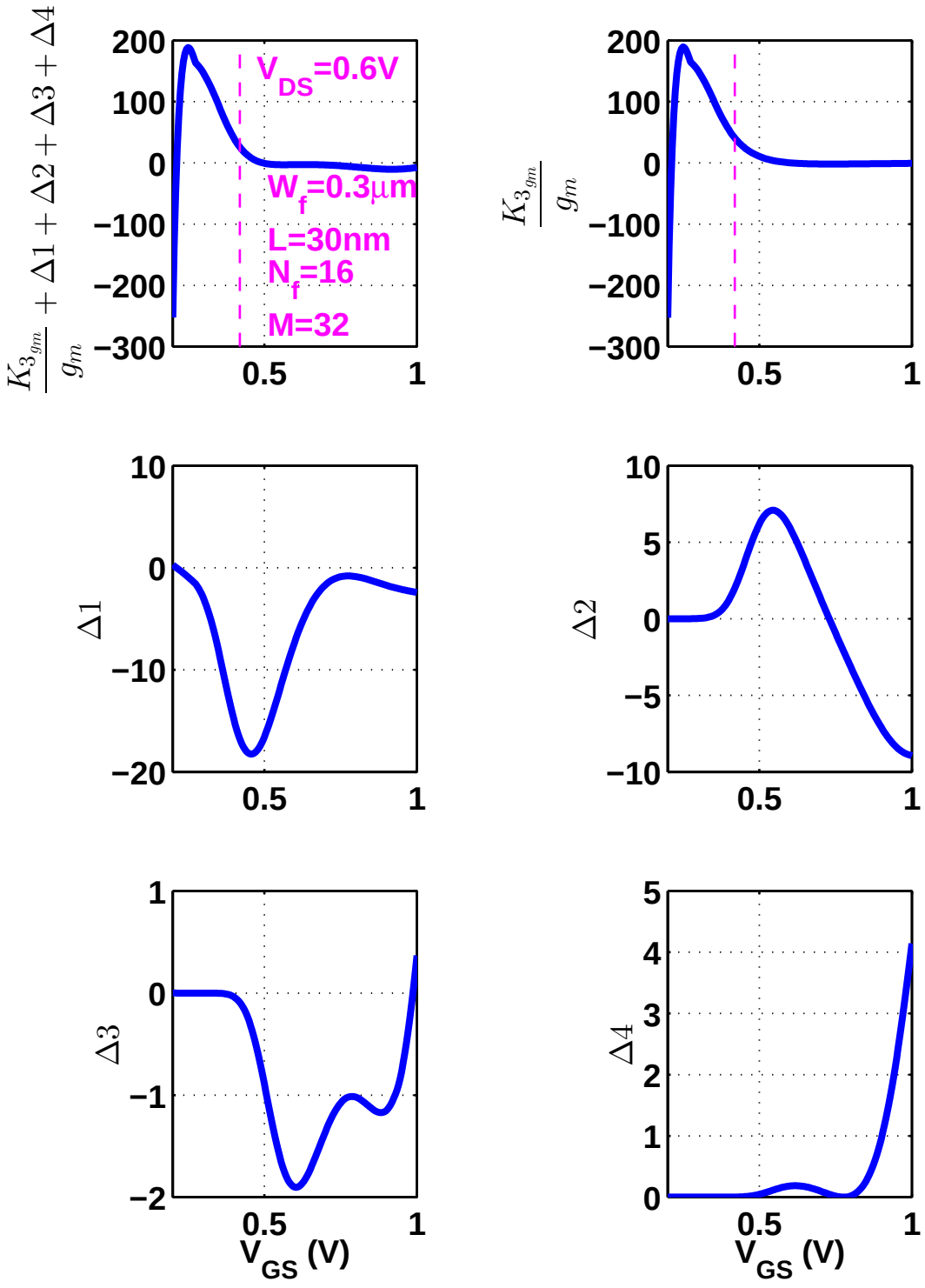

Figure 3.15: V_{GS} dependence of $\frac{K_{3_{gm}}}{g_m} + \Delta$, $\frac{K_{3_{gm}}}{g_m}$, $\Delta1$, $\Delta2$, $\Delta3$ and $\Delta4$.

Below V_{GS} =0.4 V, $\frac{K_{3g_m}}{g_m}+\Delta$ has a very similar shape to $\frac{K_{3g_m}}{g_m}$. This indicates $\frac{K_{3g_m}}{g_m}$ plays a predominant role for the denominator of (3.11). Above V_{GS} =0.4 V, the absolute values of $\Delta 1$ and $\Delta 2$ are much larger than $\Delta 3$ and $\Delta 4$, meaning $\Delta 1$ and $\Delta 2$ are more important at high V_{GS} compared to $\Delta 3$ and $\Delta 4$. At very high V_{GS} (close to $V_{GS} = 1$ V), $\Delta 4$ is a positive number and increases quickly with V_{GS} while $\Delta 1$, $\Delta 2$ and $\Delta 3$ are negative numbers. This suggests $\Delta 4$ has a strong impact in this V_{GS} range.

Figure 3.16(a) and (b) show calculated (3.11) and the denominator in (3.24) a function of V_{GS}. The black line is obtained using $\frac{K_{3g_m}}{g_m}$ only in the denominator. Green, blue and red lines are obtained using a denominator value of $\frac{K_{3g_m}}{g_m}+\Delta 1+\Delta 2$, $\frac{K_{3g_m}}{g_m}+\Delta 1+\Delta 2+\Delta 3$, and $\frac{K_{3g_m}}{g_m}+\Delta 1+\Delta 2+\Delta 3+\Delta 4$, respectively. As expected, calculated (3.11) in black by $\frac{K_{3g_m}}{g_m}$ is merged with the full (3.11) expression in red when V_{GS} is smaller than 0.4 V. Below V_{GS} =0.5 V or so, calculated (3.11) in green by $\frac{K_{3g_m}}{g_m}+\Delta 1+\Delta 2$ can accurately predict $IIP3$ shape and peak $IIP3$ V_{GS}. For V_{GS} smaller than about 0.9 V, $\Delta 3$ helps determine the $IIP3$ values. At very high V_{GS}, $\Delta 4$ is important.

Figure 3.17(a), (b) and (c) show calculated $IIP3$ by (3.11), $\frac{K_{3g_m}}{g_m}+\Delta$ and $\frac{K_{3g_m}}{g_m}+\Delta$ as functions of V_{GS} at V_{DS} = 0.6 and 1.0 V. The 0.5 V shift of peak $IIP3$ V_{GS} clearly originates from the same magnitude of shift for $\frac{K_{3g_m}}{g_m}$ and therefore $\frac{K_{3g_m}}{g_m}+\Delta$ as shown by the zoomed-in view in Figure 3.17(b) and (c). Four contributions of Δ term barely affects $IIP3$ as discussed above.

Fig. 3.18 (a)-(f) show $\frac{K_{3g_m}}{g_m}+\Delta$, $\frac{K_{3g_m}}{g_m}$, $\Delta 1$, $\Delta 2$, $\Delta 3$ and $\Delta 4$ at high V_{GS} for $V_{DS} = 0.6$ and 1.0 V. Y-axis range is scaled to the same. $\Delta 1$ and $\Delta 2$ are more sensitive to V_{DS} increase compared to $\Delta 3$ and $\Delta 4$ at same V_{GS}. Below V_{GS} = 0.7 V or so, $\Delta 1$ increases with V_{DS} while $\Delta 2$ decreases. Above 0.7 V, $\Delta 1$ becomes positive and further cancels out with negative $\frac{K3_{g_m}}{g_m}$. So, at high V_{GS}, $\Delta 1$ is the most important term in response to V_{DS} increase.

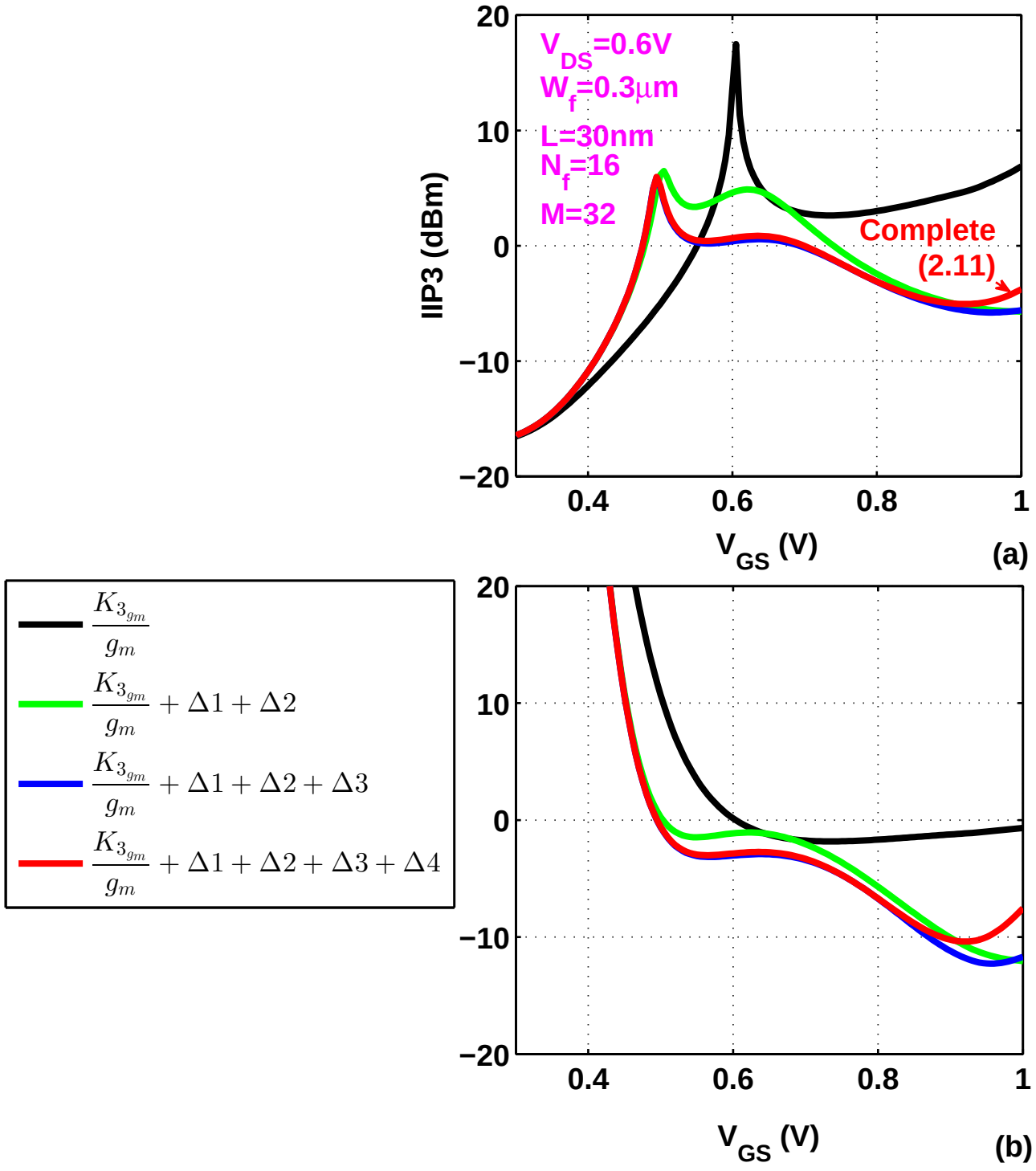

Figure 3.16: V_{GS} dependence of (a) IIP3 by (3.11) and (b) $\frac{K_{3_{g_m}}}{g_m} + \Delta$ with different Δ contributions.

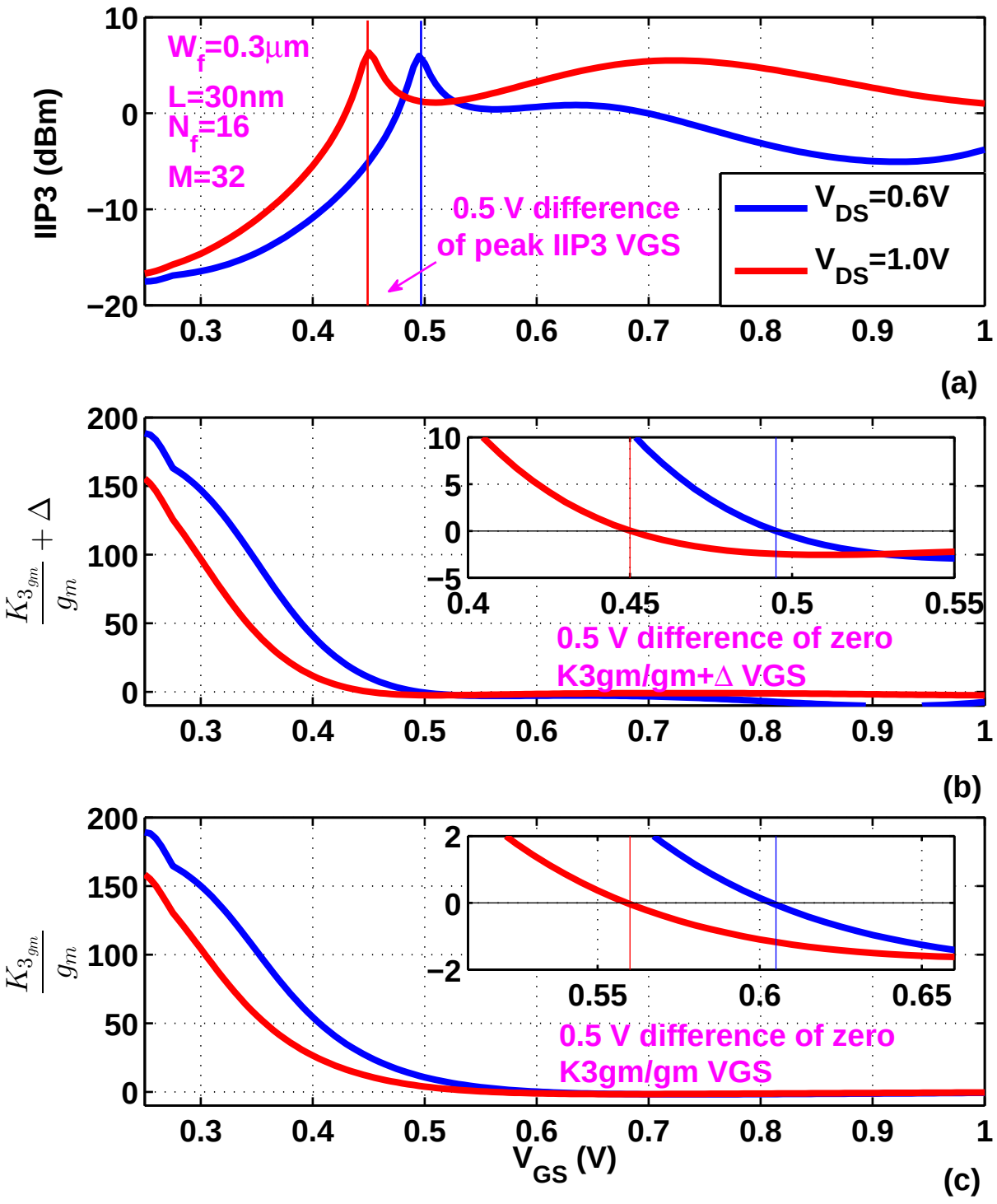

Figure 3.17: (a) $IIP3$ by (3.11) (b)$\frac{K_{3_{gm}}}{g_m}+\Delta$ and (c) $\frac{K_{3_{gm}}}{g_m}$ as a function of V_{GS} at $V_{DS}=0.6$ and 1.0 V.

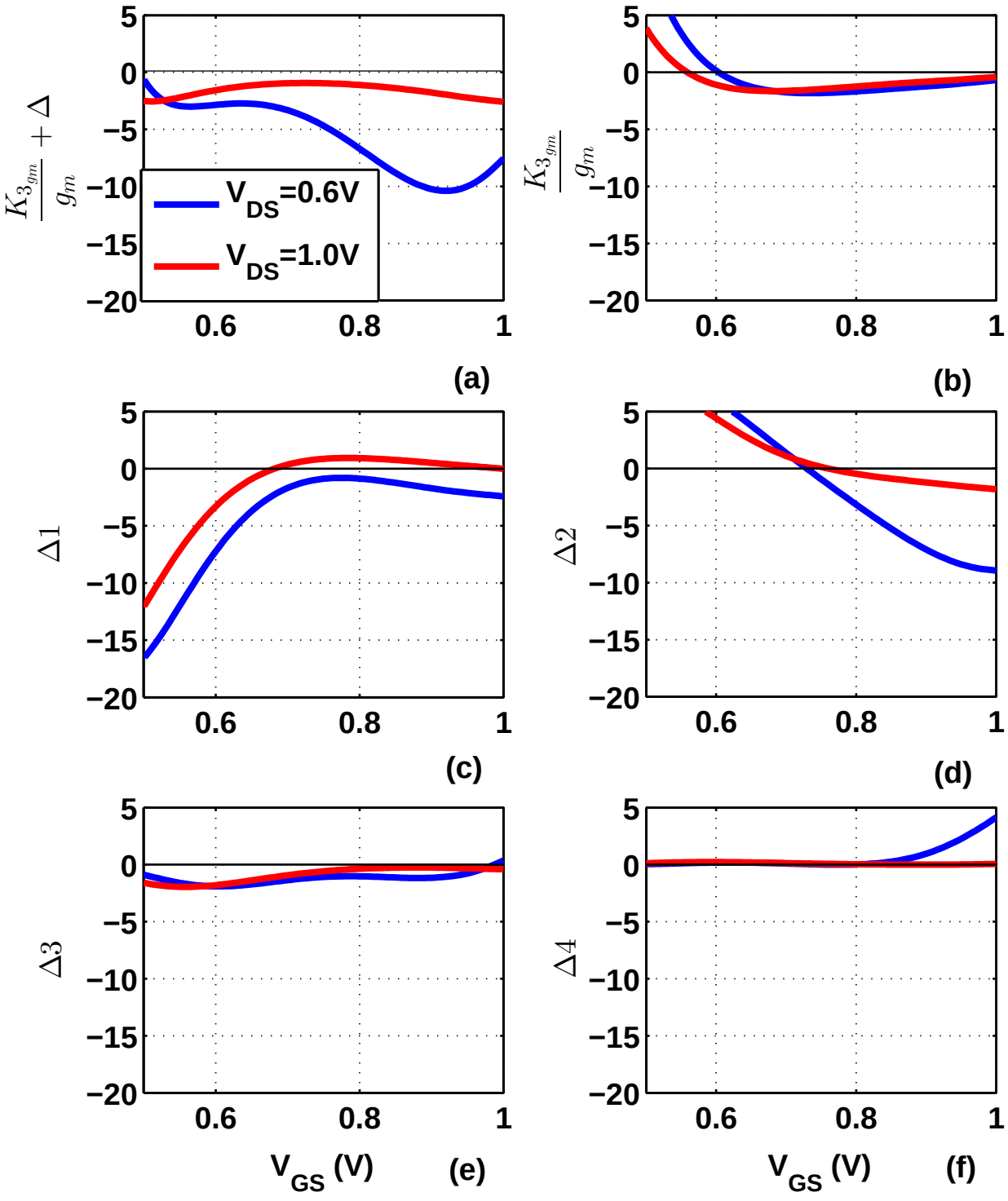

Figure 3.18: V_{GS} dependence of (a) $\frac{K_{3g_m}}{g_m} + \Delta$, (b) $\frac{K_{3g_m}}{g_m}$, (c) $\Delta 1$, (d) $\Delta 2$, (e) $\Delta 3$ and (f) $\Delta 4$ at $V_{DS} = 0.6$ and 1.0 V.

Chapter 4

CMOS model calibration

There are two 28nm high k/ metal gate RF CMOS technologies tested in this dissertation. The one for RF modeling has $BSIM$ core model, which will be discussed in chapter 5. The one for $IP3/IP2$ discussion in chapter 3 has PSP core model [22]. The initial parameter values are for base line digital CMOS transistors of the same technology and need to be tuned fit DC characteristics for tested devices.

Figure 4.1 compares measured I_{DS} versus V_{GS} with simulation using the original design kit at $V_{DS} = 0.6$ and 1.0 V for a 30nm device with a total width of 153.6μm. Gate finger width W_f is 0.3μm, number of finger N_f is 32, and multiplicity $M = 16$. Without DC tuning, the design kit totally overestimates drain current at both V_{DS}. In order to model intermodulation linearity on RF MOSFETs, netlist description and extrinsic components' parameter are optimized to accurately describe both DC and RF behavior using measured $I - V$ characteristics and S-parameters.

In this chapter, five key parameters for drain current fitting are introduced according to the compact model manual [22]. Then improved I_{DS}-V_{GS} and I_{DS}-V_{DS} are presented by tuning these model parameters to fit the measurement. With adjustment of gate resistance, reasonable Y-parameter fitting results are obtained.

4.1 Key PSP parameters to drain current fitting

The version number of the PSP compact model is 103.1. $PSP103$ is a surface potential based compact MOSFET model including important physical effects, such as mobility reduction, velocity saturation, drain-induced barrier lowering, gate current, etc. $PSP103$ can

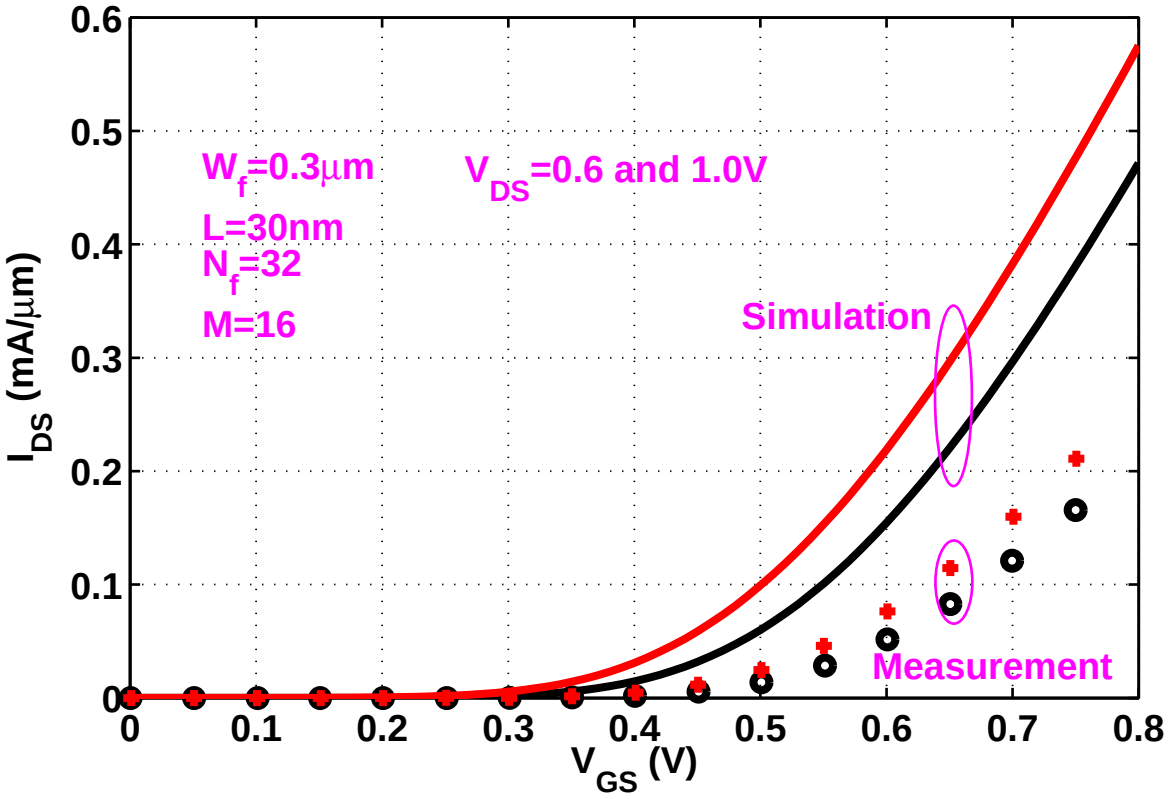

Figure 4.1: Comparison of simulated I_{DS} versus V_{GS} with measurement at V_{DS} = 0.6 and 1.0 V.

give an accurate description of currents, charges, and their first and higher order derivatives, which leads to an accurate description of electrical distortion behavior [22].

