RF LINEARITY ANALYSIS IN NANO SCALE CMOS USING HARMONIC BALANCE DEVICE SIMULATIONS

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

Deepika Kopalle

Certificate of Approval:

Adit Singh James B. Davis Professor Electrical and Computer Engineering	Guofu Niu, Chair Professor Electrical and Computer Engineering
Fa Foster Dai Associate Professor Electrical and Computer engineering	Stephen L. McFarland Acting Dean, Graduate School

RF LINEARITY ANALYSIS IN NANO SCALE CMOS USING HARMONIC BALANCE DEVICE SIMULATIONS

Deepika Kopalle

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

Auburn, Alabama
July 29, 2005

RF LINEARITY ANALYSIS IN NANO SCALE CMOS USING HARMONIC BALANCE DEVICE SIMULATIONS

Deepika Kopalle

_____
Signature of Author

_____
Date

Copy sent to:

Name Date

Vita

Deepika Kopalle, eldest daughter of Panduranga Vittal Kopalle and Purna Kopalle was born on Aug 7, 1978 in Hyderabad, India. She finished her high school from Ratna Junior College, Hyderabad in 1996 and entered Sri Krishnadevaraya University for her Bachelors in Electronics and Communication Engineering in 1997. She joined Auburn University for her Masters in Electrical and Computer Engineering in 2002 under the guidance of Dr.Guofu Niu.

Thesis Abstract

RF LINEARITY ANALYSIS IN NANO SCALE CMOS USING HARMONIC BALANCE DEVICE SIMULATIONS

Deepika Kopalle

Master of Science, July 29, 2005
(B.Tech.,SKDU,India,2001)

93 Typed Pages

Directed by Guofu Niu

In this thesis, Intermodulation Linearity characteristics of CMOS have been analyzed using power series and Harmonic Balance(HB) Method. Harmonic Balance method is a frequency domain steady state analysis method used for solving nonlinear circuits. This method is extended to semiconductor device simulation using Taurus-Device tool. Third order Input Intermodulation Product (IIP3), a measure for linearity is characterized as a function of channel length, oxide thickness, drain and gate voltages using 130nm, 100nm and 90nm MOS devices. The effect of Polysilicon gate depletion on linearity is studied and analyzed for different doping concentrations. Further, the simulated IIP3 values obtained from Harmonic Balance method are compared to the theoretical values calculated using power series.

Acknowledgments

I would like to express my deepest gratitude to Dr. Guofu Niu for his constant support and guidance. This thesis would not have been possible without his motivation and encouragement. I would like to thank my committee members Dr. Adit Singh and Dr. Foster Dai for their valuable input.

I am grateful to Dr. J.M.Wersinger from the Department of Physics and Dr. Luke Marzen from Geology and Geography Department for their financial support throughout my Master's degree.

Thanks are due to Dr. Qingqing Liang from Georgia Tech., and my colleague Ms. Yan Cui for their help in my research. I would like to extend special thanks to my friend Ms. Sailaja Chilaka for her constant support and help in my research.

I am deeply indebted to my parents Mr. K.P.R.Vittal and Mrs. Purna Vittal and my sister Ms. Manasa whose love and support gave me the strength to overcome the hurdles throughout my career. I am grateful to all my friends at Auburn for their love, support and encouragement towards achieving this goal.

Style manual or journal used Journal of Approximation Theory (together with the style known as "aums"). Bibliograpy follows van Leunen's *A Handbook for Scholars*.

Computer software used The document preparation package TEX (specifically LATEX) together with the departmental style-file `aums.sty`.

Table of Contents

List of Figures

Chapter 1

INTRODUCTION

With rapid growth of wireless communication systems, the use of CMOS technologies has been extended to RF applications. Many RF figures of merit like cutoff frequency and noise figure have been improved with scaling. Semiconductor device simulations have proven to play a key role in the design and development of Analog and RF systems. Thus, there has been an increasing need for analysis and modelling of RF figures of merit at semiconductor device level.

In this work, the effects of technology scaling on linearity and RF distortion have been analyzed using power series and Harmonic Balance(HB) method using Taurus, a process and device simulation tool from Synopsys.

1.1 Background

Linearity is one of the key parameters for RFIC design and it refers to the ability of a device, circuit or a system to amplify the input signal in a linear fashion [1]. In an ideal system, the output is linearly related to the input. However, in any real-time device or system the transfer function is more complicated, which can be due to presence of active or passive devices in the circuit or signal swing limitation of the power supply rails [1]. While all electronic circuits are mostly nonlinear, some circuits such as small signal amplifiers are very weakly nonlinear, hence are used in systems as if they are linear [2].