Table 4.1 lists five key parameters to account for sub-threshold, mobility and velocity saturation calculations. (4.1) through (4.5) show corresponding model equations without temperature scaling.

Model parameter	Affected region or effect	Description
CT	Sub-threshold	Interface states factor
$DPHIB$	Threshold voltage	Offset of ψ_B
CS	Weak inversion	Coulomb scattering parameter
MUE	Phonon scattering	Mobility reduction coefficient
$THESAT$	Strong inversion	Velocity saturation parameter

Table 4.1: Five key DC parameters in PSP.

$$CT = [CTO + CSL(\frac{L_{EN}}{L_E})^{CTLEXP}](1 + CTW\frac{W_{EN}}{W_E}) \tag{4.1}$$

$$\begin{aligned} DPHIB &= DPHIBO + DPHIBL(\frac{L_{EN}}{L_E})^{DPHIBLEXP} + \\ &\quad DPHIBW\frac{W_{EN}}{W_E} + DPHIBW\frac{W_{EN}L_{EN}}{W_EL_E} \end{aligned} \tag{4.2}$$

$$CS = [CSO + CSL(\frac{L_{EN}}{L_E})^{CSLEXP}](1 + CSW\frac{W_{EN}}{W_E}) \tag{4.3}$$

$$MUE = MUEO(1 + MUEW)\frac{W_{EN}}{W_E} \tag{4.4}$$

$$\begin{aligned} THESAT &= [THESATO + THESATL(\frac{G_{W,E}}{G_{P,E}}\frac{L_{EN}}{L_E})^{THESATLEXP}] \\ &\quad (1 + THESATW\frac{W_{EN}}{W_E})(1 + THESATLW\frac{W_{EN}L_{EN}}{W_EL_E}) \end{aligned} \tag{4.5}$$

Meanings of the various symbols can be found in [22]. In Figure 4.2, the effective regions in which the five key parameters have a big impact are highlighted by using I_{DS} versus V_{GS} on both linear and log scales. For sub-threshold region, CT determines the slope on log scale. $DPHIB$ affects threshold voltage and body effect. CS describes Coulomb scattering in moderate inversion. MUE is responsible for phonon scattering in strong inversion. $THESAT$ is important to fit drain current at very high V_{GS}.

4.2 Improved DC and Y-parameter fittings

With the observation above, DC tuning is done with many trials. Figure 4.3 compares simulated I_{DS} versus V_{DS} with measurement at V_{GS} = 0.4, 0.6 and 0.8 V for the device in Figure 4.1. For each V_{GS}, good agreement of simulated I_{DS} is achieved compared to measurement. For the same device in Figure 4.3, Figure 4.4(a) and (b) compare simulated I_{DS} versus V_{GS} with measurement using linear and log I_{DS} scales, respectively.

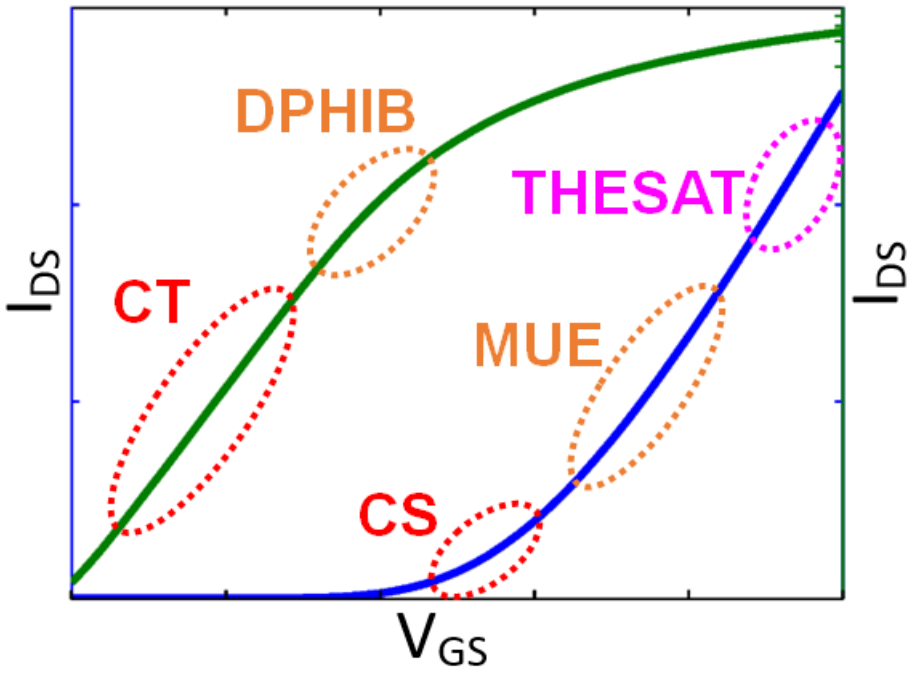

Figure 4.2: Effective regions for model parameters in affecting drain current.

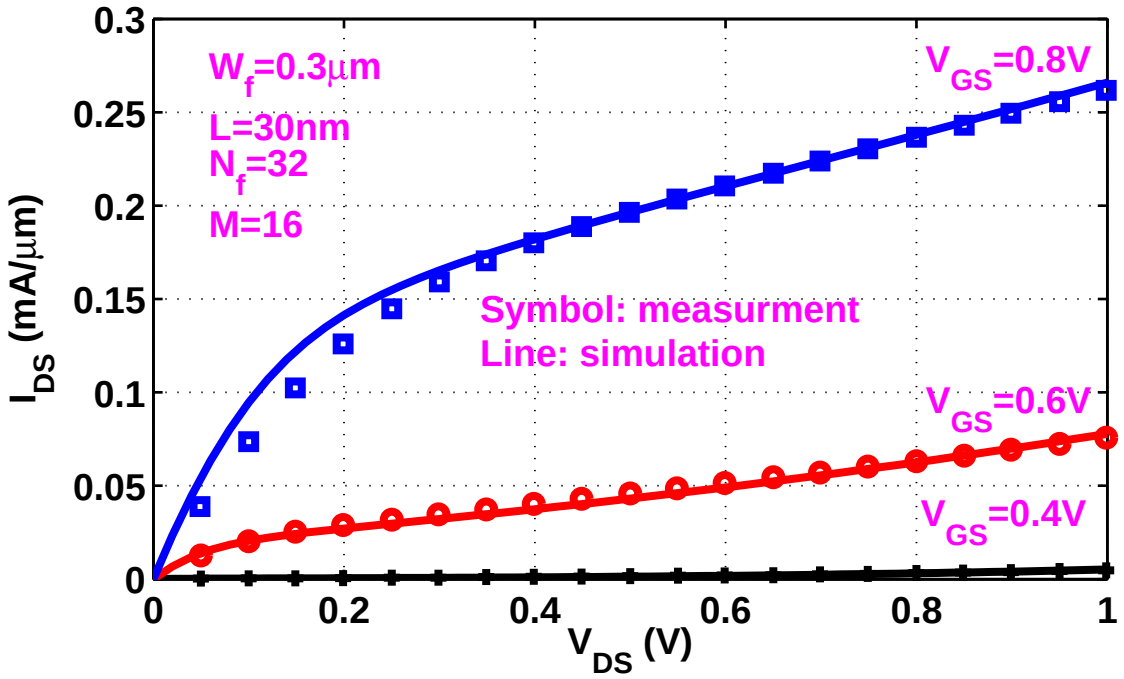

Figure 4.3: Comparison of simulated I_{DS} versus V_{DS} with measurement at V_{GS} = 0.4, 0.6 and 0.8 V.

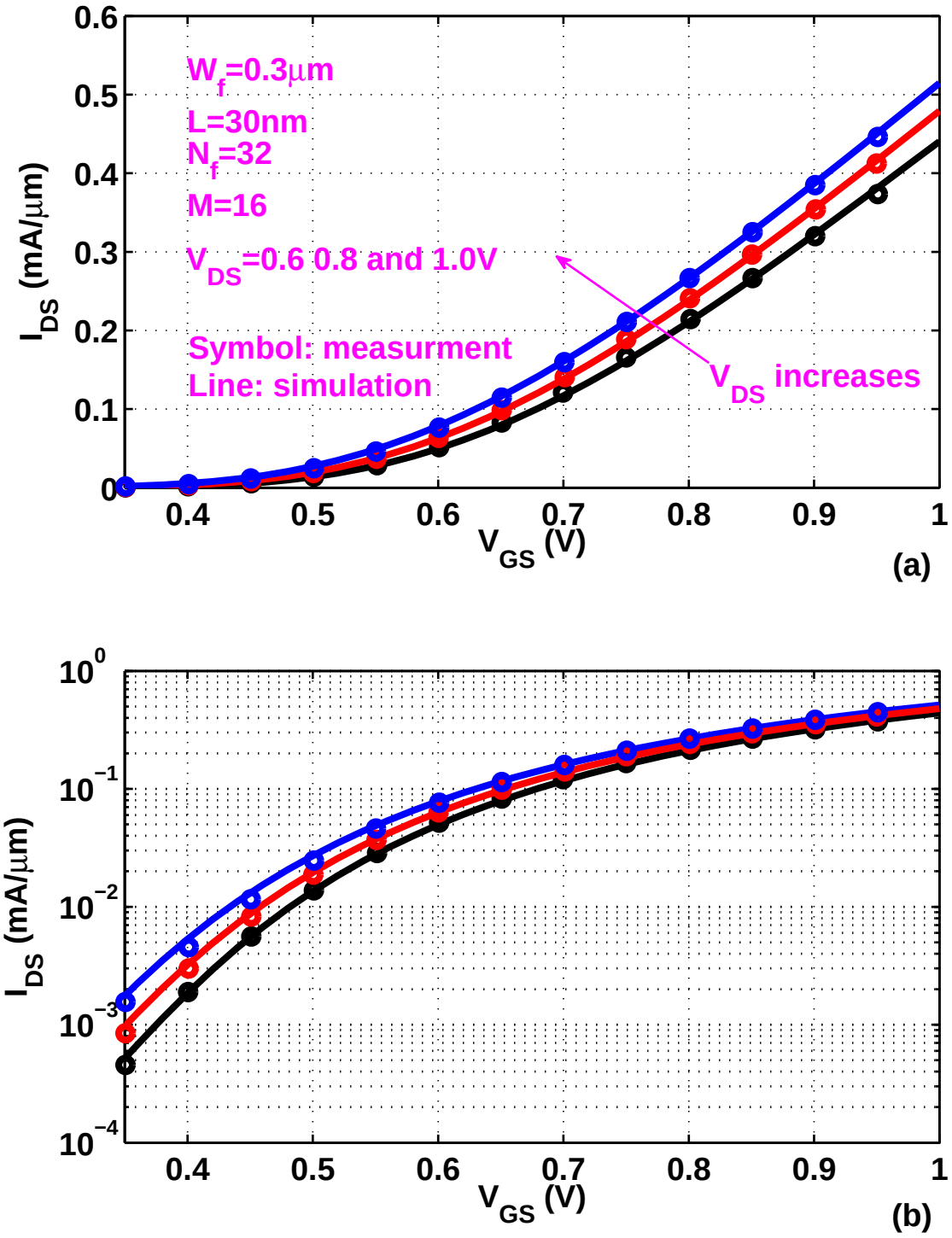

Figure 4.4: Comparison of simulated I_{DS} versus V_{GS} with measurement on (a) linear and (b) log I_{DS} scales at $V_{DS} = 0.6$, 0.8 and 1.0 V.

For Y-parameter fitting, extrinsic gate resistance is found to be insufficient and therefore is increased. For the same device in Figure 4.3, Figure 4.5 (a), (b), (c) and (d) compare simulated Y11, Y21, Y22 and Y12 as a function of frequency at $V_{GS} = 0.6$ V and $V_{DS} = 0.6$ V. For the frequency of interest (below 10 GHz), the improved design kit is able to provide reasonable fitting result.

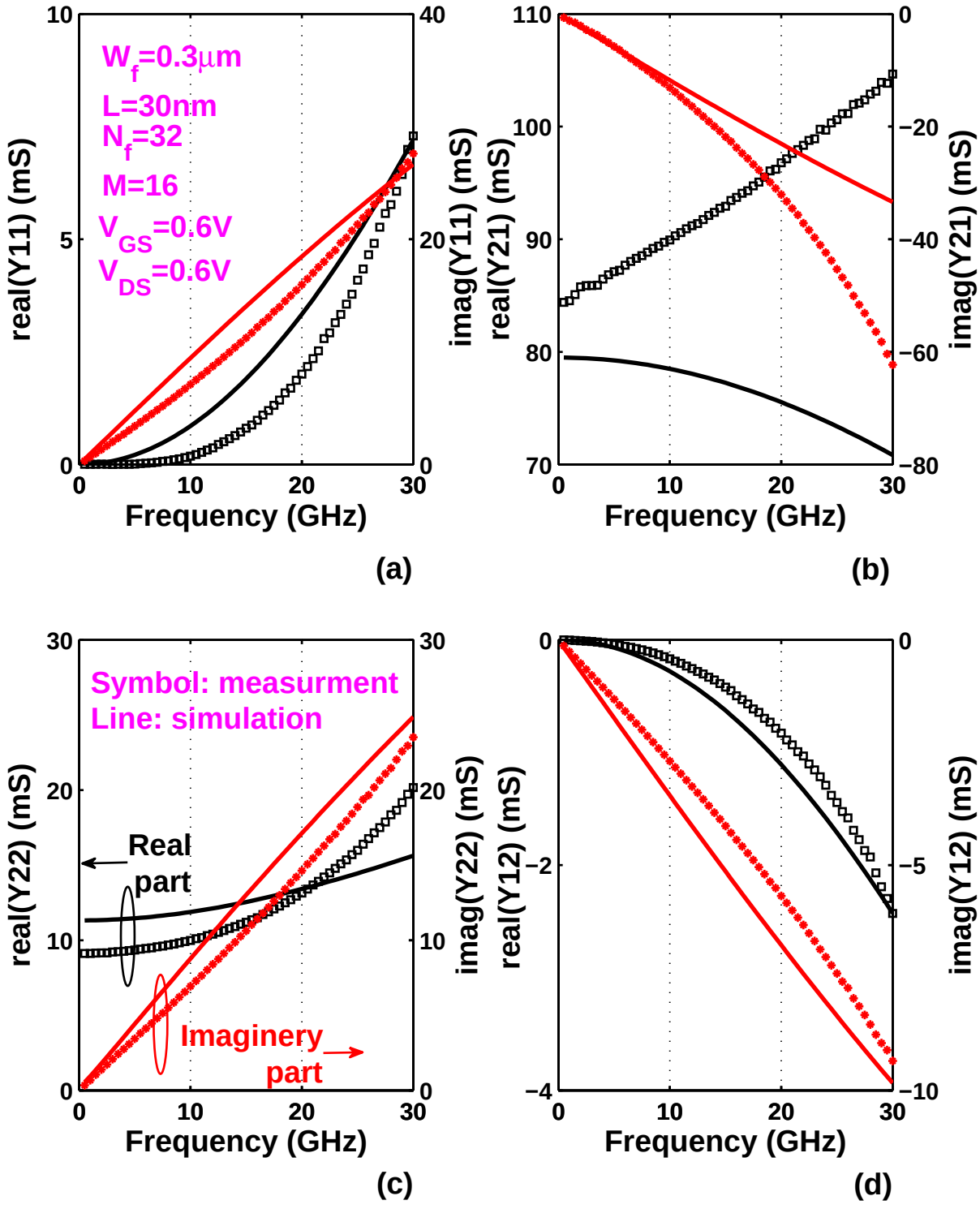

Figure 4.5: Comparison of simulated (a) Y11 (b) Y21 (c) Y22 and (d) Y12 versus frequency at $V_{GS} = 0.6$ V and $V_{DS} = 0.6$ V.

Chapter 5

Y22 modeling

High frequency (HF) modeling of MOSFET is more challenging compared to low frequency application, because it requires accurate prediction of bias dependence of small-signal parameters, correct description of nonlinear behavior, reasonable resistance networks, a bias dependent overlap capacitance model, and HF noise model [23].

In this chapter, both frequency and bias dependence of $real(Y22)$ of 28nm HKMG RF MOSFETs are shown to deviate from measurement data. According to the derived $real(Y22)$ expression, the body resistance and the capacitance between drain and body nodes are identified to help improve frequency dependence of $real(Y22)$ and therefor are tuned in the same time. A technique to normalize bias dependence at all frequency is proposed. Back gate transconductance is found to be responsible for weak bias dependence of $real(Y22)$ and should be strengthened. By increasing the body resistance and the capacitance between drain and body nodes, $Y22$ fitting is improved considerably.

5.1 $real(Y22)$ problem

Figure 5.1 compares measured frequency dependence of $real(Y22)$ with simulation using the original design kit at different V_{GS} for a 30nm device. V_{GS} = 0, 0.2, 0.4, 0.6, 0.8 and 1.0 V. V_{DS} = 0.6 V. The device has a drawn gate length of 30nm with a total width of 8 μm. $W_f = 1\mu$m. $N_f = 8$. Multiplicity $M = 1$. For all biases, simulated $real(Y22)$ fails to respond to frequency increase. At high frequency, the difference between simulation and measurement is even larger compared to low frequency. Body resistance is initially raised to increase $real(Y22)$ to some extent. In simulation, this stops working at high R_b values.

So key components in the design kit need to be identified and tuned to increase simulated $real(Y22)$.

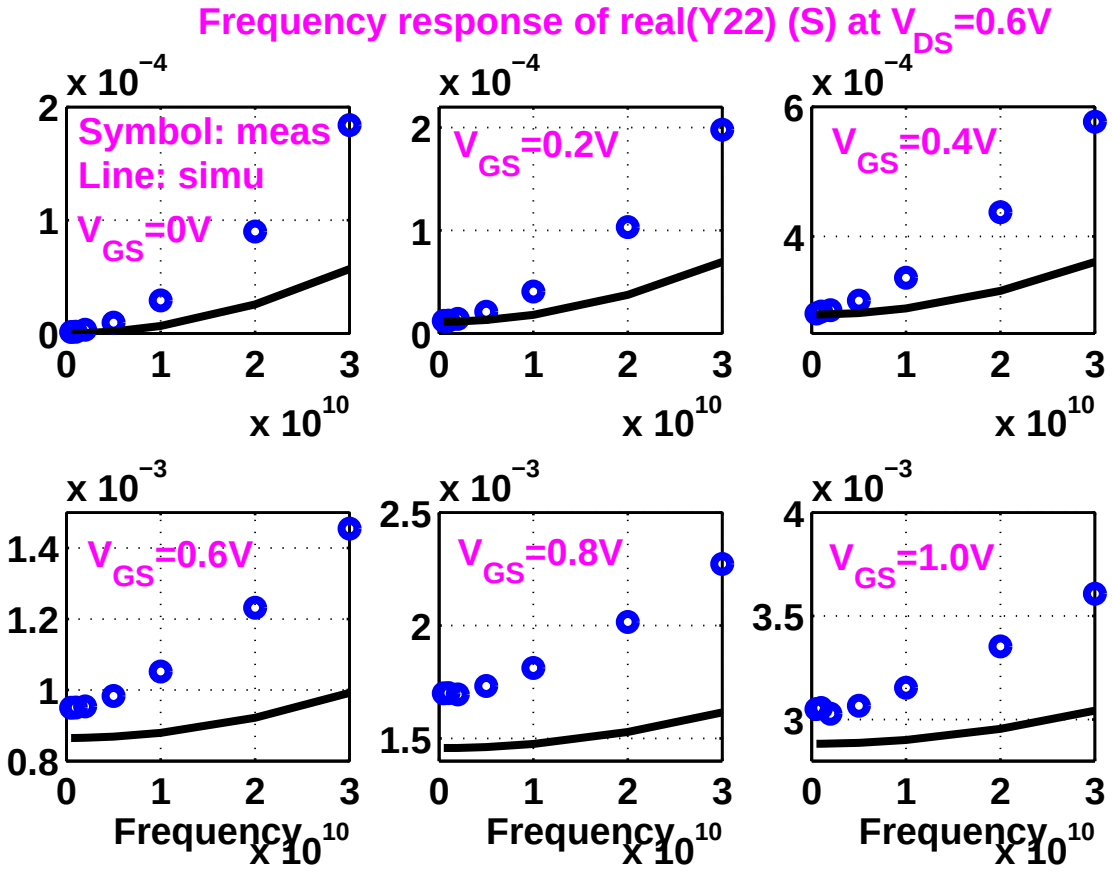

Figure 5.1: Comparison of simulated $real(Y22)$ as a function of frequency with measurement for a 30nm MOSFET at V_{GS} = 0, 0.2, 0.4, 0.6, 0.8 and 1.0 V. V_{DS} = 0.6 V.

5.2 DNW model

RF MOSFET structure has a dedicated region to suppress noise cross talk between neighboring body regions [24]. For nMOSFET, it is called deep N-well (DNW). Figure 5.2 shows the diagram for RF nMOSFET. N-type DNW, together with N-well, keeps P-type body regions separated one another. There are with six terminals in total: source (S), gate (G), dain (D), body (B), DNW and p-type substrate (PSUB). Shallow trench (STI) is used to isolate P-type regions near Ohmic contact from N-type regions.

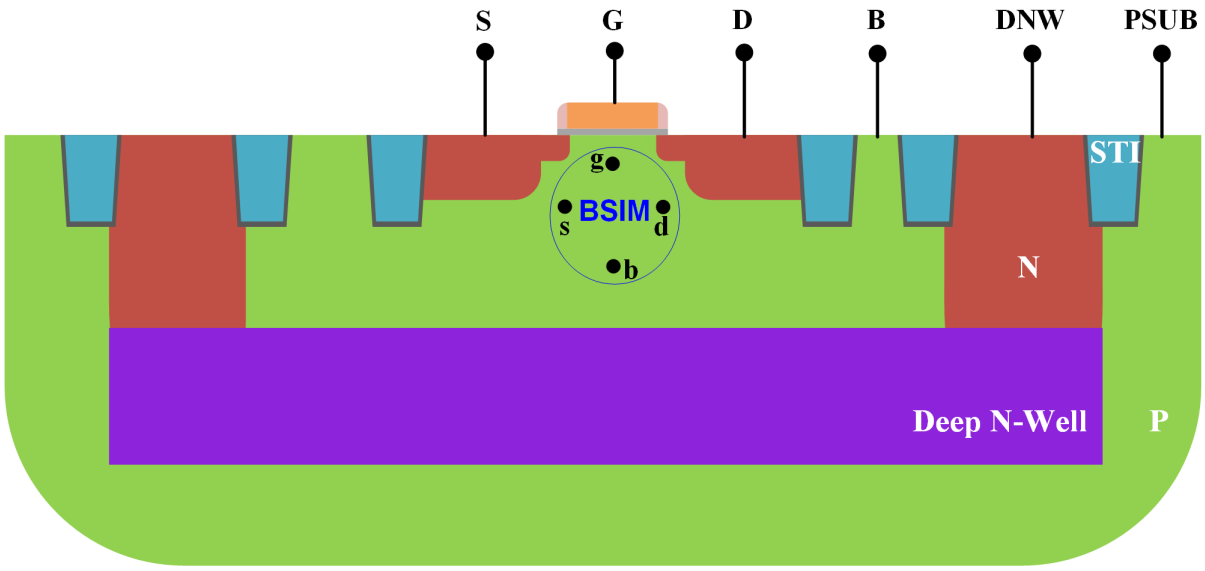

Figure 5.2: RF nMOSFET structure.

As highlighted by the blue circle, $BSIM$ is used as core model in the design kit to describe four-terminal intrinsic transistor behavior. Internal nodes are denoted with lower case letters: g, s, b and d. The $BSIM$ compact model used has a version number of 4.6.2. Compared to previous versions, $BSIM4.6.2$ has a unique strength to account for Coulombic scattering in high-k/metal gate transistors by adding a new option ($mobmod = 3$) in mobility model [25].

Figure 5.3 illustrates the implementation of DNW model in netlist. Extrinsic components with subscript of "$\_ex$" are added outside of $BSIM$ core model. R_{g_ex}, R_{s_ex}, R_{b_ex}, R_{d_ex}, R_{dnw_ex} and R_{pw_ex} are the parasitic resistance between internal nodes and external terminals. C_{gs_ex}, C_{gd_ex}, C_{sb_ex} and C_{db_ex} account for parasitic capacitance that the intrinsic $BSIM$ model does not cover. Two more diodes are reserved to describe the P-N junctions that DNW forms with body region (P-well) and P-type substrate. These parasitics can be optimized to minimize the difference between measurement and simulation.

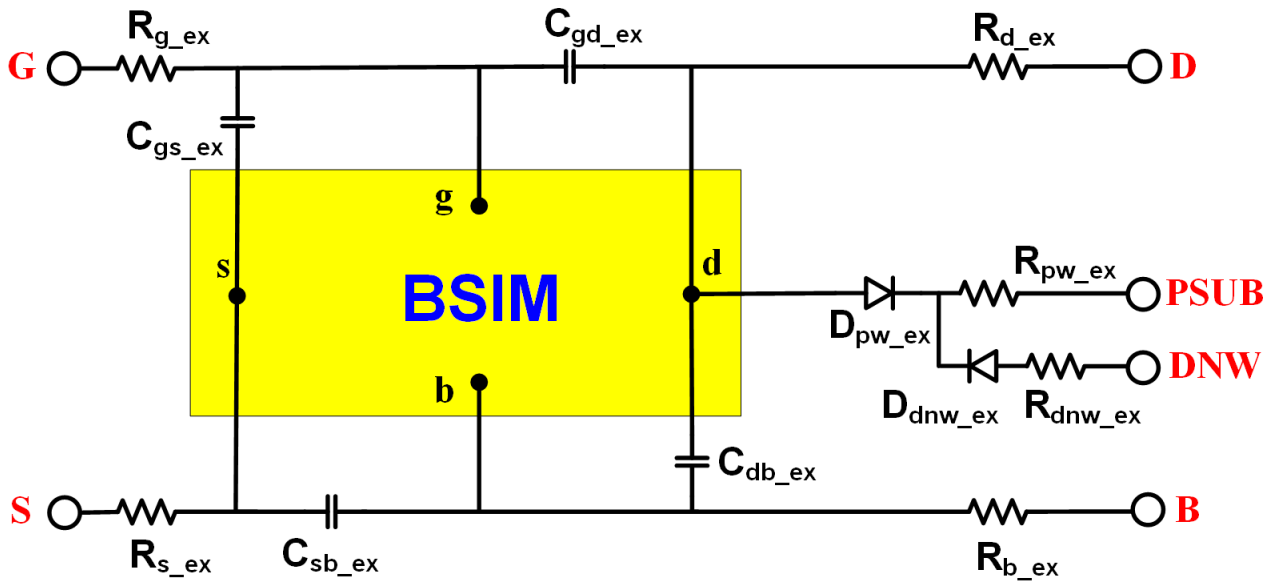

Figure 5.3: Implementation illustration of DNW model.

5.3 Expression of $real(Y22)$

Figure 5.4 shows an equivalent circuit for $(Y22)$. Previous six terminal case in Figure 5.3 is simplified to four terminal one. It is justified by the fact that DNW and PSUB are tied together. We have verified that the variation of parasitics related to these two terminals has a negligible impact on $real(Y22)$.

Following the derivation in [26], $real(Y22)$ can be expressed as:

$$\begin{aligned} real(Y22) &= g_{ds} + \\ &\omega^2 \frac{R_g C_{gd} C_{dg} + g_m R_g^2 C_{gd}(C_{gs} + C_{gd})}{1 + \omega^2 R_g^2 (C_{gs} + C_{gd})^2} + \\ &\omega^2 \frac{R_b C_{db}^2 + g_{mb} R_b^2 C_{db}(C_{sb} + C_{db})}{1 + \omega^2 R_b^2 (C_{sb} + C_{db})^2} \end{aligned} \tag{5.1}$$

Output conductance g_{ds} is a frequency-independent term. We have tested that the second term, $\omega^2 \frac{R_g C_{gd} C_{dg} + g_m R_g^2 C_{gd}(C_{gs}+C_{gd})}{1+\omega^2 R_g^2 (C_{gs}+C_{gd})^2}$ has a relatively weak impact on $real(Y22)$ compared to the third term $\omega^2 \frac{R_b C_{db}^2 + g_{mb} R_b^2 C_{db}(C_{sb}+C_{db})}{1+\omega^2 R_b^2 (C_{sb}+C_{db})^2}$.

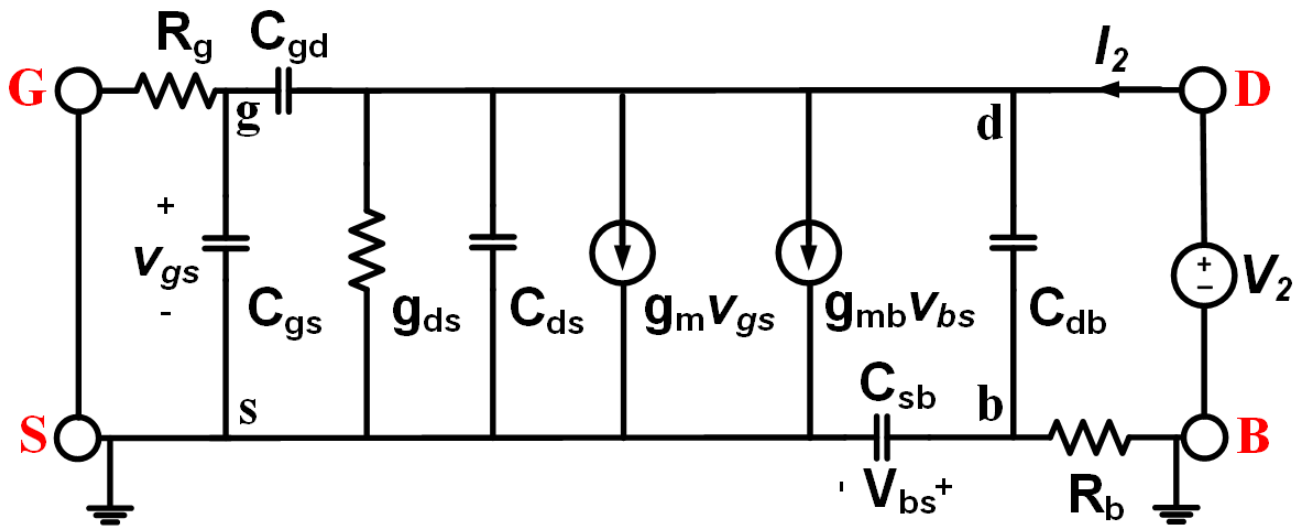

Figure 5.4: Equivalent circuit for $(Y22)$.

Figure 5.5 shows a simplified equivalent circuit used for $real(Y22)$ discussion. In addition to g_{ds}, the capacitance between drain and body (C_{db}), the capacitance between source and body (C_{sb}), and body resistance (R_b) are kept. Therefore,

$$real(Y22) \approx g_{ds} + \omega^2 \frac{R_b C_{db}^2 + g_{mb} R_b^2 C_{db}(C_{sb} + C_{db})}{1 + \omega^2 R_b^2 (C_{sb} + C_{db})^2} \tag{5.2}$$

According to (5.2), the frequency dependence are mainly affected by R_b and C_{db}. In particular, C_{db} has a square law impact so that a small increase of C_{db} helps $real(Y22)$ fast respond to frequency increase. C_{sb} may help, too, at large bias.

5.4 V_{GS} dependence

We propose the term $\frac{real(Y22)-g_{ds}}{\omega^2}$ to normalize bias dependence of $real(Y22)$ for different frequency. Figure 5.6(a) compares measured V_{GS} dependence of $\frac{real(Y22)-g_{ds}}{\omega^2}$ with simulation using the original kit for the same device in Figure 5.1 at 5, 10, 20 and 30 GHz. $V_{DS} =$ 0.6 V. Figure 5.6(b) shows simulated g_{mb} versus V_{GS} at the same V_{DS}. Observe that simulated $\frac{real(Y22)-g_{ds}}{\omega^2}$ at different frequency lags behind of measurement data to respond to V_{GS}

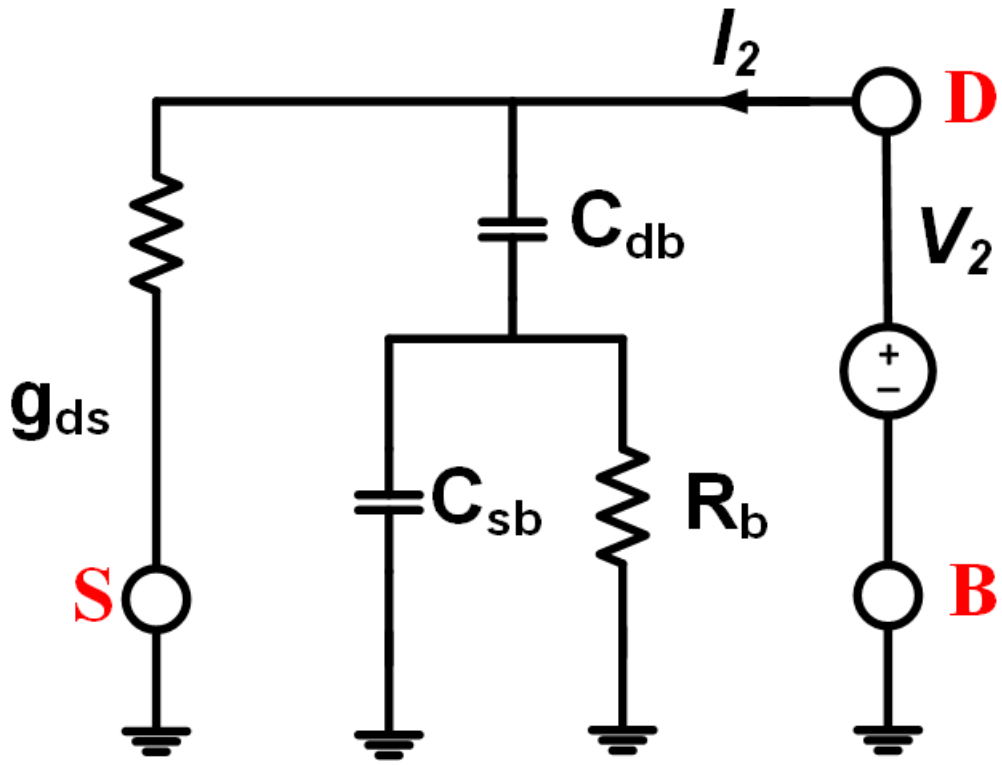

Figure 5.5: Simplified equivalent circuit used for $real(Y22)$ discussion.

increase and tracks simulated V_{GS} dependence of g_{mb}. This indicates g_{mb} produces extra increase of $real(Y22)$. The bias dependence of g_{mb} should be improved.

5.5 Improved modeling

With the observations mentioned above, we know increasing R_b together with C_{db} will improve frequency dependence of $real(Y22)$. At the same time, V_{GS} dependence of $real(Y22)$ will also be improved due to equivalently strengthened V_{GS} dependence of g_{mb} according to (5.2).

Figure 5.7 compares measured V_{GS} dependence of $real(Y22)$ $\frac{real(Y22)-g_{ds}}{\omega^2}$ and $real(Y22)$ with simulations using both the original and modified kits. Frequency= 5, 10 and 30 GHz. R_b and C_{jd} are increased by 1629 Ω and 1 fF. At both low and high V_{GS}, the improved kit is able to predict measurement data.

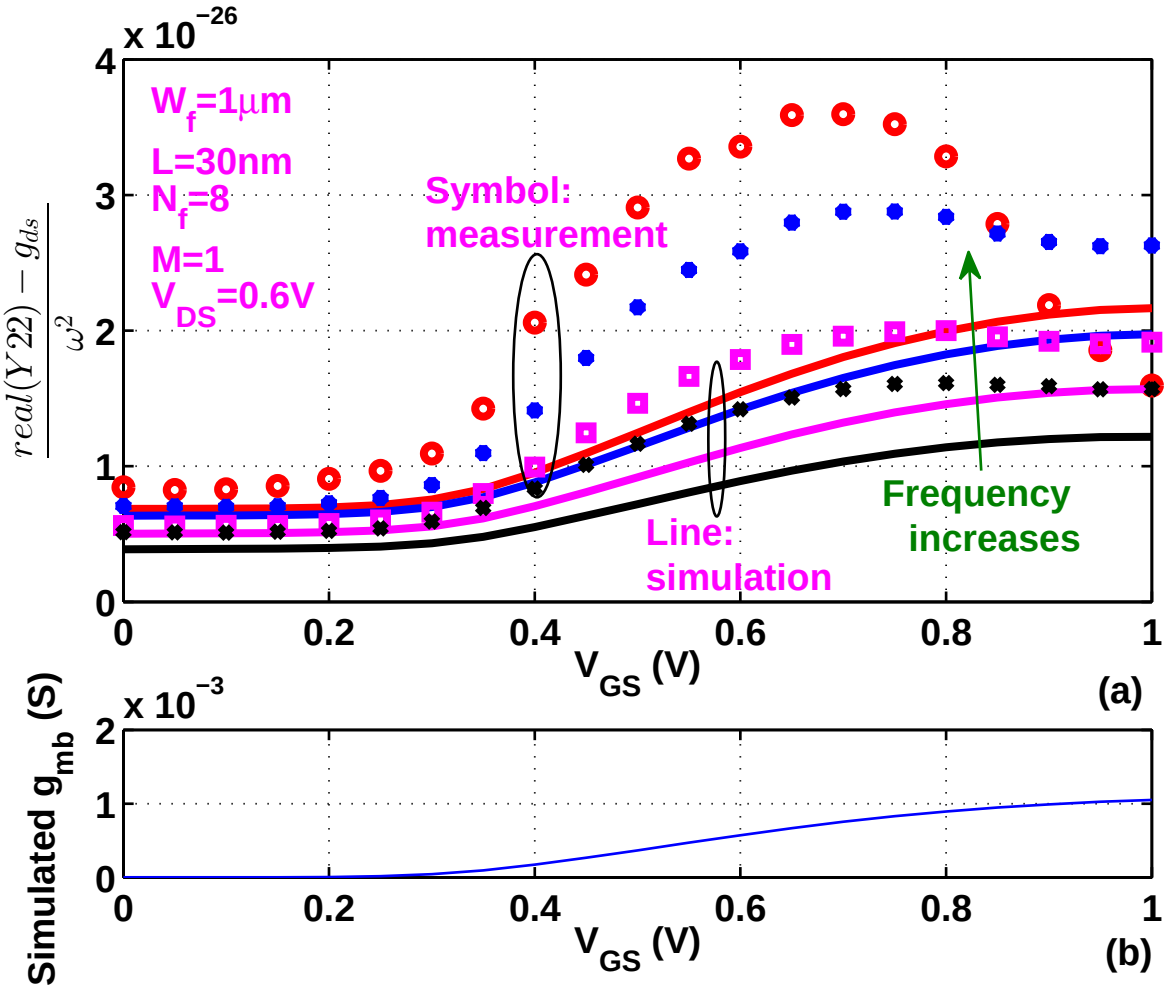

Figure 5.6: (a) Comparison of simulated V_{GS} dependence of $\frac{real(Y22)-g_{ds}}{\omega^2}$ with measurement result at 5, 10, 20 and 30 GHz. (b)Simulated g_{mb} as a function V_{GS} using the original kit.

Figure 5.8 - Figure 5.13 compare measured frequency dependence of Y-parameters with simulations using both the original and modified kits. V_{GS} = 0, 0.2, 0.4, 0.6, 0.8 and 1.0 V. V_{DS} = 0.6 V. For all V_{GS} values, the improved kit greatly improves $real(Y22)$ fitting without sacrificing other Y-parameters.

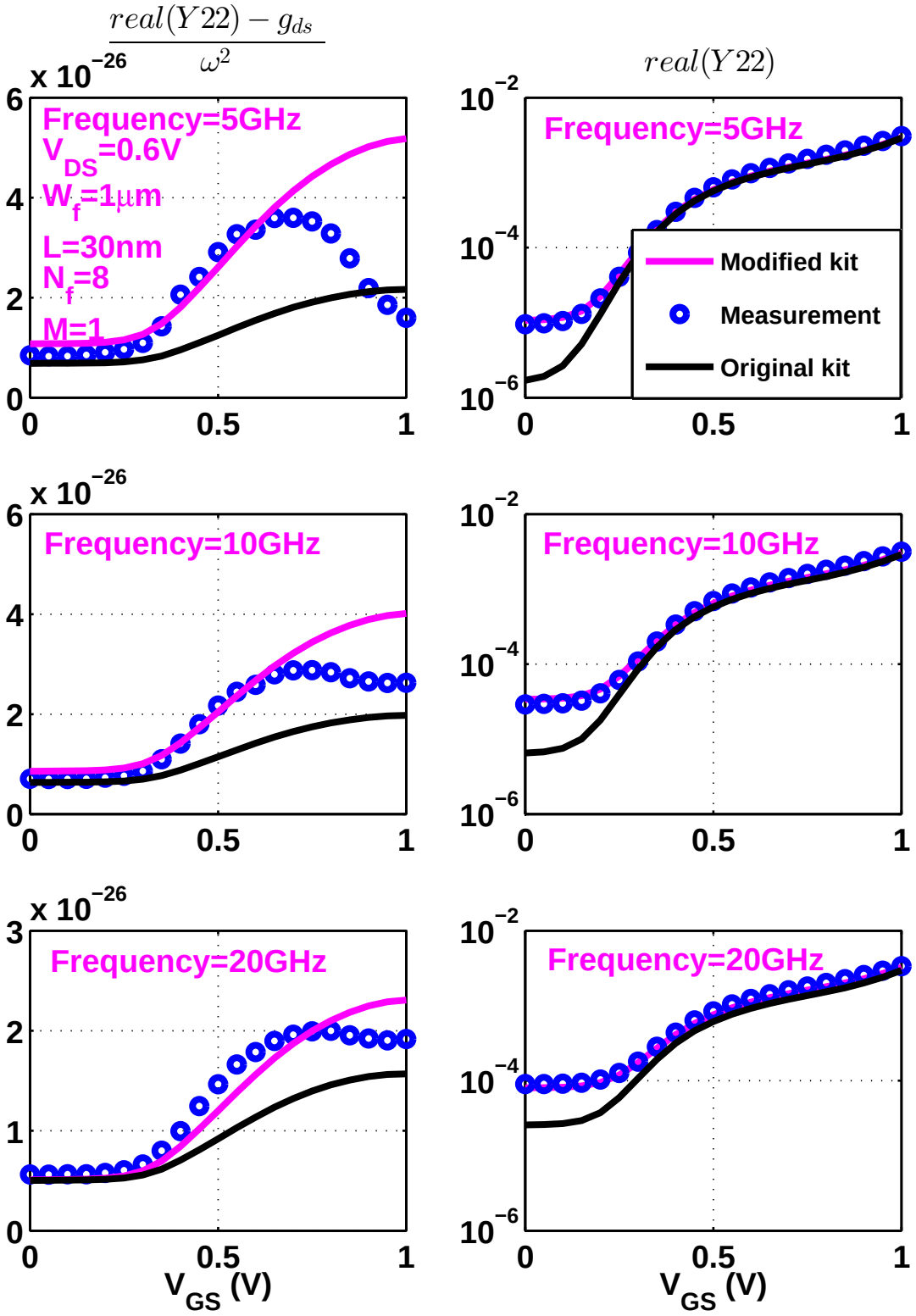

Figure 5.7: Comparisons of simulated V_{GS} dependences of $\frac{real(Y22)-g_{ds}}{\omega^2}$ and $real(Y22)$ using both the original and modified kits with measurement.