Nonlinearities in circuits have both advantages and disadvantages. Nonlinearity is required to translate frequency from baseband to RF and vice versa depending on whether the signal is transmitted or received. Nonlinearity is also necessary to build an oscillator and to realize frequency multiplication used in frequency synthesis. Inspite of the mentioned advantages, nonlinearity in circuits may also create distortion in the desired signals. It causes intermodulation of two adjacent strongly interfering signals at the input of a receiver, which can corrupt the nearby desired weak signal [1]. Nonlinear circuits usually generate a large number of frequencies and hence are more complicated to analyze when compared to linear circuits.

Nonlinear circuits are often characterized as either strongly nonlinear or weakly nonlinear [2]. Strongly nonlinear circuits are very complicated and are analyzed using HB or time domain methods. Weakly nonlinear circuits can be described by Taylor series expansion of their nonlinear current-voltage(I-V) characteristics. Most of the transistors and passive components are weakly nonlinear and are analyzed using power series or Volterra series [2].

1.2 Motivation

Traditional time domain approaches though extremely effective, often fall short when applied to simulating steady state quantities such as harmonic distortion, due to long time constants or widely separated spectral components. Harmonic balance, a nonlinear frequency domain analysis technique has emerged as a widely accepted solution to many of the shortcomings that conventional time

domain simulators have in high frequency analog arena [3]. With the development of commercial harmonic balance simulator and compact software, nonlinear frequency domain analysis has assumed its current position as the method of choice for simulating most nonlinear microwave and RF circuits.

HB simulation has been in use for quite some time to simulate harmonic and intermodulation distortion at device level. In this work, HB method is applied to the semiconductor device simulation using Taurus process and device simulation tool.

1.3 Overview and Organization

This thesis is organized into 8 chapters. Chapter 2, discusses the basic concepts of Harmonics, Intermodulation and the figure of merits for linearity. In Chapter 3, the advantages of HB over transient analysis are discussed. A brief description of the solution methods used for HB simulation and its implementation in Taurus is presented. Chapter 4 gives an overview of RF CMOS linearity. Transconductance g_m and the effect of third order g_m nonlinearity coefficient g_{m3} on linearity is described. In chapter 5, 100nm MOS HB simulation results are presented and parametric analysis of IIP3 with channel length, oxide thickness is presented. Simulation results are compared to the theoretical values obtained from the power series. Chapter 6 discusses CMOS scaling and polysilicon gate depletion effects. Effect of polydepletion on linearity is analyzed for different doping concentrations of polygate. Chapter 7, extends the analysis for 90nm MOS device

with halo doping. Chapter 8, summarizes the work presented in this thesis and also extends the scope of this work.

Chapter 2

HARMONICS AND INTERMODULATION BASICS

One of the important properties of a nonlinear system is its generation of harmonics of the excitation/fundamental frequency. In narrow band systems, such harmonics may not be a serious problem as they are far from the signals of interest and are rejected by filters [4]. In others, such as transmitters harmonics may interfere with desired weak signal or other communication systems and must be reduced using filters.

Most of the nonlinearity concepts can be briefly analyzed using simple power series. This technique is relatively simple but requires an unrealistic assumption that the circuit contains only ideal memoryless transfer nonlinearities. However, power series approach is useful in some instances and gives a good intuition of a nonlinear circuit behavior. It is a simple mathematical representation, which gives direct response of a nonlinear device or system in frequency domain. Hence, can be easily applied to most analog, RF and microwave applications. For a small signal input $x(t)$, the output voltage $y(t)$ of memoryless nonlinear circuit can be expressed using power series as

$$y(t) = k_1 x(t) + k_2 x^2(t) + k_3 x^3(t) + \cdots \tag{2.1}$$

for simplicity, higher order nonlinearity terms are not considered. Using power series, the concepts of harmonics and intermodulation are discussed briefly in the following sections. Most of the concepts presented here are directly extracted from [1], [4], [5] with pertinent changes.

2.1 Harmonics

If a sinusoidal input $x(t) = A\cos\omega t$ is applied to a nonlinear circuit, the output $y(t)$ is given by

$$y(t) = k_1 A\cos\omega t + k_2 A^2\cos^2\omega t + k_3 A^3\cos^3\omega t. \tag{2.2}$$

Equation (2.2) can further be expressed as

$$y(t) = \frac{k_2 A^2}{2} + (k_1 A + \frac{3k_3 A^3}{4})\cos\omega t + \frac{k_2 A^2}{2}\cos 2\omega t + \frac{k_3 A^3}{4}\cos 3\omega t. \tag{2.3}$$

In equation (2.3), first term is the dc shift, second term with the input frequency is the "fundamental", and other higher order terms are the "Harmonics", which are integral multiples of the fundamental frequency. For small A, higher powers of A can be neglected and therefore, the n^{th} harmonic is proportional to A^n.

2.2 Intermodulation

When two signals with different frequencies are applied to a nonlinear system, the output exhibits some components that are harmonics of neither input frequencies. Such frequencies, called Intermodulation (IM) products arise from the mixing of two signals.