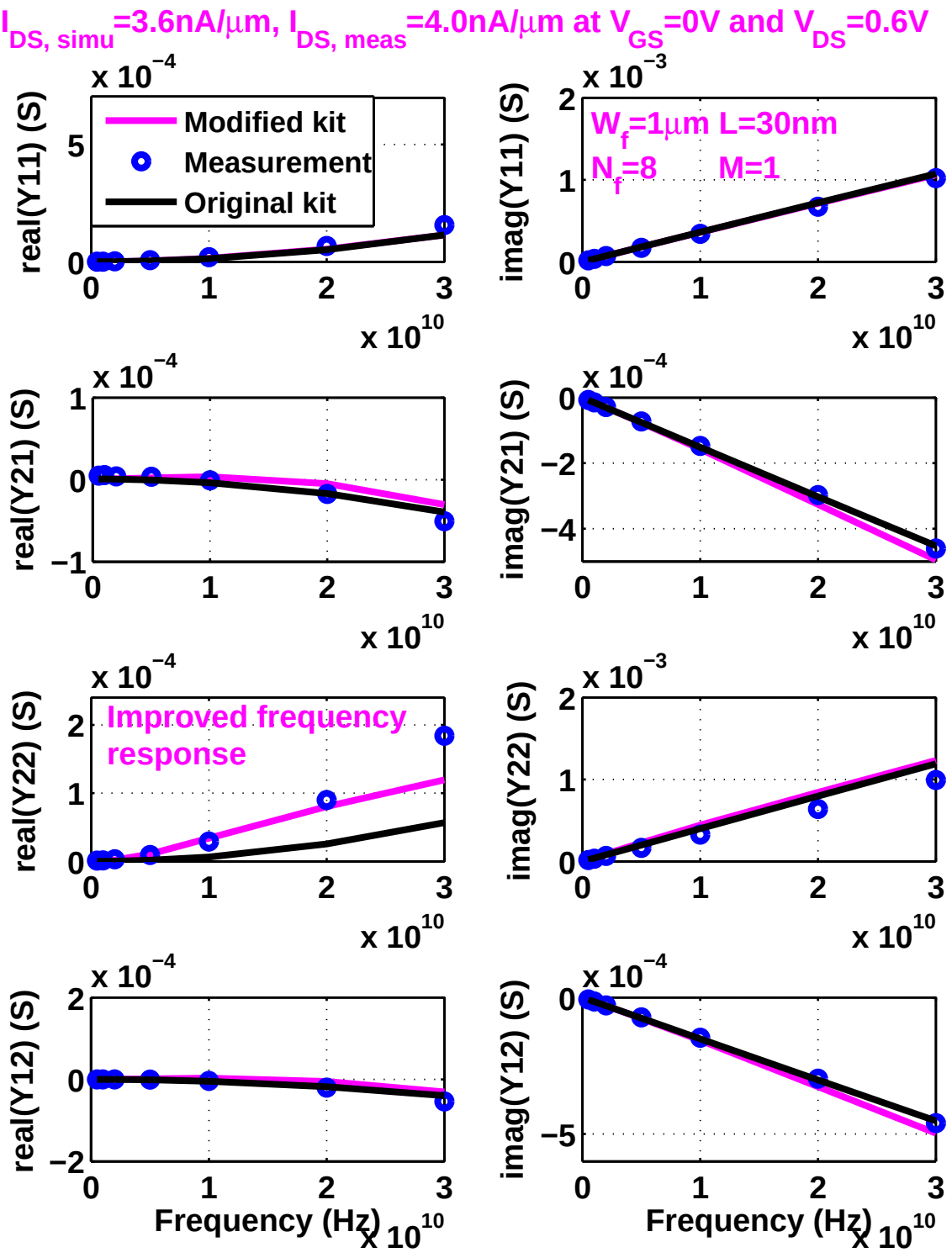

Figure 5.8: Comparison of simulated frequency dependence of Y-parameters using both modified and original kits with measurement. $V_{GS} = 0$ V. $V_{DS} = 0.6$ V.

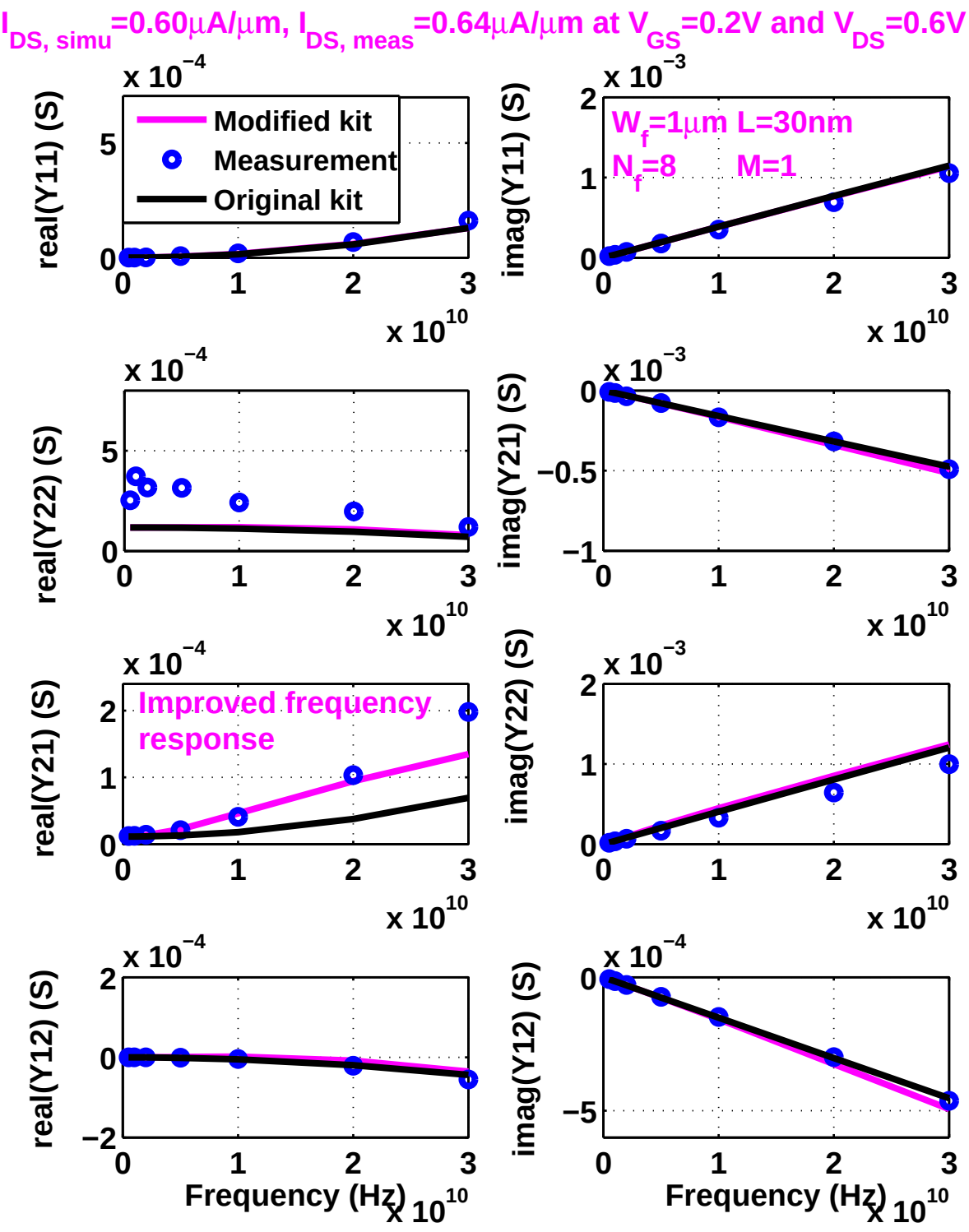

Figure 5.9: Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 0.2$ V. $V_{DS} = 0.6$ V.

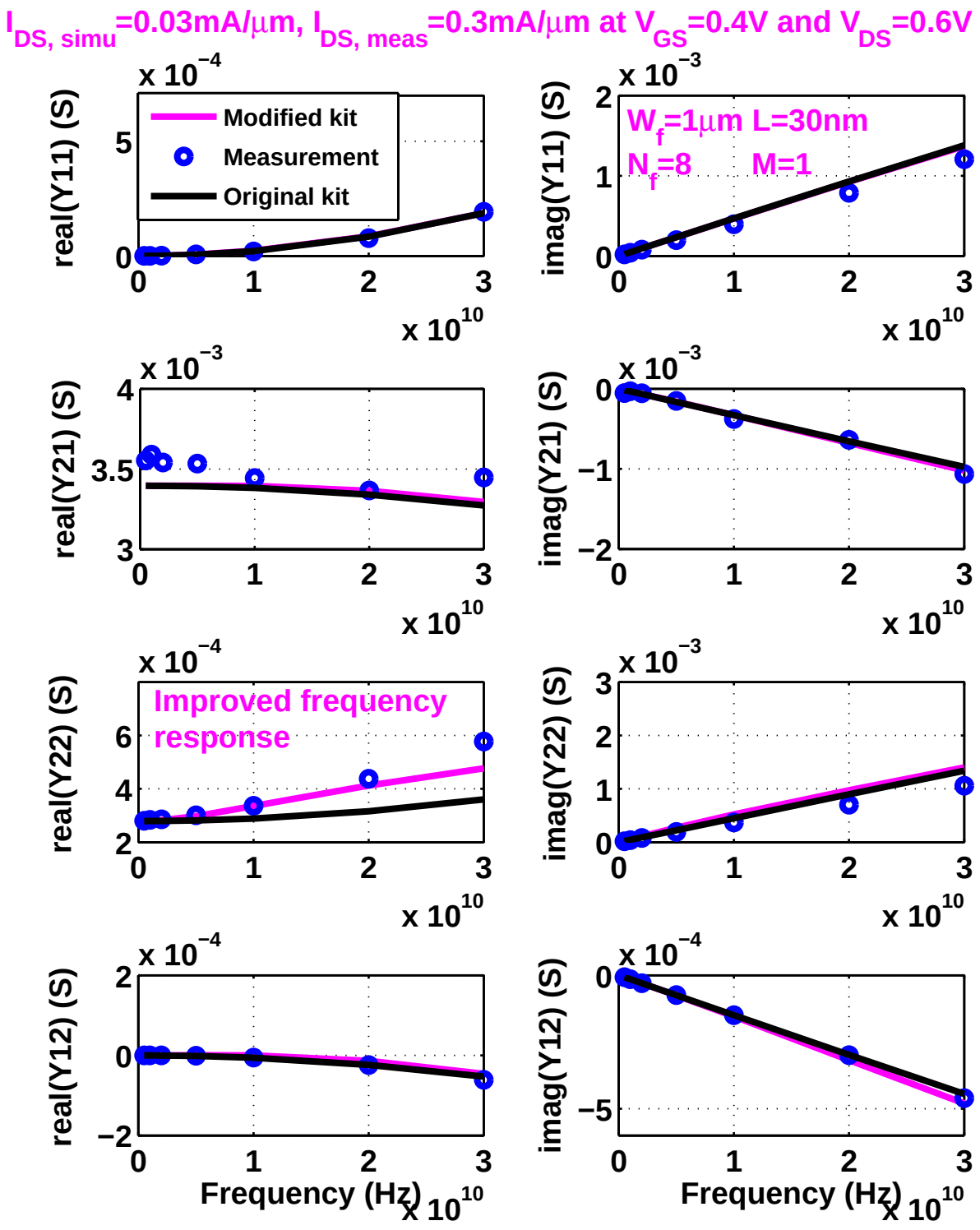

Figure 5.10: Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 0.4$ V. $V_{DS} = 0.6$ V.

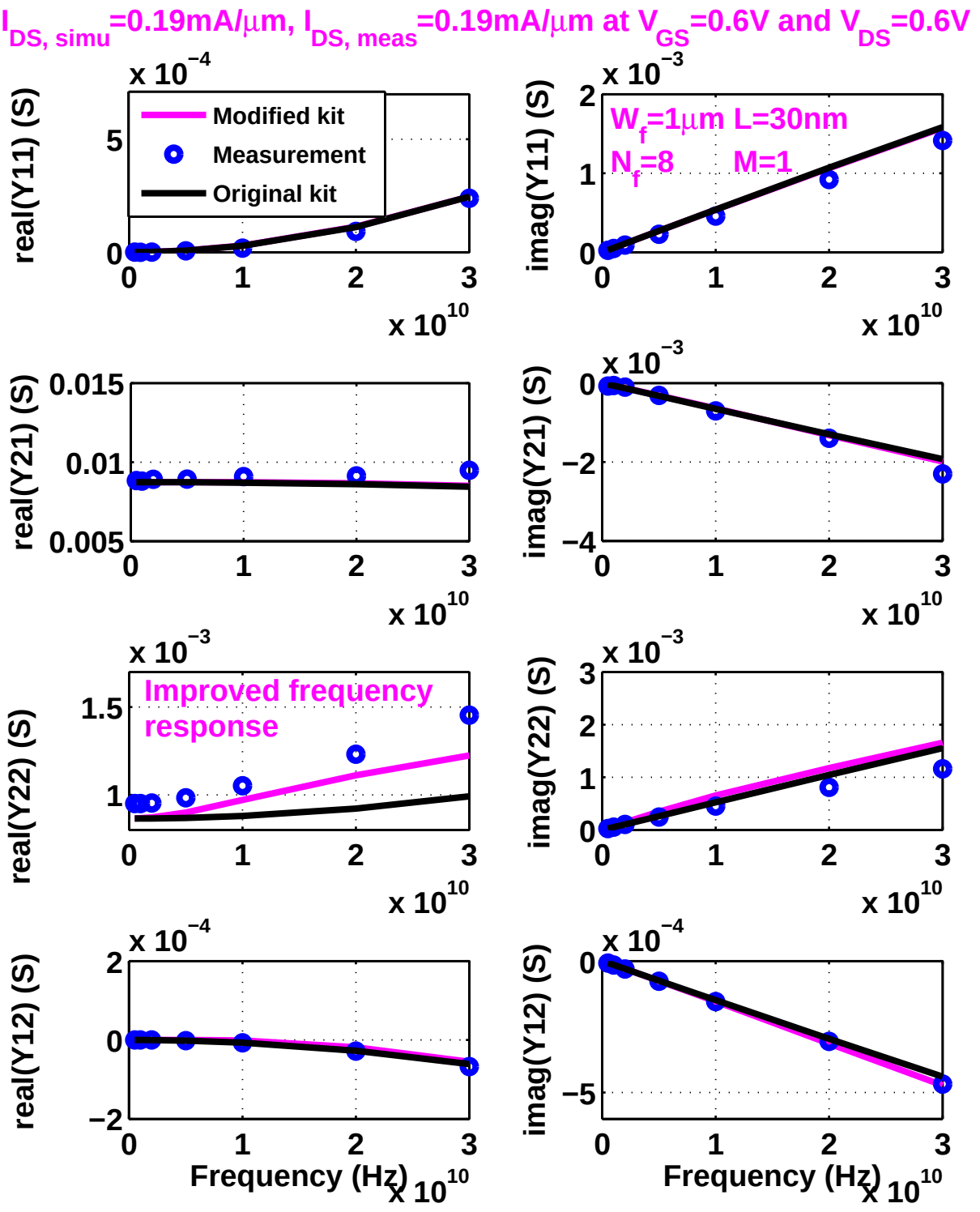

Figure 5.11: Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 0.6$ V. $V_{DS} = 0.6$ V.

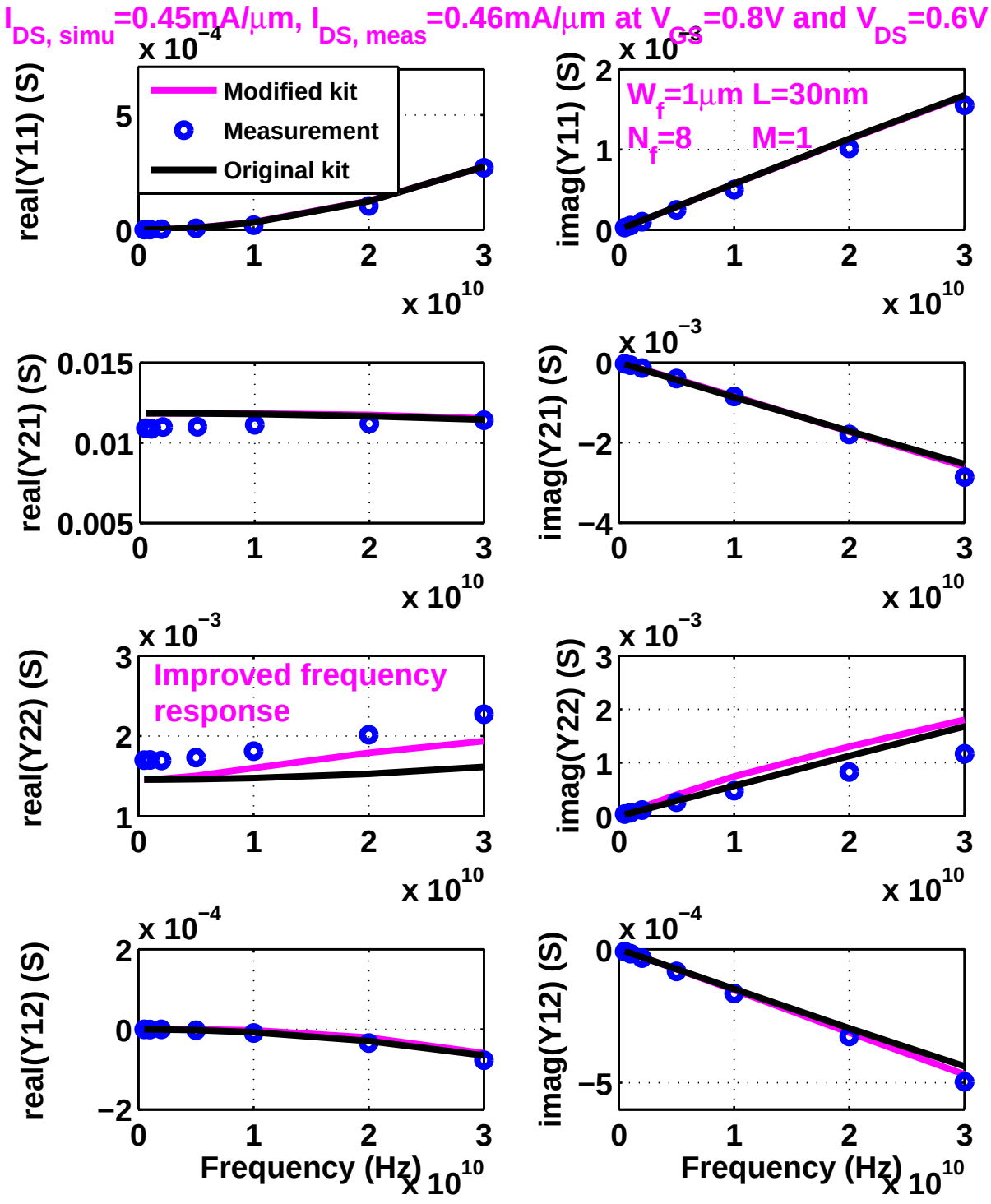

Figure 5.12: Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 0.8$ V. $V_{DS} = 0.6$ V.

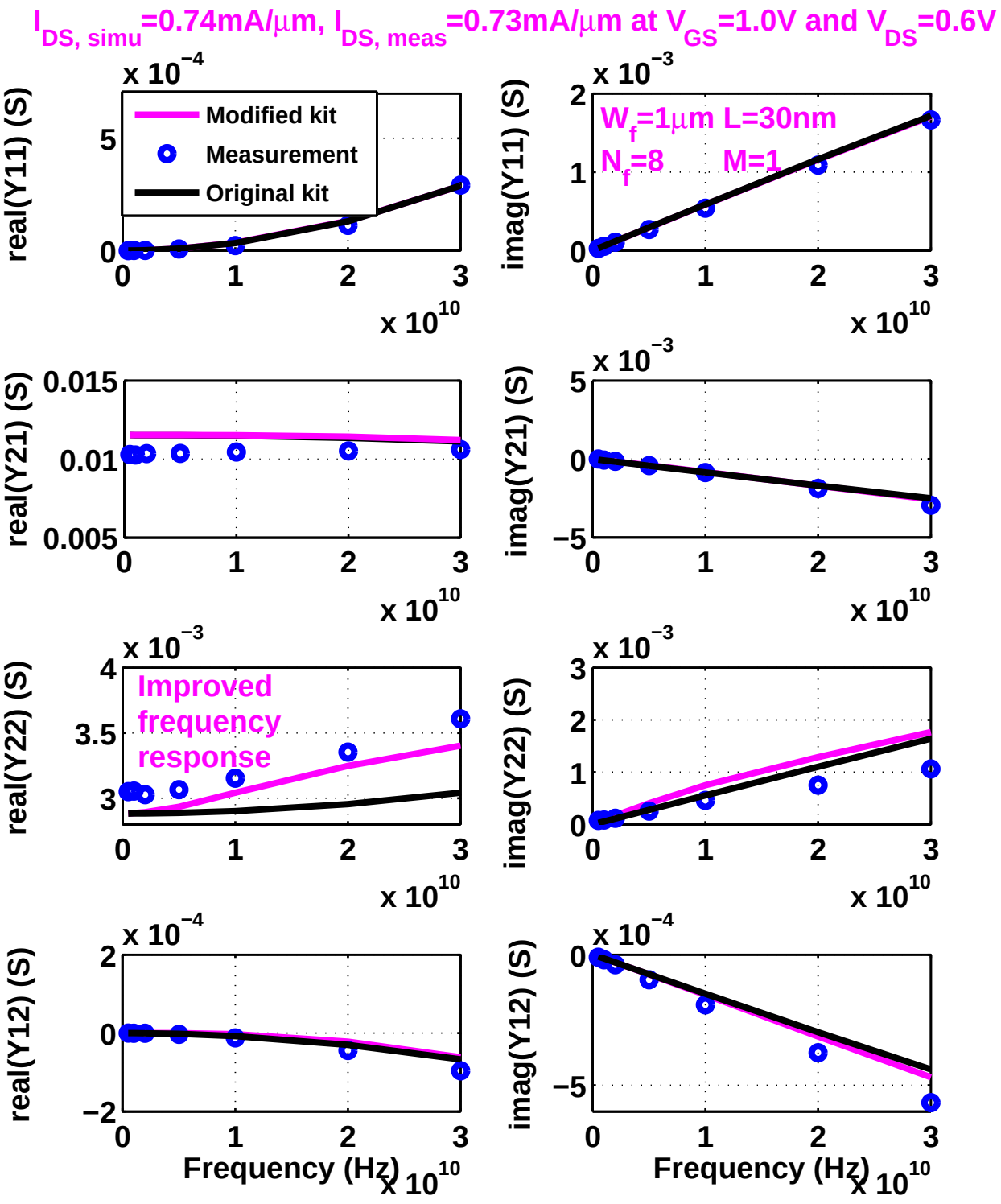

Figure 5.13: Comparison of simulated Y-parameters using both modified and original kits with measurement. $V_{GS} = 1.0$ V. $V_{DS} = 0.6$ V.

Chapter 6

IP2 of 200 GHz SiGe HBT

In this chapter, the input second order intercept point, $IIP2$, is characterized on $I_C - V_{CE}$ plane for SiGe HBTs with different breakdown voltage. In addition to bias impact, the influence of emitter size on $IIP2$ is also explored. By using a simplified small signal model, current dependence of $IP2$ is discussed.

6.1 Analytical model

As detailed in [27], there are many physical nonlinearities in a bipolar transistor (BJT), including transport current I_{CE}, emitter injection current I_B, junction capacitances C_{BE} and C_{BC} and avalanche current I_{CE}. All of them produce intermodulation products at transistor output, with different magnitude and phase in a complex manner. In many cases, however, the transport current nonlinearity dominates. For intuitive understanding, it is highly desirable to derive analytical expressions of $IIP2$ and $OIP2$. Here we use the simplified small signal equivalent circuit in Figure 6.1, where $R_L = R_S = 50\,\Omega$. Further, we assume ideal exponential I_C-V_{BE}. Two-tone input $v_s = A(cos_1 t + cos_2 t)$. $i_c \approx g_m v_{be}(1 + \frac{v_{be}}{2\phi_t})$.

Following the Volterra series based circuit analysis approach of [27] [14], we can derive A_{IP2}, the input amplitude at the $IP2$ point as:

$$A_{IP2} = 2\phi_t \frac{v_s}{v_{be}} \tag{6.1}$$

where $\phi_t = \frac{kT}{q}$ is the thermal voltage. Here the two tones are very close, and the ratio $\frac{v_s}{v_{be}}$ is the evaluated at the fundamental frequency $\omega_1 \approx \omega_2$. $IIP2$, the RF input power at the

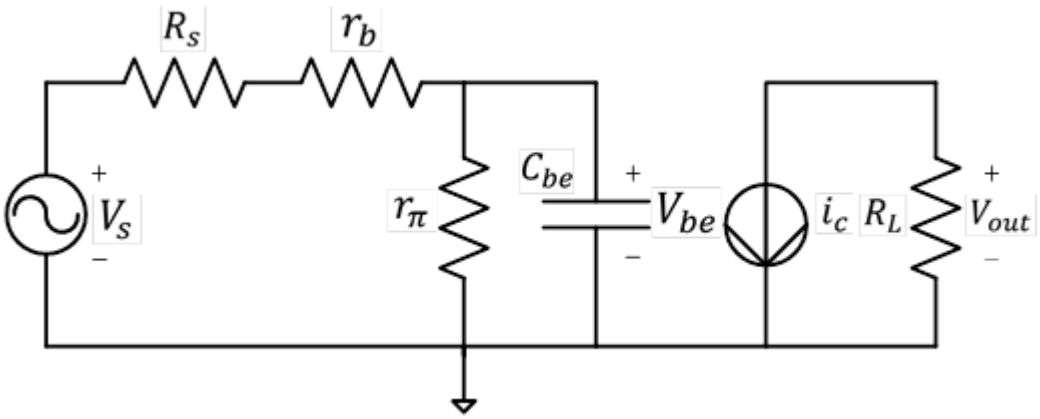

Figure 6.1: Simplified small signal equivalent circuit for BJT.

second order intermodulation intercept, is given by:

$$IIP2 = \frac{A_{IP2}^2}{8R_S} \tag{6.2}$$

Equation (6.1) shows that increasing biasing current I_C should increase $IIP2$ as it decreases v_{be} for a given v_s due to the increase of C_{be} and the decrease of r_π. This is consistent with experimental observation before high injection.

As $v_{out} = -g_m v_{be} R_L$, output power $P_{out} = \frac{v_{out}^2}{2R_L}$, input power $P_{in} = \frac{v_s^2}{8R_S}$, we can relate power gain $G = \frac{P_{out}}{P_{in}}$ to $\frac{v_s}{v_{be}}$ by:

$$(\frac{v_s}{v_{be}})^2 = \frac{4g_m^2 R_L^2}{G} \tag{6.3}$$

Substituting Equation 6.3 into Equation (6.1) and Equation (6.2), we obtain:

$$IIP2 = \frac{2I_C^2 R_L^2}{R_S}\frac{1}{G} = 2\frac{I_C^2 R_S}{G} \tag{6.4}$$

$OIP2$ is obtained from $IIP2 \times G$:

$$OIP2 = 2I_C^2 R_S \tag{6.5}$$

Thus $OIP2$ depends on biasing current only, and is independent of device size, V_{CE} or breakdown voltage, which is consistent with experimental observations that we will see below.

6.2 SBV device

Figure 6.2(a), (b) and (c) show the measured $IIP2$, f_T and $OIP2$ as a function of I_C for V_{CE} = 1.0, 1.3, 1.6, and 1.9 V on a SBV HBT. For comparison, $IIP3$ is shown in Figure 6.2(d). At lower I_C, both $IIP2$ and f_T increase with I_C. However, $IIP2$ starts to roll off while f_T is on the plateau of the $f_T - I_C$ curve. Therefore, for high $IIP2$, a higher I_C should be used, but I_C should be below peak f_T I_C. A higher V_{CE} effectively delays $IIP2$, f_T and $IIP3$ roll-off overall. The impact is more obvious for $IIP2$ and $IIP3$ than for f_T. Observe that the $IIP3$ peak is reached at a much smaller I_C, well before peak f_T is reached. $IIP2$ peaks either reside inside the flat f_T region or closely lag behind f_T roll-off. At high I_C, higher V_{CE} shifts $IIP2$ to higher level, implying that a higher V_{CE} can be used to effectively improve $IIP2$ at higher currents. For different V_{CE}, $OIP2$ have a very similar current dependence to $IIP2$. When collector current is lower than peak f_T I_C, $OIP2$ keeps the same for different V_{CE}.

As Equation (6.5) accounts for only the I_C or g_m nonlinearity, neglects C_{cb}, and assumes ideal $I_C - V_{BE}$ relation, it should only be used as a reference for understanding the impact of the transport current or transconductance nonlinearity. The measured $OIP2$ is generally higher than predicted by Equation (6.5), as can be seen from Figure 6.2(c). This difference is likely due to cancellation between different nonlinearities as well as the use of ideal $I_C - V_{BE}$ relation. Actual $I_C - V_{BE}$ is less nonlinear.

6.3 MBV device

Figure 6.3(a), (b) and (c) show $IIP2$, f_T and $IIP3$ versus I_C for MBV device, respectively. V_{CE} = 1.3, 1.6, 1.9 and 2.5 V. Overall, $IIP2$ has a similar current dependence to

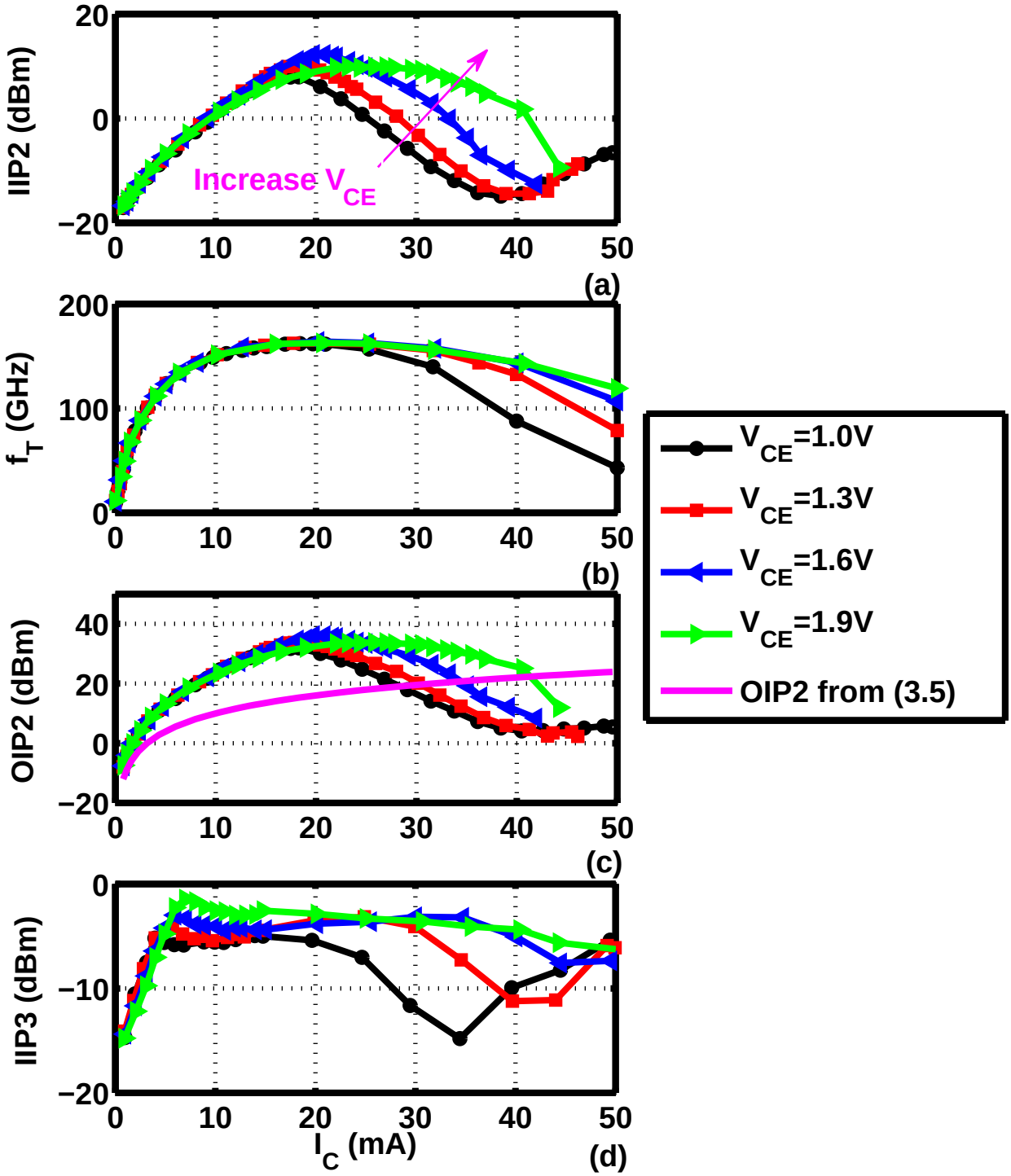

Figure 6.2: (a) $IIP2 - I_C$ (b) $f_T - I_C$ (c) $OIP2 - I_C$ (d) $IIP3 - I_C$. Standard breakdown voltage HBT.

$IIP3$. $IIP2$ increases with I_C, reaches a peak, and then drops down to a valley. Similar to SBV case, at low I_C, there is no big difference of $IIP2$ for different V_{CE} while $IIP3$ peaks at much smaller I_C than $IIP2$. $IIP2$ peak occurs at a higher I_C for larger V_{CE}.

6.4 HBV device

For HBV device, $IIP2$ versus I_C, f_T versus I_C, and $IIP2$ versus I_C are plotted in Figure 6.4(a), (b) and (c), respectively. V_{CE} = 1.3, 1.9, 2.5 and 3.1 V. Like SBV and MBV devices, $IIP2$ increases with I_C and peaks around where f_T rolls off. Increasing V_{CE} is much more effective in extending low injection to higher I_C in HBV device than in SBV and HBV devices.

6.5 Comparison of breakdown voltages

In Figure 6.5(a) and (b), the I_C dependence of $IIP2$ and f_T are compared for three breakdown versions at the same V_{CE}. Emitter area is the same. V_{CE} = 1.9 V is chosen as this is close to the open base breakdown voltage (BV_{CEO}) of the SBV device. $IIP2$ first increases with I_C, peaks, and then decreases sharply before rising again. The $IIP2$ peak occurs approximately at the same current as the f_T peak for all breakdown versions. Before 1 mA, $IIP2$ is the same for all, indicating that the collector-base junction difference does not affect $IIP2$ so long as high injection does not occur.

Figure 6.6(a), (b) and (c) compare $IIP2$, $OIP2$ and f_T of the three breakdown voltages as a function of I_C in a more fair manner. A higher V_{CE} is used for higher breakdown voltage. $OIP2$ calculated from Equation (6.5) is included in Figure 6.6(b). Before I_C = 4 mA, $OIP2$ is the same for three breakdown versions.

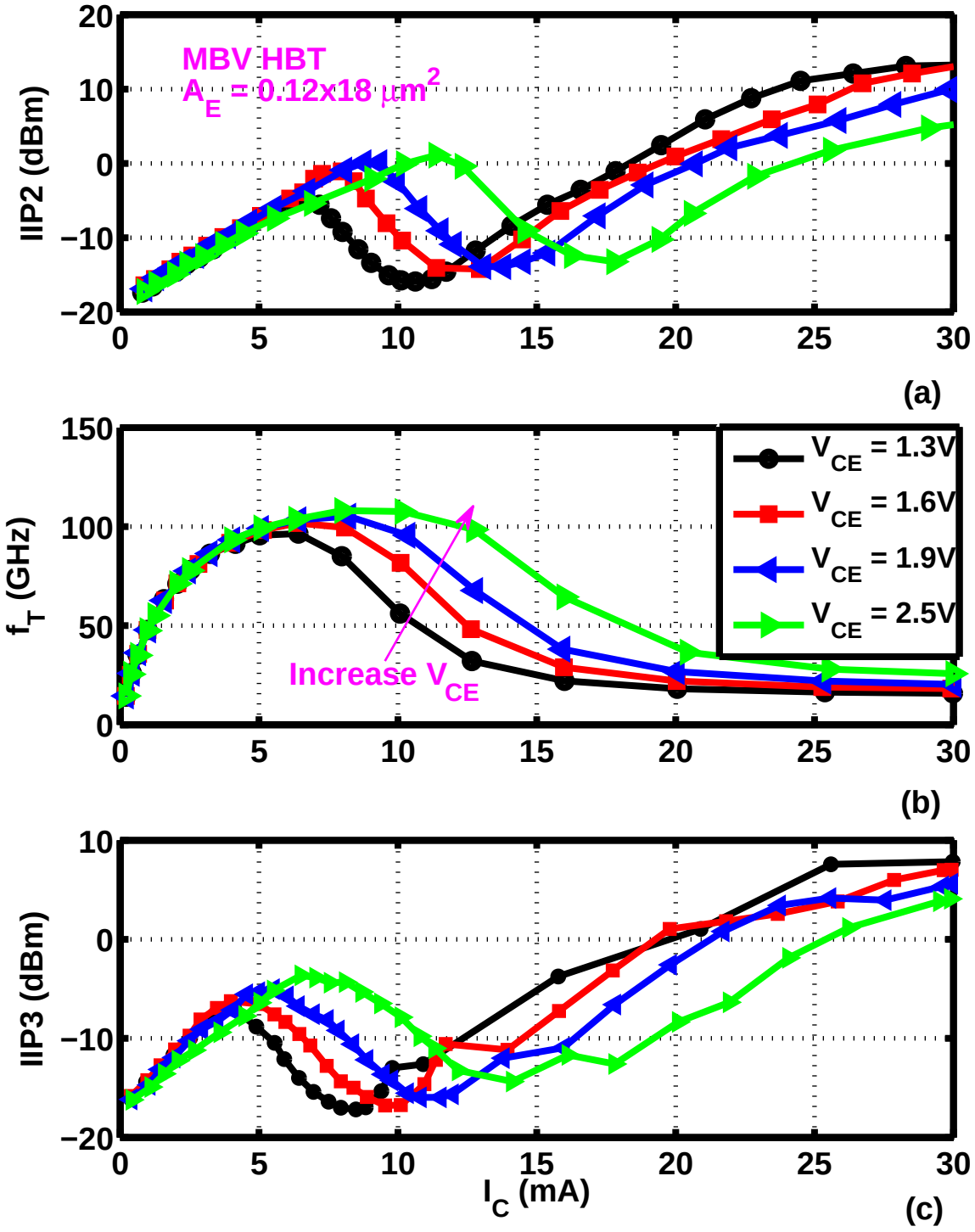

Figure 6.3: (a) $IIP2$-I_C (b) f_T-I_C (c) $IIP3$-I_C for V_{CE} = 1.3, 1.6, 1.9, 2.5 V. MBV HBT.

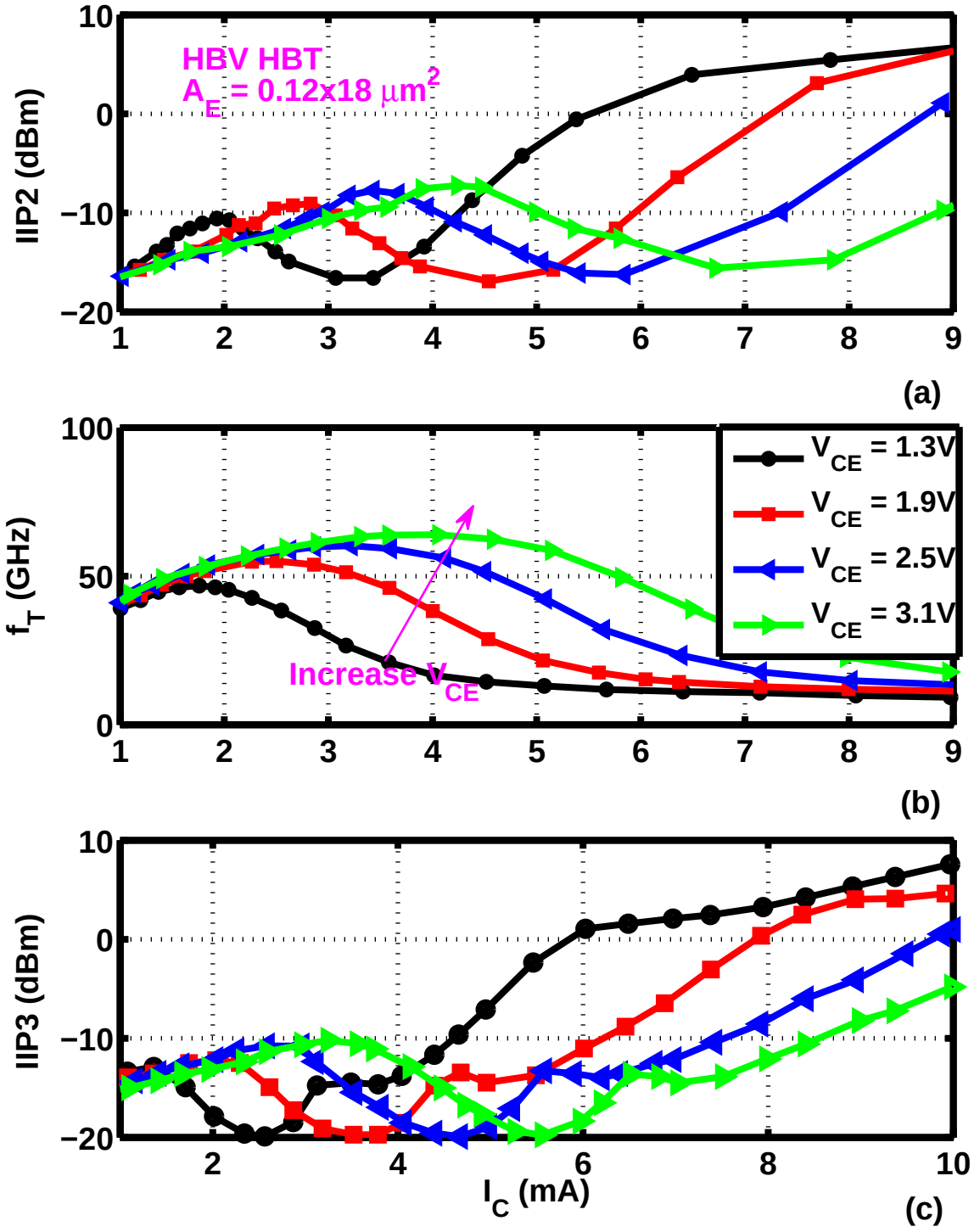

Figure 6.4: (a) $IIP2$-I_C (b) f_T-I_C (c) $IIP3$-I_C for V_{CE} = 1.3, 1.9, 2.5, 3.1 V. HBV HBT.

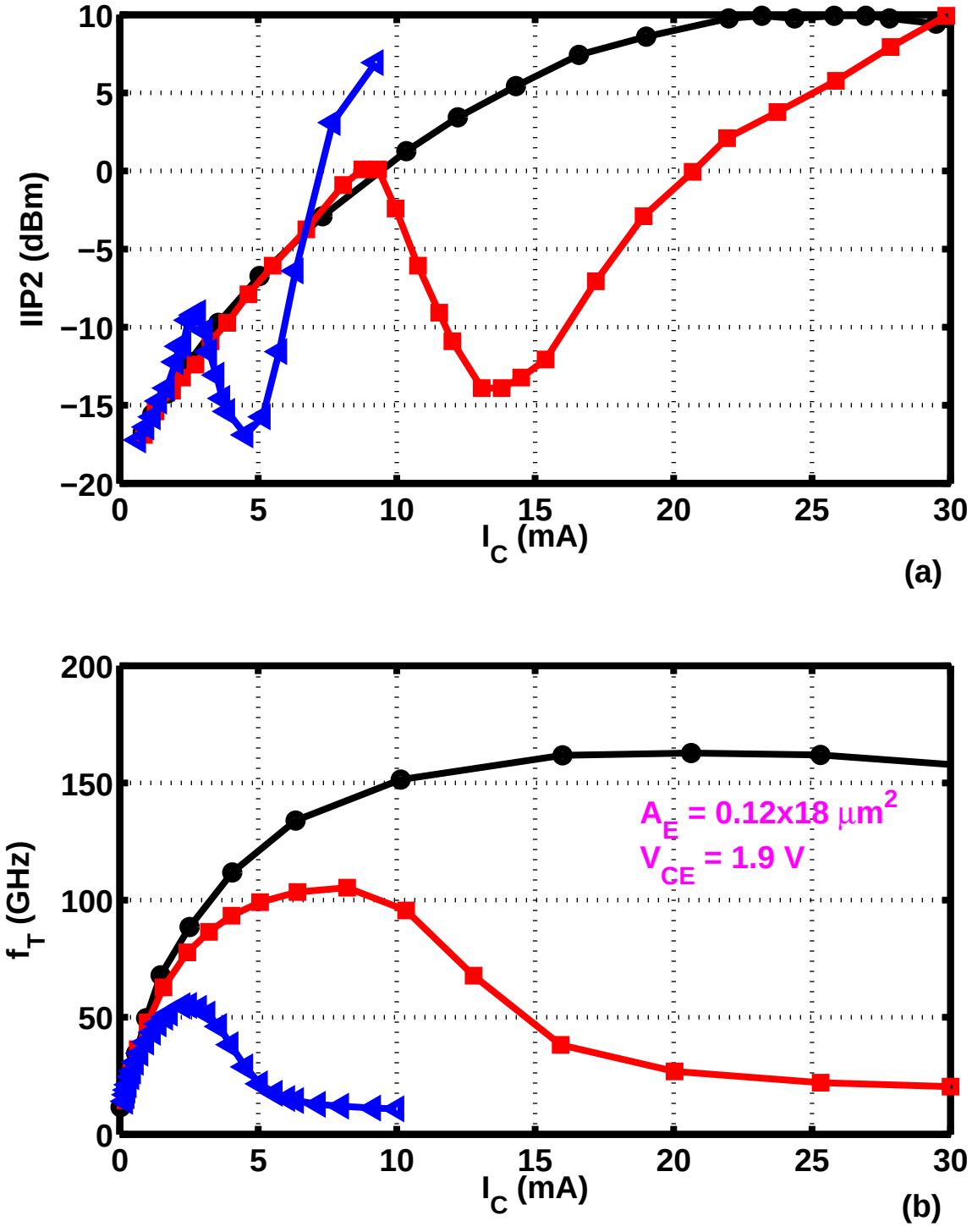

Figure 6.5: (a) $IIP2$-I_C (b) f_T-I_C for different breakdown versions. $V_{CE} = 1.9$ V.

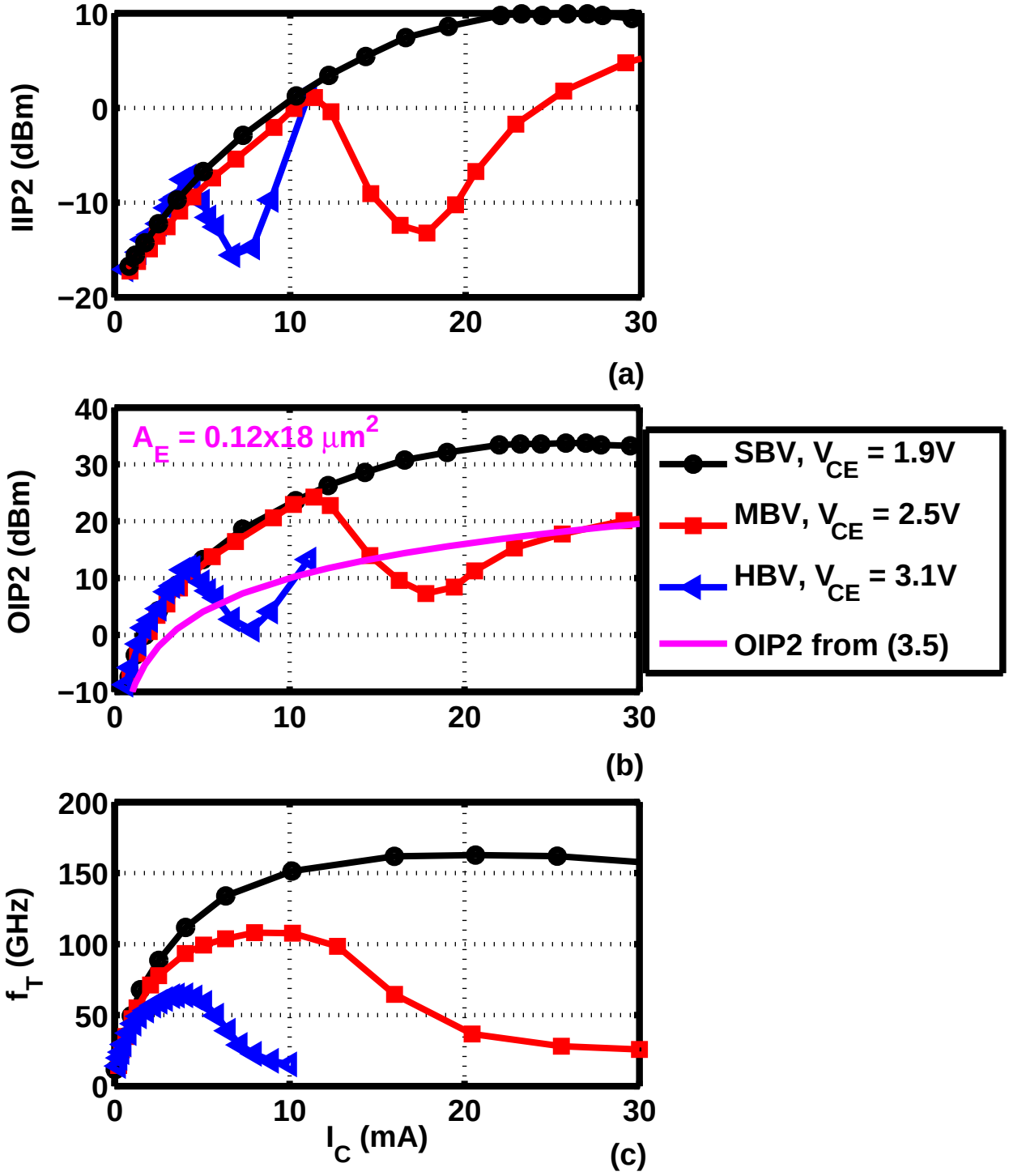

Figure 6.6: (a) $IIP2$-I_C (b) $OIP2$-I_C and (c) f_T-I_C for different breakdown versions.

6.6 Impact of device size

For SBV HBTs, size impact of $IIP2$ is investigated for three emitter sizes, 0.12x6, 0.12x12 and 0.12x18μm^2. Figure 6.7(a), (b) and (c) illustrate I_C dependence of $IIP2$, gain, and $OIP2$ with $V_{CB} = 0.5$ V. At low I_C when injection level is low, $OIP2$ does not vary with emitter area, as predicted by the Volterra series analysis results. A larger device means a larger I_C range for low injection operation, which naturally results in higher peak $OIP2$.

6.7 Impact of frequency

Figure 6.8(a) and (b) show measured $OIP2$ and gain as a function of I_C at different frequency for a SBV HBT at $V_{CE} = 1.5$ V. The tone spacing is 1 MHz. Before reaching a peak, $OIP2$ is similar to three frequencies. This verifies the validity of (6.5) in Section 6.1. Beyond $IIP2$ peak, a lower frequency gives a higher $OIP2$ at high I_C. For the whole I_C range, a higher frequency gives a lower gain.

6.8 Model comparison

This work is not aimed about modeling second order intermodulation linearity for the examined 200 GHz HBT technology. A quick comparison of simulated $IIP2$ versus I_C using different device models for a SBV HBT is shown in Figure 6.9. $V_{CE} = 1.3$ V. All three models give similar result before and near peak $IIP2$. Beyond the peak, HIgh CUrrent Model ($HICUM$) gives the best fitting to measurement result. This is due to $HICUM$ strength at high current description compared to the rest two candidates Vertical Bipolar Inter-Company ($VBIC$) model and Most EXquisite TRAnsistor Model ($Mextram$) [28]. In particular, $HICUM$ is the only model that captures the rising trend of $IIP2$ after reaching a valley at high I_C. Accuracy, however, is limited.

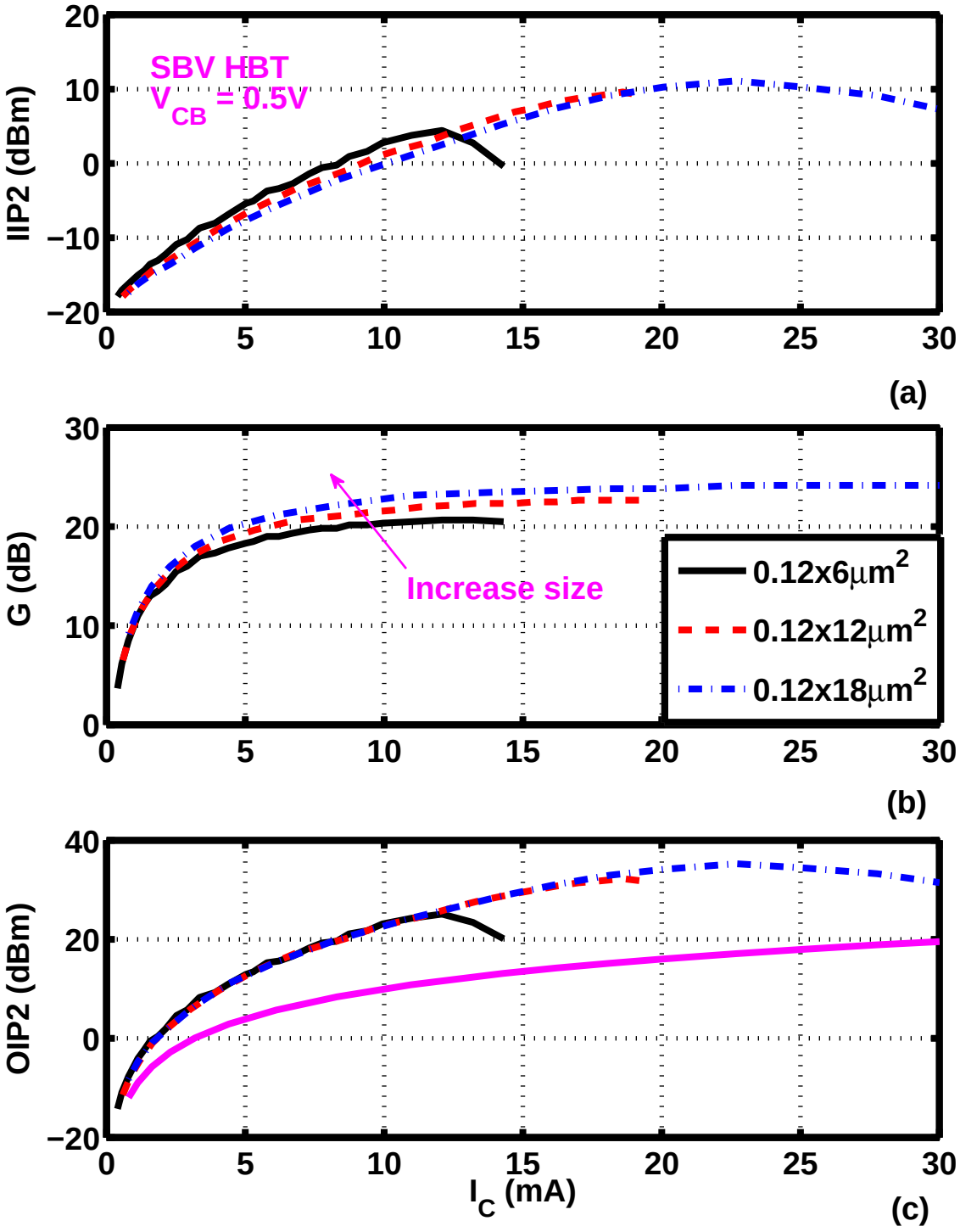

Figure 6.7: (a) $IIP2$-I_C (b) G-I_C and (c) $OIP2$-I_C. SBV HBT at $V_{CB} = 0.5$ V.

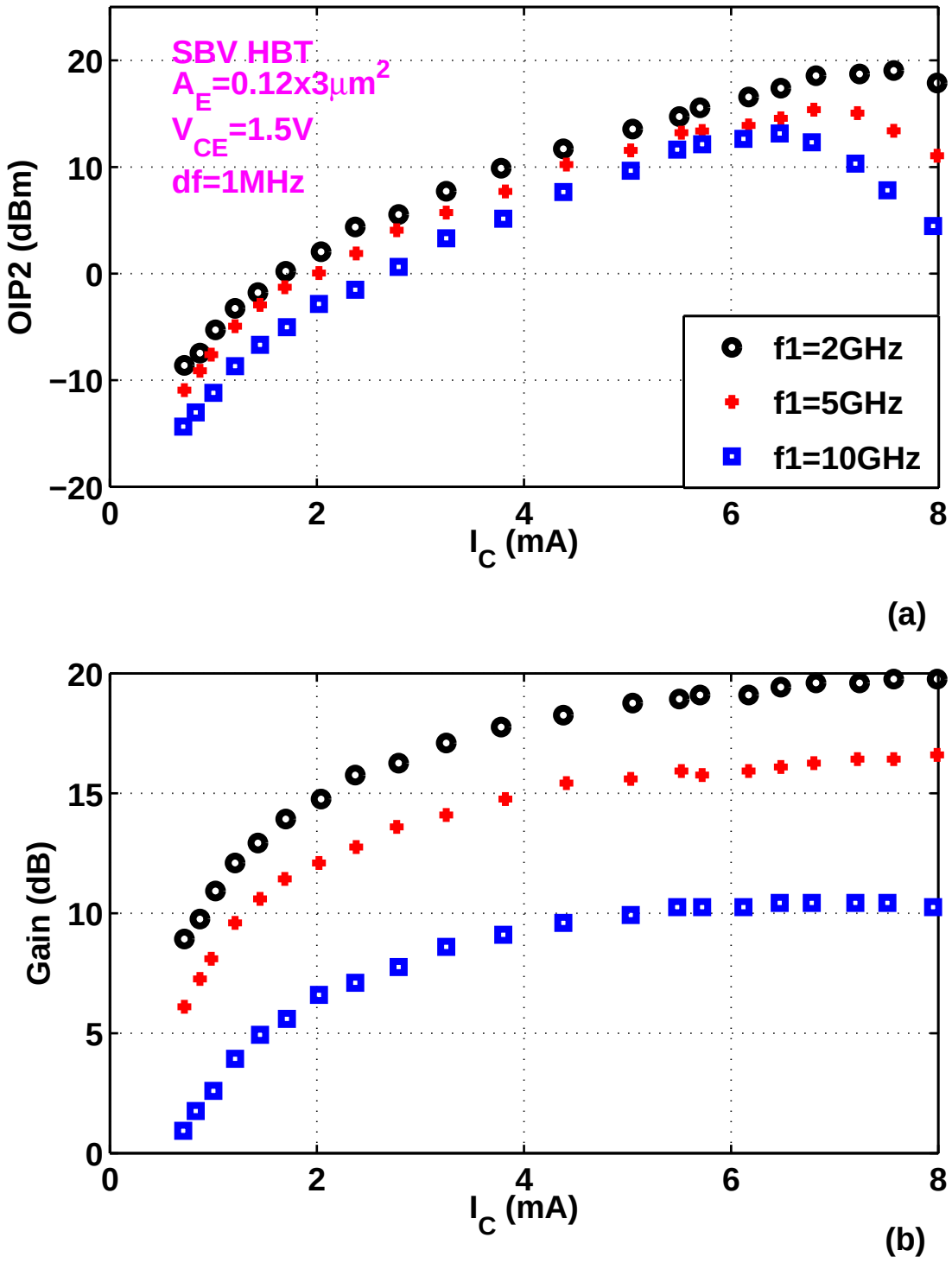

Figure 6.8: Measured (a) $OIP2$ and (b) Gain as a function of I_C at different frequency for a SBV HBT at $V_{CE} = 1.5$ V.

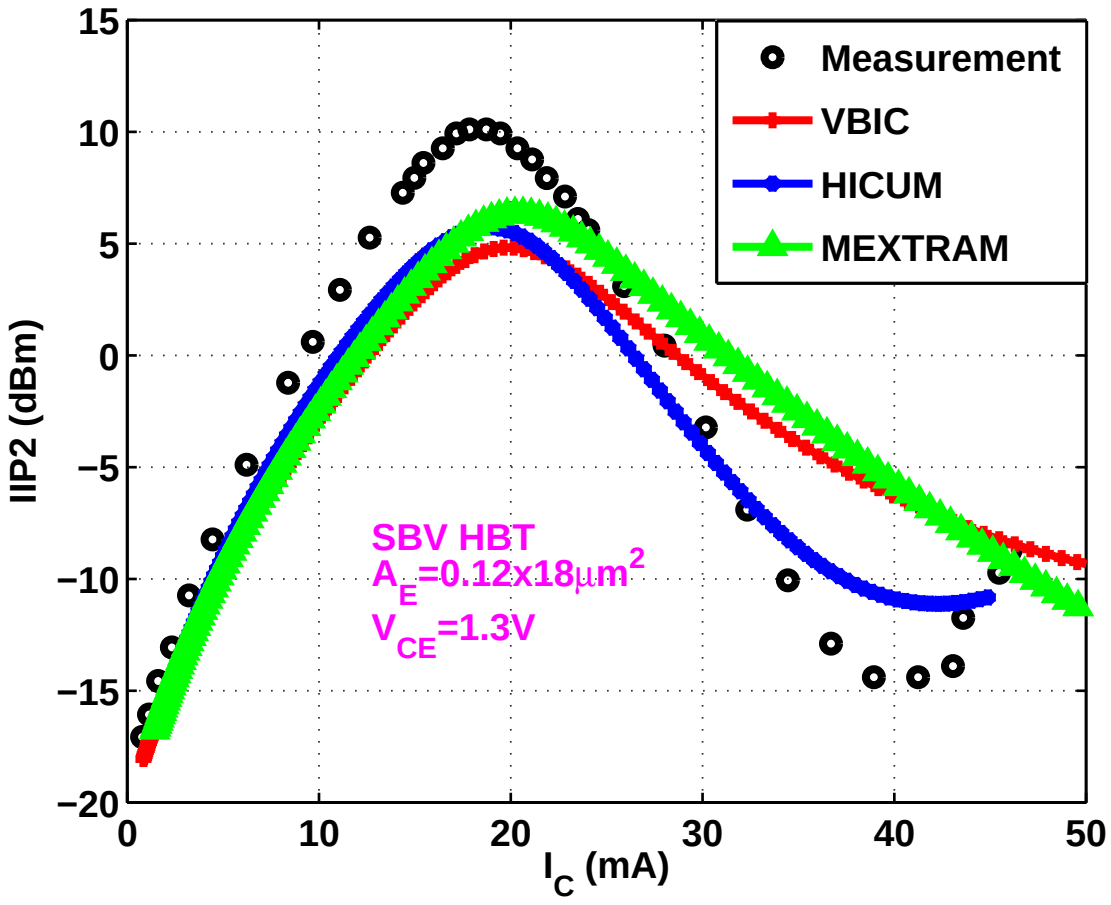

Figure 6.9: Comparison of measured $IIP2$ with simulation using three compact models at 5 GHz for a SBV HBT. $V_{CE} = 1.3$ V.

Chapter 7

Conclusion

For RF MOSFETs, the amplitude of the first order v_{gs} for $IP3$, $V_{GS,IP3} = \sqrt{\frac{1}{|\frac{3}{4}\frac{K_3 g_m}{g_m} + \Delta|}}$, is proposed for $IP3$ linearity as it is not affected by frequency and is comparable to traditional harmonic gate voltage $IP3$ ($VIP3$). Peak $VIP3$ V_{GS} is 55 mV higher than peak $IIP3$ V_{GS} for lack of output conductance related nonlinearities. Therefore $VIP3$ should not be used as a fast indicator for $IP3$ linearity. A high V_{DS} shifts $IIP3$ curve toward low V_{GS} value, resulting a high $IIP3$ at low V_{GS}. Peak $IIP2$ V_{GS} is near the second peak $IIP3$ V_{GS}. Fixed total width gives the same $IIP3$. For both thick oxide and thin oxide devices, a smaller channel length gives a higher $IIP3$ due to threshold voltage change. At high current, channel length does not affect $IIP3$ much. High frequency gives a high $IIP3$ while $OIP3$ is same same for different frequency.

For PSP compact model, parameters CT, $DPHIB$, CS, MUE and $THESAT$ are tuned to fit drain current. For sub-threshold region, CT determines the slope on log scale. $DPHIB$ affects threshold voltage and body effect. CS describes Coulomb scattering in moderate inversion. MUE is responsible for phonon scattering in strong inversion. $THESAT$ is important to fit drain current at very high V_{GS}. Gate resistance contributed by extrinsic components helps RF tuning.

For $real(Y22)$ modeling, body resistance (R_b) has a larger impact than gate resistance. Increasing R_b alone cannot solve the frequency dependence problem for $real(Y22)$. The capacitance between drain and body (C_{db}) needs to be tuned together with R_b, which also improves bias dependence of $real(Y22)$. $Imag(Y22)$ is not affected. V_{GS} dependence of g_{mb} results in extra $real(Y22)$ increase with V_{GS}. g_{mb} in design kit needs to be strengthened in future work.

For SiGe HBTs, $IP2$ behavior is much simpler than $IP3$. $IIP2$ and $OPI2$ both increase with I_C before high injection. $OIP2$ is independent of frequency, tone spacing , transistor size and breakdown voltage. SBV HBT has a smaller V_{CE} impact on $IIP2$ due to heavier doping on collector side compared to other breakdown versions. A higher V_{CE} leads to higher peak $IP2$ as it delays high injection. High current, standard breakdown version and large size are helpful to achieve high $IP2$.

Bibliography

[1] C.-H. Choi, Z. Yu, and R. Dutton, "Impact of poly-gate depletion on MOS RF linearity," *IEEE Electron Device Letters*, vol. 24, no. 5, pp. 330–332, May 2003.

[2] G. Niu, Q. Liang, J. D. Cressler, C. S. Webster, and D. L. Harame, "Systematic analysis of RF distortion in SiGe HBT's," *IEEE RFIC Dig.*, pp. 147–150, May 2001.

[3] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*. Springer, 1998.

[4] B. Razavi, *RF Microelectronics, 2nd ed.* Prentice Hall, 2011.

[5] J. Tang, G. Niu, Z. Jin, J. Cressler, S. Zhang, A. Joseph, and D. Harame, "Modeling and characterization of SiGe HBT low-frequency noise figures-of-merit for RFIC applications," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 11, pp. 2467–2473, Nov. 2002.

[6] A. J. Joseph, J. D. Cressler, D. M. Richey, and G. Niu, "Optimization of SiGe HBT's for operation at high current densities," *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1347 –1354, Jul. 1999.

[7] F. Sischka, "Checklist for successful DC measurements," Jul. 2015.

[8] ——, "Checklist for successful S-parameter measurements," Dec. 2014.

[9] G. Niu, Y. Li, Z. Feng, J. Pan, and D. Sheridan, "Characterization and modeling of intermodulation linearity in a 200 GHz SiGe HBT technology," *in Proc. IEEE BCTM*, pp. 1–4, Oct. 2006.

[10] J. Pan, "Systematic characterization and modeling of small and large signal performance of 50-200 GHz SiGe HBTs," Ph.D. dissertation, Auburn University, Aug. 2005.

[11] C. Rauscher, *Fundamentals of Spectrum Analysis*. Rohde & Schwarz, 2001.

[12] Cadence Design Systems, Inc., "Virtuoso Spectre circuit simulator and accelerated parallel simulator RF analysis user guide, product version 13.1," 2013.

[13] D. Liebl, "Measuring with modern spectrum analyzers," Feb. 2013.

[14] J. D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*. Artech House, 2003.

[15] Z. Li, G. Niu, Q. Liang, and K. Imura, "Intermodulation linearity in high-k/metal gate 28nm RF CMOS transistors," *Electronics*, vol. 4, no. 3, pp. 614–622, Sep. 2015.

[16] M.-T. Yang, K. Liao, R. Welstand, C. Teng, W. Sy, Y. Chen, R. Dutta, P. Chidambaram, M. Han, Y. Du, and G. Yeap, "RF and mixed-signal performances of a low cost 28nm low-power CMOS technology for wireless system-on-chip applications," *IEEE VLSIT Dig.*, pp. 40–41, Jun. 2011.

[17] X. Wei, G. Niu, Y. Li, M.-T. Yang, and S. Taylor, "Modeling and characterization of intermodulation linearity on a 90nm RF CMOS technology," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 4, pp. 965–971, Apr. 2009.

[18] X. Wei, G. Niu, Y. Li, M. Yang, and S. Taylor, "Experimental characterization and simulation of RF intermodulation linearity in a 90nm RF CMOS technology," *IEEE RFIC Dig.*, pp. 251–254, Jun. 2008.

[19] X. Wei, "On-wafer S-parameter measurement using four-port technique and intermodulation linearity of RF CMOS," Ph.D. dissertation, Auburn University, Dec. 2008.

[20] P. H. Woerlee, M. J. Knitel, R. van Langevelde, D. B. M. Klaassen, L. F. Tiemeijer, A. J. Scholten, and A. T. A. Z. van Duijnhoven, "RF-CMOS performance trends," *IEEE Trans. Electron Devices*, vol. 48, pp. 1776–1782, Aug. 2001.

[21] G. Niu, J. Pan, X. Wei, S. Taylor, and D. Sheridan, "Intermodulation linearity characteristics of CMOS transistors in a 0.13μm process," *IEEE RFIC Dig.*, pp. 65–68, Jun. 2005.

[22] X. Li, W. Wu, G. Gildenblat, G. Smit, A. Scholten, , D. Klaassenand, and R. van Langevelde, "PSP 103.1, Technical Note NXP-R-TN-2008/00299," 2009.

[23] Y. Cheng, M. J. Deen, and C.-H. Chen, "MOSFET modeling for RF IC design," *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1286–1303, Jul. 2005.

[24] T. Ytterdal, Y. Cheng, and T. A. Fjeldly, *Device Modeling for Analog and RF CMOS Circuit Design*. John Wiley & Sons, 2003.

[25] W. Yang, M. V. Dunga, X. Xi, J. He, W. Liu, Kanyu, M. Cao, X. Jin, J. J. Ou, M. Chan, A. M. Niknejad, and C. Hu, "BSIM4.6.2 MOSFET model, user manual," 2008.

[26] I. Kwon, M. Je, K. Lee, and H. Shin, "A simple and analytical parameter-extraction method of a microwave MOSFET," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 6, pp. 1503–1509, Jun. 2002.

[27] G. Niu, Q. Liang, J. D. Cressler, C. S. Webster, and D. L. Harame, "RF linearity characteristics of SiGe HBTs," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 9, pp. 1558–1565, Sep. 2001.

[28] M. Schroter and A. Chakravorty, *Compact Hierarchical Bipolar Transistor Modeling With Hicum*. World Scientific Publishing Company, 2010.

Appendix A

Taylor series expansion of output voltage for a two-tone input

The output voltage of a nonlinear two-port (v_{out}) can be expressed as a function of the input voltage ($v_{in}(t)$) by using a power series like below:

$$v_{out} = \sum_{n=1}^{\infty} a_n v_{in}(t) = a_1 v_{in}(t) + a_2 v_{in}^2(t) + a_3 v_{in}^3(t) + \cdots \tag{A.1}$$

where a_n is the nonlinear coefficient.

For a two-tone input with the same amplitude (A), the input voltage can be expressed as

$$v_{in}(t) = Acos(\omega_1 t) + Acos(\omega_2 t) \tag{A.2}$$

In addition to original fundamental signals, a DC component, harmonics and intermodulation products exist at output. Table A.1 lists the frequencies for corresponding output spectrum components up to third order.

Plugging (A.2) into (A.1), up to third power, the output voltage can be expressed as

	Symbolic Frequency	Spectrum Components
First order harmonic	ω_1, ω_2	Fundamental
Second order harmonic	$2\omega_1$, $2\omega_2$	$HD2$
Second order intermodulation products	$\omega_2 - \omega_1$, $\omega_2 + \omega_1$	$IM2$
Third order harmonic	$3\omega_1$, $3\omega_2$	$HD3$
Third order intermodulation products	$2\omega_1 - \omega_2$, $2\omega_2 - \omega_1$	$IM2$

Table A.1: Output spectrum components up to third order

$$
\begin{aligned}
v_{out} \;=\; & a_2A^2 + && DC\ component \\
& +(a_1A + \frac{9}{4}a_3A^3)cos(\omega_1 t) && Fundamental \\
& +(a_1A + \frac{9}{4}a_3A^3)cos(\omega_2 t) && Fundamental \\
& +\frac{1}{2}a_2A^2cos(2\omega_1 t) && HD2 \\
& +\frac{1}{2}a_2A^2cos(2\omega_2 t) && HD2 \\
& +a_2A^2cos((\omega_2 - \omega_1)t) && IM2 \\
& +a_2A^2cos((\omega_2 + \omega_1)t) && IM2 \\
& +\frac{1}{4}a_3A^3cos(3\omega_1 t) && HD3 \\
& +\frac{1}{4}a_3A^3cos(3\omega_2 t) && HD3 \\
& +\frac{3}{4}a_3A^3cos((2\omega_1 - \omega_2)t) && IM3 \\
& +\frac{3}{4}a_3A^3cos((2\omega_1 + \omega_2)t) && IM3 \\
& +\frac{3}{4}a_3A^3cos((2\omega_2 - \omega_1)t) && IM3 \\
& +\frac{3}{4}a_3A^3cos((2\omega_2 + \omega_1)t) && IM3
\end{aligned}
\tag{A.3}
$$

According to (A.3), the amplitude difference in dB between $IM2$ and $HD2$ is:

$$
20log|\frac{a_2A^2}{\frac{1}{2}a_2A^2}| = 6\ dB. \tag{A.4}
$$

This 6 dB difference is a very helpful rule-of-thumb to monitor time-consuming second-order intermodulation linearity measurement as shown in Figure 2.14.

Appendix B

Volterra Series based IP3 derivation for MOSFET

For weak nonlinear system, the output can be described as the sum of the transfer functions below order three. The first order transfer function $H1(s)$ is essentially the transfer function of the linearized circuit. The second and third order transfer functions, $H2(s1, s2)$ and $H3(s1, s2, s3)$, can be solved in increasing order by repeatedly solving the linearized circuit using different excitation. $s1 = j\omega 1$, $s2 = j\omega 2$, and $s3 = j\omega 3$.

$IIP3$ is derived below using nonlinear current source method. A simplified equivalent circuit as shown in Figure B.1 is used. Gate-drain capacitance (C_{gd}) is omitted, as the result is much simpler and sufficient for most purposes. $R_S = 50\ \Omega$. C_{gs} is gate-to-source capacitance. C_d is drain capacitance. $R_L = 50\ \Omega$ is load resistance.

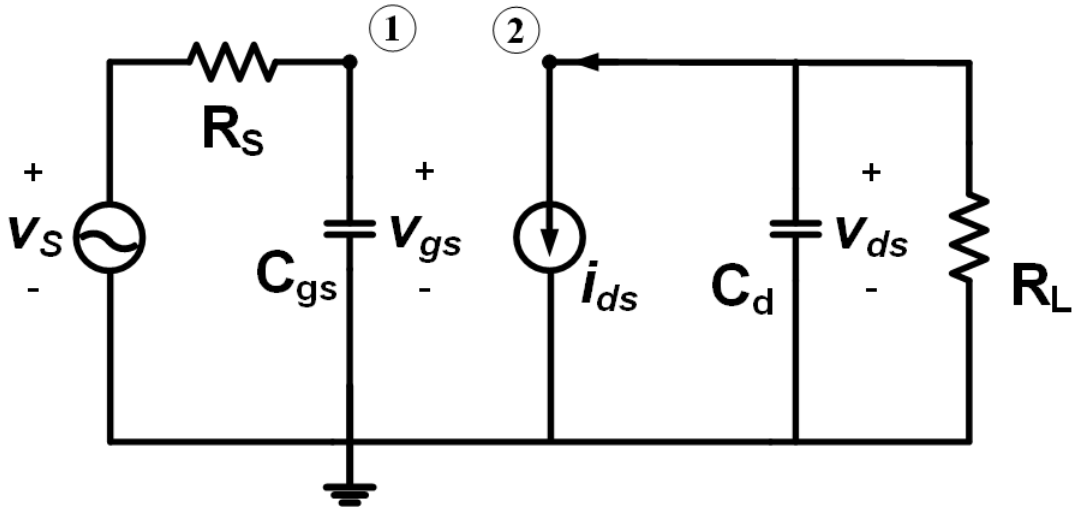

Figure B.1: Simplified equivalent circuit used for $IP3$ derivation using Volterra series.

Up to third order, nonlinear drain current i_{ds} can be expressed as:

$$\begin{aligned} i_{ds} &= g_m v_{gs} + K_{2_{g_m}} v_{gs}^2 + K_{3_{g_m}} v_{gs}^3 + g_0 v_{ds} + K_{2_{g_0}} v_{ds}^2 + K_{3_{g_0}} v_{ds}^3 + K_{2_{g_m g_0}} v_{gs} v_{ds} \\ &\quad + K_{3_{2_{g_m} g_0}} v_{gs}^2 v_{ds} + K_{3_{g_m 2 g_0}} v_{gs} v_{ds}^2 \end{aligned} \tag{B.1}$$

g_m and g_0 are transconductance and output conductance.

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} \tag{B.2}$$

$$g_0 = \frac{\partial I_{DS}}{\partial V_{DS}} \tag{B.3}$$

$K_{2_{g_m}}, K_{3_{g_m}}, K_{2_{g_0}}, K_{3_{g_0}}, K_{2_{g_m g_0}}, K_{3_{2_{g_m g_0}}}$ and $K_{3_{g_m 2 g_0}}$ are nonlinearity coefficients that relate to higher order partial derivatives as defined in [3] using Taylor expansion, which are repeated below:

$$K_{2_{g_m}} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2} \tag{B.4}$$

$$K_{3_{g_m}} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3} \tag{B.5}$$

$$K_{2_{g_0}} = \frac{1}{2} \frac{\partial^2 I_{DS}}{\partial V_{DS}^2} \tag{B.6}$$

$$K_{3_{g_0}} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{DS}^3} \tag{B.7}$$

$$K_{2_{g_m g_0}} = \frac{1}{2} \frac{\partial I_{DS}}{\partial V_{GS}} \frac{\partial I_{DS}}{\partial V_{DS}} \tag{B.8}$$

$$K_{3_{2_{g_m} g_0}} = \frac{1}{6} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2} \frac{\partial I_{DS}}{\partial V_{DS}} \tag{B.9}$$

$$K_{3_{g_m 2 g_0}} = \frac{1}{6} \frac{\partial I_{DS}}{\partial V_{GS}} \frac{\partial^2 I_{DS}}{\partial V_{DS}^2} \tag{B.10}$$

Applying Kircoffs current law at node 1 and 2 in Figure B.1 gives:

$$\begin{bmatrix} Y_S(s) & 0 \\ g_m & Y_L(s) \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{V_S}{R_S} \\ 0 \end{bmatrix} \tag{B.11}$$

where admittance $Y_S(s) = \frac{1}{R_S} + sC_{gs}$ and $Y_L(s) = \frac{1}{R_L} + g_0 + sCd$.

B.1 First order kernels

The first order kernels can be obtained from the response of the linearized circuit to excitation of V_S according to the equivalent circuit in Figure B.2.

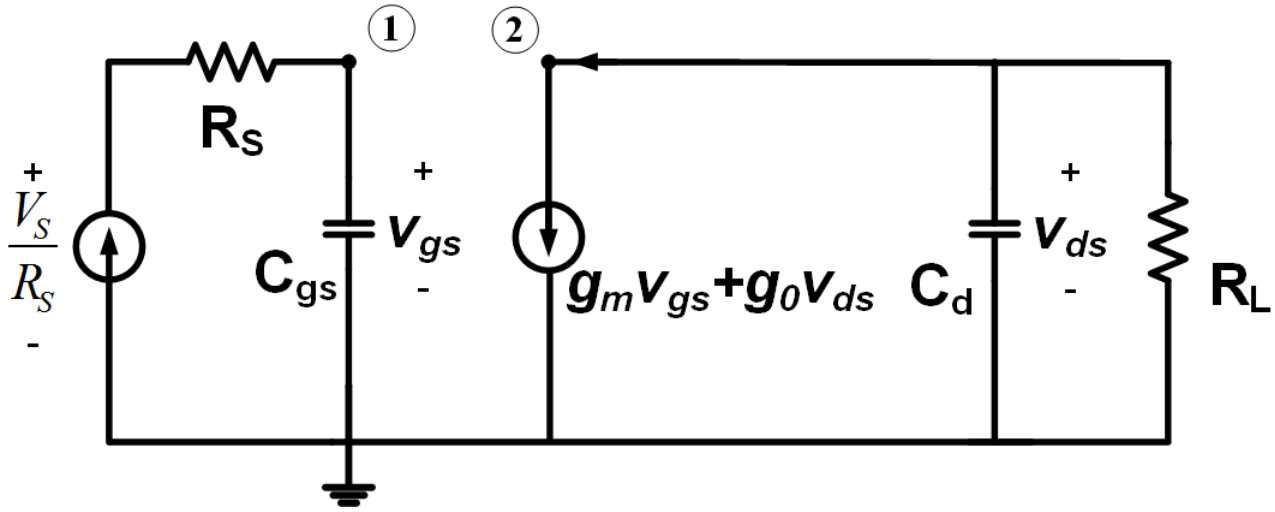

Figure B.2: Equivalent circuit to solve the first order kernels.

The transfer functions at nodes 1 and 2 are denoted as $H1_1(s)$ and $H1_2(s)$ with the subscript for corresponding node. So matrix equation for $H1_1(s)$ and $H1_2(s)$ is

$$\begin{bmatrix} Y_S(s) & 0 \\ g_m & Y_L(s) \end{bmatrix} \begin{bmatrix} H1_1(s) \\ H1_2(s) \end{bmatrix} = \begin{bmatrix} \frac{1}{R_S} \\ 0 \end{bmatrix} \tag{B.12}$$

So $H1_1(s)$ and $H1_2(s)$ can be obtained by sloving (B.12) as below:

$$H1_1(s) = \frac{1}{Y_S(s)} \frac{1}{R_S} \tag{B.13}$$

and

$$H1_2(s) = \frac{-g_m}{Y_S(s)Y_L(s)} \frac{1}{R_S}. \tag{B.14}$$

B.2 Second order kernels

The second order transfer functions at nodes 1 and 2 are denoted as $H2_1(s1, s2)$ and $H2_2(s1, s2)$. Figure B.3 shows the equivalent circuit to calculate the second order kernels. A virtual excitation i_{NL2} is placed in parallel with the linearized response $g_m v_{gs} + g_0 v_{ds}$ with following expression:

$$\begin{aligned} i_{NL2} &= K_{2_{g_m}} H1_1(s1)H1_1(s2) + \\ &\quad K_{2_{g_0}} H1_2(s1)H1_2(s2) + \\ &\quad \frac{1}{2} K_{2_{g_m g_0}} [H1_1(s1)H1_2(s2) + H1_1(s2)H1_2(s1)]. \end{aligned} \tag{B.15}$$

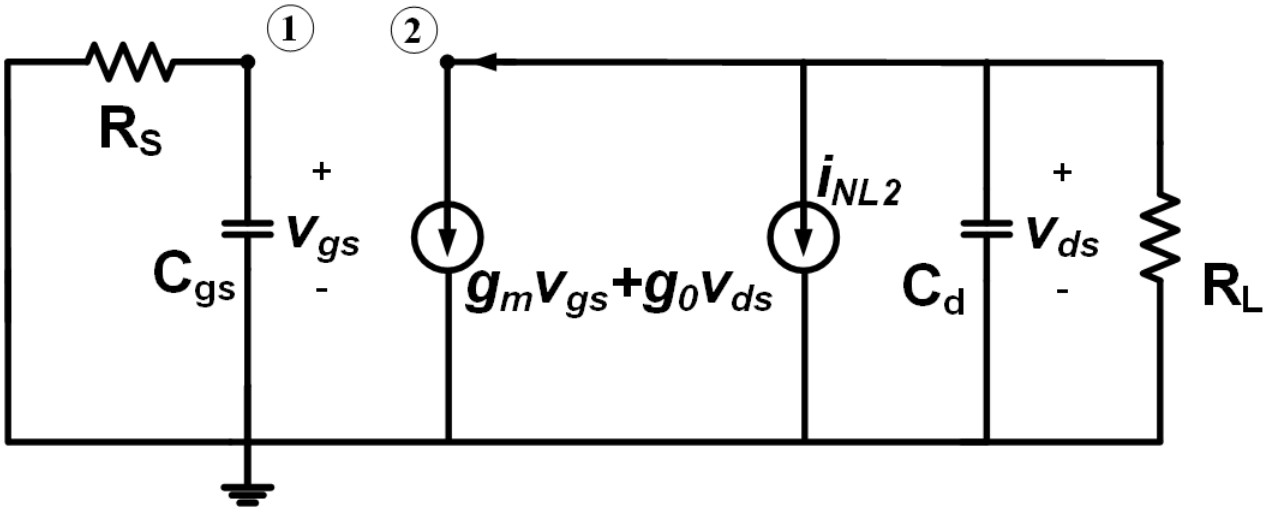

Figure B.3: Equivalent circuit to solve the second order kernels.

Matrix equation for $H2_1(s1, s2)$ and $H2_2(s1, s2)$ is listed as

$$\begin{bmatrix} Y_S(s1+s2) & 0 \\ g_m & Y_L(s1+s2) \end{bmatrix} \begin{bmatrix} H2_1(s1,s2) \\ H2_2(s1,s2) \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL2} \end{bmatrix}. \tag{B.16}$$

So $H2_1(s1, s2)$ and $H2_2(s1, s2)$ can be obtained by sloving (B.16) as below:

$$H2_1(s1, s2) = 0 \tag{B.17}$$

and

$$H2_2(s1, s2) = \frac{-i_{NL2}}{Y_L(s1 + s2)}. \tag{B.18}$$

B.3 Third order kernels

The third order transfer functions at nodes 1 and 2 are denoted as $H3_1(s1, s2, s3)$ and $H3_2(s1, s2, s3)$. Figure B.4 shows the equivalent circuit to calculate the third order kernels. A virtual excitation i_{NL3} is placed in parallel with the linearized response $g_m v_{gs} + g_0 v_{ds}$ with following expression:

$$\begin{aligned}
i_{NL3} \;=\; & K_{3_{g_m}} H1_1(s1)H1_1(s2)H1_1(s3) + \\
& K_{3_{g_0}} H1_2(s1)H1_2(s2)H1_2(s3) + \\
& \frac{2}{3} K_{2_{g_m}} [H1_1(s1)H2_1(s2, s3) + H1_1(s2)H2_1(s1, s3) + H1_1(s3)H2_1(s1, s2)] + \\
& \frac{2}{3} K_{2_{g_0}} [H1_2(s1)H2_2(s2, s3) + H1_2(s2)H2_2(s1, s3) + H1_2(s3)H2_2(s1, s2)] + \\
& \frac{1}{3} K_{2_{g_m g_0}} [H1_1(s1)H2_2(s2, s3) + H1_1(s2)H2_2(s1, s3) + H1_1(s3)H2_2(s1, s2) + \\
& H1_2(s1)H2_1(s2, s3) + H1_2(s2)H2_1(s1, s3) + H1_2(s3)H2_1(s1, s2)] + \\
& \frac{1}{3} K_{3_{2g_m g_0}} [H1_1(s1)H1_1(s2)H1_2(s3) + H1_1(s2)H1_1(s3)H1_2(s1) \\
& + H1_1(s1)H1_1(s3)H1_2(s2)] + \\
& \frac{1}{3} K_{3_{g_m 2g_0}} [H1_1(s1)H1_2(s2)H1_2(s3) + H1_1(s2)H1_2(s1)H1_2(s3) \\
& + H1_1(s3)H1_2(s1)H1_2(s2)]
\end{aligned} \tag{B.19}$$

Matrix equation for $H3_1(s1, s2, s3)$ and $H3_2(s1, s2, s3)$ is listed as

$$\begin{bmatrix} Y_S(s1 + s2 + s3) & 0 \\ g_m & Y_L(s1 + s2 + s3) \end{bmatrix} \begin{bmatrix} H3_1(s1, s2, s3) \\ H3_2(s1, s2, s3) \end{bmatrix} = \begin{bmatrix} 0 \\ -i_{NL3} \end{bmatrix} \tag{B.20}$$

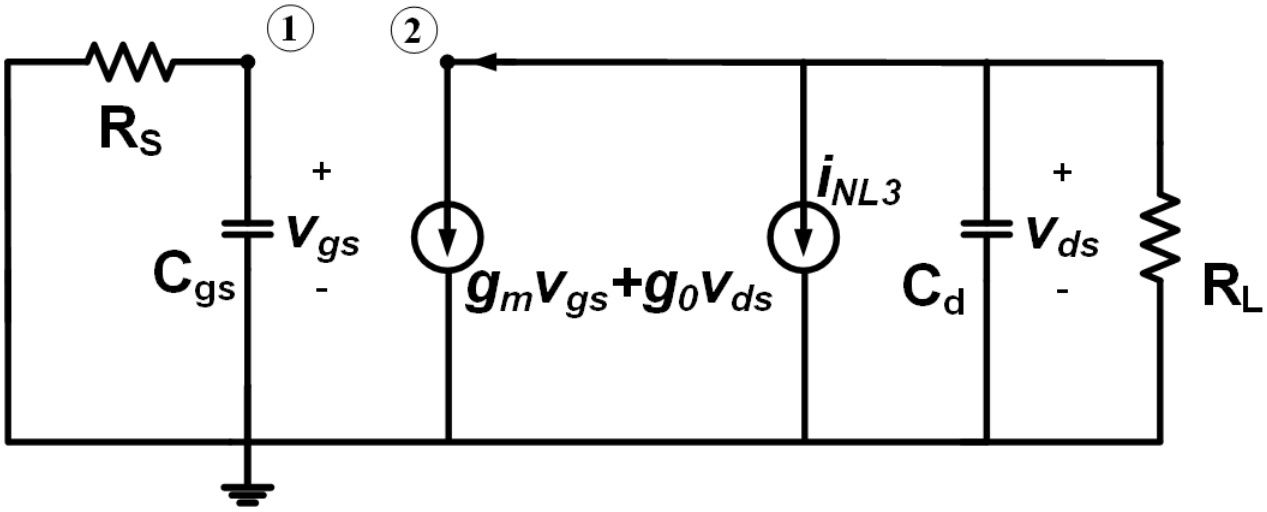

Figure B.4: Equivalent circuit to solve the third order kernels.

So $H3_1(s1, s2, s3)$ and $H3_2(s1, s2, s3)$ can be obtained by sloving (B.20) as below:

$$H3_1(s1, s2, s3) = 0 \tag{B.21}$$

and

$$H3_2(s1, s2, s3) = \frac{-i_{NL3}}{Y_L(s1 + s2 + s3)}. \tag{B.22}$$

B.4 Expression of $IIP3$

For a two-tone input V_S with an amplitude of A, $IIP3$ is defined as the input power at which the third order intermodulation product ($\frac{3}{4}A^3|H3_2(-s1, s2, s2)|$) equals to fundamental tone output ($A|H1_2(s2)|$). According to (B.14) and (B.22), The first and third order Volterra kernels

$$H1_2(s2) = \frac{-g_m}{Y_S(s2)Y_L(s2)}\frac{1}{R_S} \tag{B.23}$$

and

$$H3_2(-s1, s2, s2) = \frac{-i_{NL3}}{Y_L(2s2 - s1)}. \tag{B.24}$$

Since $2s2 - s1 \approx s2$,

$$H3_2(-s1, s2, s2) = \frac{-i_{NL3}}{Y_L(s2)}. \tag{B.25}$$

Therefore,

$$\frac{\frac{3}{4}A^3|H3_2(-s1, s2, s2)|}{A|H1_2(s2)|} = \frac{3}{4}A^2 \left| \frac{-i_{NL3}}{\frac{-g_m}{Y_S(s2)}\frac{1}{R_S}} \right| \tag{B.26}$$

$IIP3$ can be expressed as:

$$IIP3 = \frac{A^2}{8R_S} = \frac{1 + \omega^2 C_{gs}^2 R_S^2}{6R_S} \frac{1}{|\frac{K_{3g_m}}{g_m} + \Delta|}. \tag{B.27}$$

where $\Delta = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$. Δ_1 through Δ_4 are functions of nonlinear output conductance, its high order terms and cross terms with transconductance nonlinearity as follows:

$$\Delta_1 = -\frac{1}{3}K_{2_{g_m}g_0}\frac{K_{2_{g_m}}}{g_m}Z_1 - \frac{1}{3}K_{3_{2g_m}g_0}Z_2 \tag{B.28}$$

$$\Delta_2 = \frac{2}{3}K_{2_{g_m}}K_{2_{g_0}}Z_3 + \frac{1}{3}K_{3_{g_m 2g_0}}g_m Z_4 + \frac{1}{3}K_{2_{g_m g_0}}^2 Z_5 \tag{B.29}$$

$$\Delta_3 = -K_{3_{g_0}}g_m^2 Z_6 - \frac{1}{3}K_{2_{g_m g_0}}K_{2_{g_0 g_m}}Z_7 \tag{B.30}$$

$$\Delta_4 = \frac{2}{3}K_{2_{g_0}}^2 g_m^2 Z_8 \tag{B.31}$$

Z_1 through Z_8 are given by:

$$Z_1 = Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2) \tag{B.32}$$

$$Z_2 = Z_L(\omega_1) + [Y_S(-\omega_2)Y_S^{-1}(\omega_1) + 2] \tag{B.33}$$

$$Z_3 = 2Z_L(\omega_1 - \omega_2)Z_L(\omega_1) + Z_L(2\omega_1)Z_L(-\omega_2) \tag{B.34}$$

$$Z_4 = Z_L^2(\omega_1)[2Y_S(-\omega_2)Y_S^{-1}(\omega_1) + 1] \tag{B.35}$$

$$Z_5 = 2Z_L(\omega_1 - \omega_2)Z_L(-\omega_2) + Z_L(2\omega_1)Z_L(-\omega_1) \tag{B.36}$$

$$Z_6 = Z_L^2(\omega_1)Z_L(-\omega_2) \tag{B.37}$$

$$Z_7 = Z_L^2(\omega_1)[Z_L(2\omega_1) + 2Z_L(2\omega_1) + 6Z_L(\omega_1 - \omega_2)] \tag{B.38}$$

$$Z_8 = Z_L^2(\omega_1)Z_L(-\omega_2)[Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2)] \tag{B.39}$$

with $Z_L(\omega) = \frac{1}{\frac{1}{R_L} + g_0 + j\omega C_d}$ and $Y_S(\omega) = \frac{1}{R_S} + j\omega C_{gs}$.

Appendix C

Power calibration procedures for IP2/IP3 measurement

Due to varying power loss with frequency, the attenuation on input branch needs to be added to the amplitude of signal generators while the attenuation on output branch needs to be deducted from the raw data from spectrum analyzer using following 6 steps.

1. Power on DC power supply, signal generator and spectrum analyzer. Wait for 30 minutes.

2. Initialize and calibrate the power meter on its front panel as shown in Figure C.1.

   - Press LINE switch to ON;
   - press "ZERO" key to initialize power meter;
   - Press "CAL" key and confirm with "ENTER" for "REF CF 100%";
   - Connect the other end of power sensor to "POWER REF" port in order to make use internal 1dBm signal at 50 MHz;
   - Press "PWR REF" to see if it returns with a steady reading of "0 dBm"; If not, repeat above steps starting from "ZERO" key

3. Place the input branch between signal generator and power meter; Adjust "Frequency" to the desired one and set "Amplitude" as 0 dBm; Toggle "RF On/Off" key to trigger input signal and record the reading on power meter as $P_{i,cal}$.

4. Repeat Step 3 for output branch and record the reading as $P_{i,cal}$.

5. Connect both input and output branches, probe an on-wafer "through" structure, and repeat Step 3; Record the reading as $P_{t,cal}$.

6. Verify if $P_{i,cal} + P_{o,cal} < P_{t,cal}$; If true, for $IP2/IP3$ raw measurement data, increase all raw output power levels on spectrum analyzer by $(P_{o,cal} + \frac{P_{t,cal} - P_{i,cal} - P_{o,cal}}{2})$ dB and decrease all raw input power levels by $(P_{i,cal} + \frac{P_{t,cal} - P_{i,cal} - P_{o,cal}}{2})$ dB;

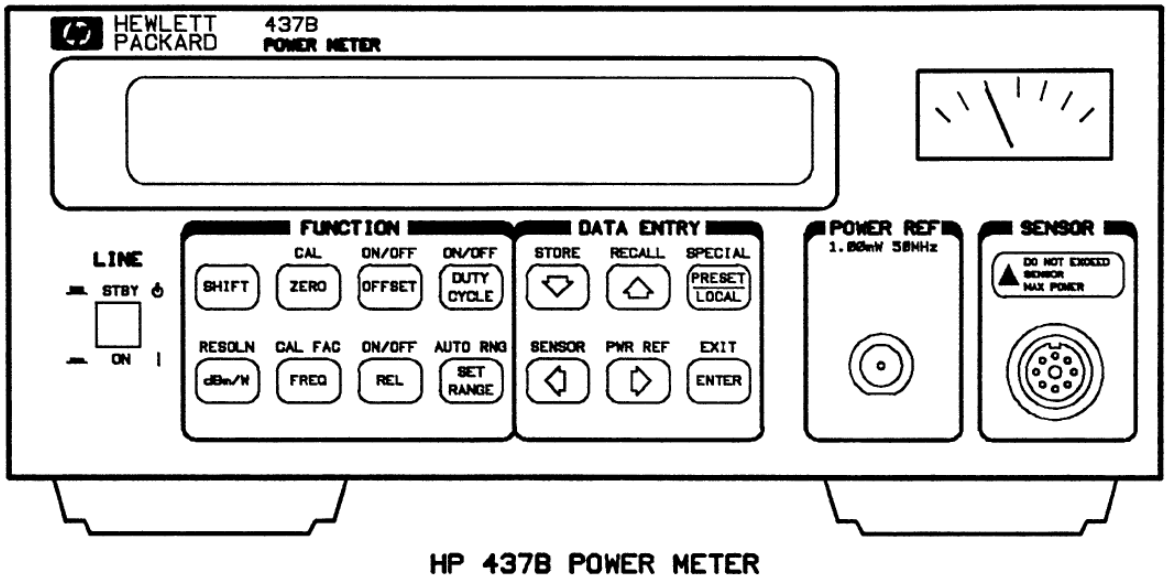

Figure C.1: Front panel of HP 437B power meter.

Appendix D

S-parameter setup

A typical S-parameter using PNA series network analyzer is described in following 13 steps. Assume Keysight's IC-CAP is used to record and process S-parameter data.

1. Press Line switches on DC supply, PNA and power meter to ON and wait for 30 minutes.

2. Finish planarization of RF probes and alignment check.

3. Set up "Sweep type" and "Average" on PNA;

4. Open "GPIB" on IC-CAP main interface to see GPIB interface is working properly as shown on Figure D.1.

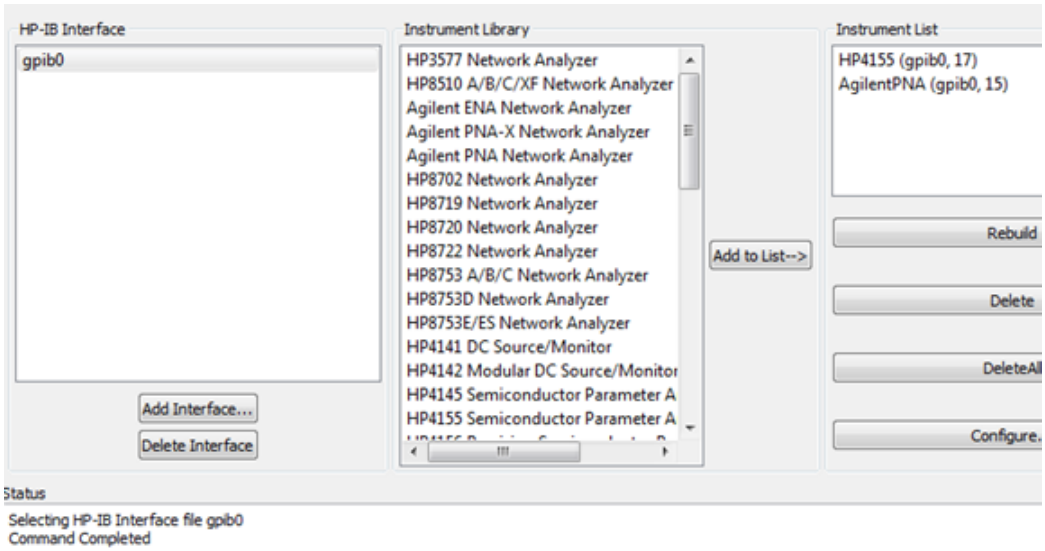

Figure D.1: Check and configure HP-IB interface.

5. Verify if GPIB addresses for DC supply and PNA are listed correctly as shown on Figure D.2.

Instrument	PAD	SAD	Identification
Instrument 0	15	None	Agilent Technologies,E8364B,MY43040665,A.06.04.32
Instrument 1	17	None	HEWLETT-PACKARD,4155A,0,01.00:01.00:01.00

Figure D.2: Check and configure HP-IB interface.

6. Check if power meter is listed on PNA's master GPIB interface as shown in Figure D.3.

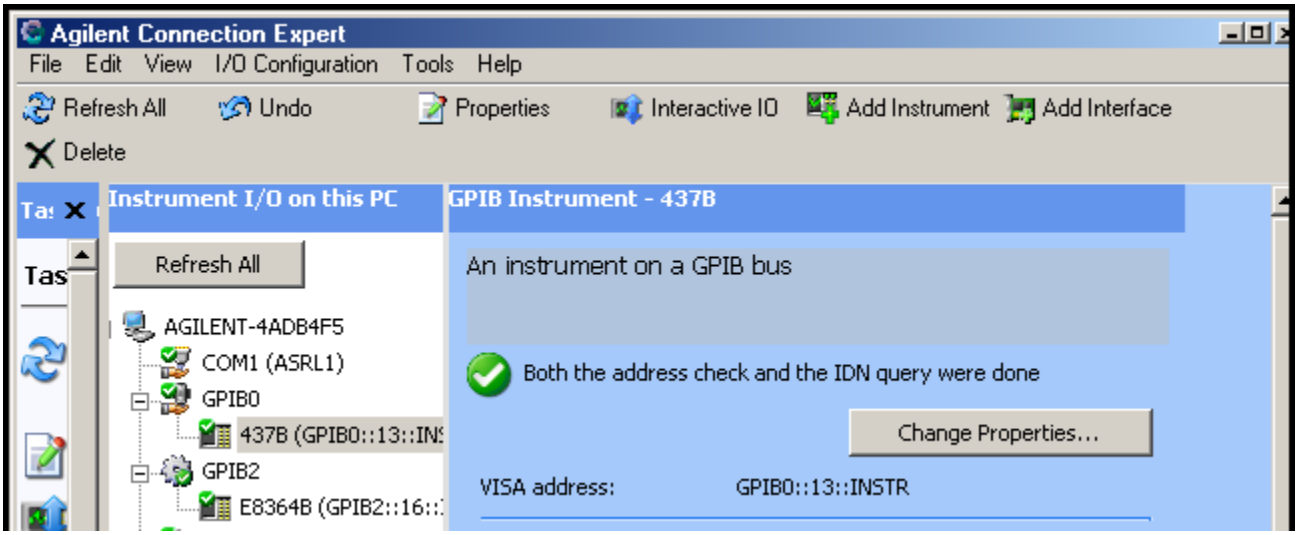

Figure D.3: Check power meter's communication with PNA.

7. Open "Source Power Cal" dialog window on PNA and edit "Cal Factor" of "Power Sensor Settering" if needed.

8. Place input RF cable between PNA's "PORT1" and power meter, press "Take Cal Sweep", and wait until "PASS" sign shows up.

9. Select proper calibration kit for corresponding RF probes from "Advanced Modify Cal Kit" on PNA.

10. Try a series of output characteristics measurement with different PNA input power values to figure out the maximum allowed input signal, above which output curve with RF input is different from that without RF input. Take the HBT measurement in

Figure D.4 as an example, the best input power should between -45 and -35 dBm as the latter disturbs DC biasing of S-parameter measurement.

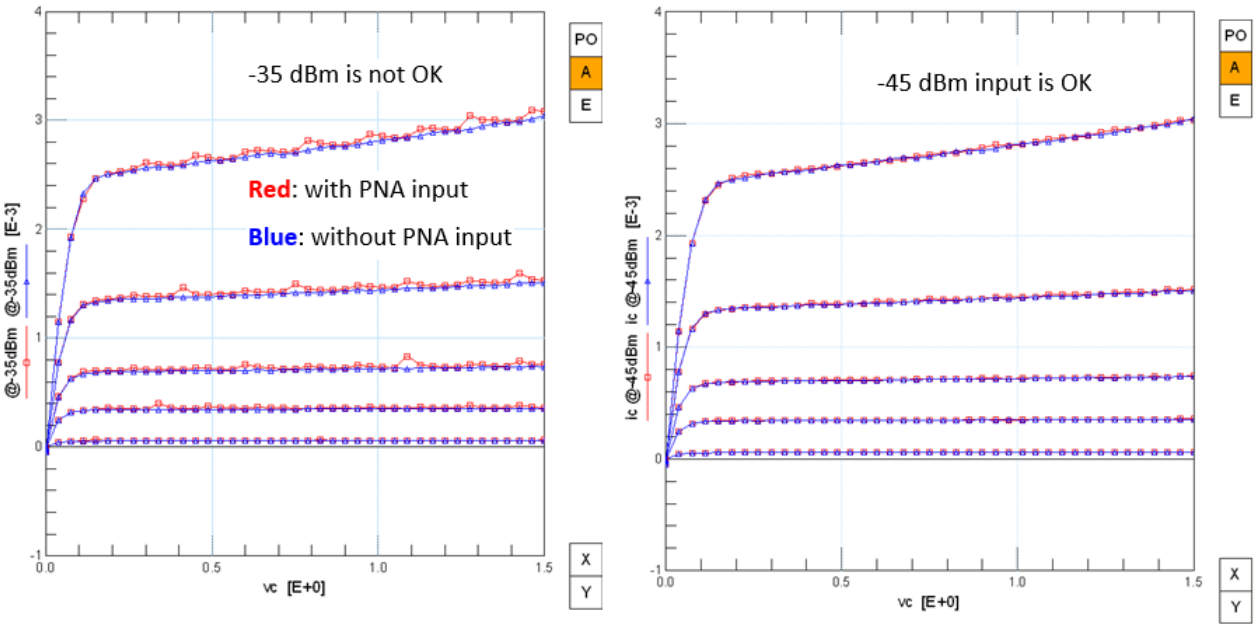

Figure D.4: Verify if a chosen input signal is appropriate.

11. Select "UNGUIDED Calibration" with "2 Port Solt" on PNA and follow instructions to measure SOLT calibration structures of "OPEN", "SHORT", "LOADS" and "THROUGH" for "PORT1" and "PORT2"; Save calibration result as a ".csa" file; Make sure "SrcPwrCal" status appears at bottom of PNA screen.

12. Following the example in Figure D.5, type the same ".csa" file name in "Cal/State File Name" in IC-CAP and measure four calibration structures to verify if SOLT calibration is valid as shown in Figure D.6.

13. Measure on-wafer de-embedding structures and device under test by using IC-CAP as controlling program.

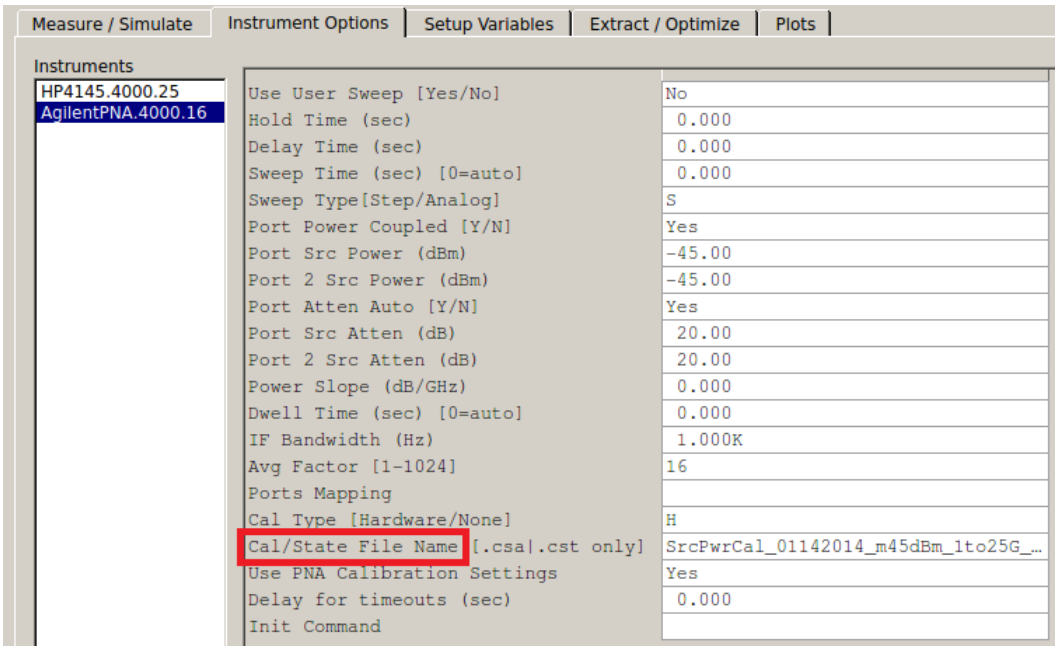

Figure D.5: Use PNA calibration settings in IC-CAP.

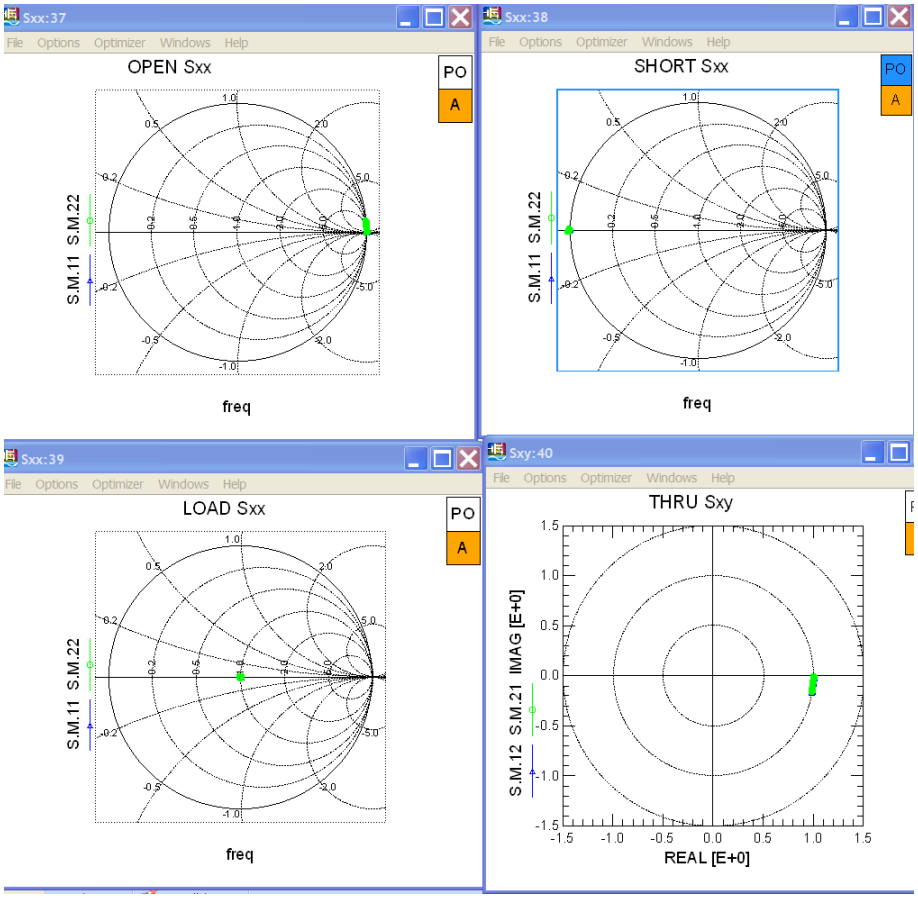

Figure D.6: Verify SOLT calibration.