IM products in an amplifier or communication receiver create serious problems since they represent spurious signals that interfere with and can be mistaken for desired signals. If a two tone input voltage $x(t) = A\cos\omega_1 t + A\cos\omega_2 t$, is applied to a nonlinear system then the output $y(t)$ is given by

$$y(t) = k_1(A\cos\omega_1 t + A\cos\omega_2 t) + k_2(A\cos\omega_1 t + A\cos\omega_2 t)^2 + k_3(A\cos\omega_1 t + A\cos\omega_2 t)^3. \tag{2.4}$$

Expanding the above equation

$$\begin{aligned} y(t) = &\left(k_1 A + \frac{3k_3A^3}{4} + \frac{3k_3A^3}{2}\right)\cos\omega_1 t + \ldots \qquad \text{fundamental} \\ &+ \frac{3k_3A^3}{4}\cos(2\omega_2 - \omega_1)t + \ldots \qquad \text{intermodulation.} \end{aligned} \tag{2.5}$$

It can be observed that the IM products are generally much weaker than the signals that generate them. However two strong signals outside the passband may generate an IM product which is within the passband, and which in turn may obscure the desired weak signal at the same frequency [4].

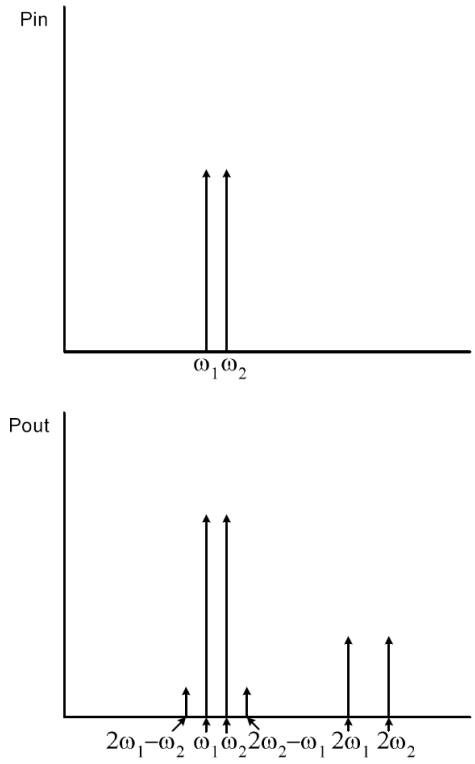

Figure 2.1: Harmonics and third order intermodulation products.

Fig. 2.1 shows the fundamental and harmonics generated for strong two tone interference. Even order IM products usually occur in frequencies well above or below the signals that generate them, hence are of little concern. The IM products of greatest concern are the third order IM products that occur at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ frequencies. They are the strongest of all odd order products and are close to the signals that generate them and often cannot be rejected by filters.

The corruption of signal due to the third order intermodulation has troublesome effects in RF systems and is very critical. Hence, a performance metric has

been defined to characterize this behavior and is called "Third order intercept point"(IP3) and can be measured by a two tone test. From equation 2.5 it can be noticed that fundamentals increase proportional to A, whereas the third order IM products increase proportional to A^3. The third order intermodulation distortion (IM3)is defined as

$$IM3 = \frac{3k_3A^3}{4}/k_1A = \frac{3}{4}\frac{k_3}{k_1}A^2. \tag{2.6}$$

Thus a 1-dB increase in input results in 1-dB increase in fundamental output while 3-dB increase in IM product. Fundamental output and IM3 product are plotted versus input on a logarithm scale as shown in Fig.2.2.

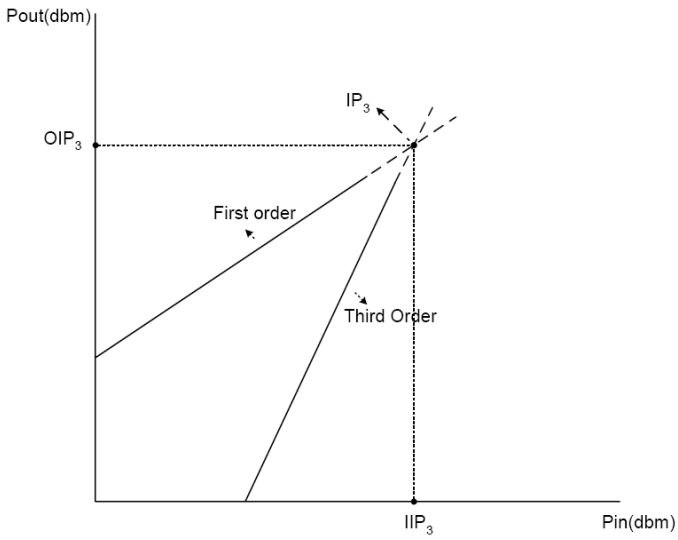

Figure 2.2: First and third order output powers vs input power.

The third order intercept point is defined as the intersection of the two lines. The horizontal coordinate of this point is called the input IP3(IIP3) and the vertical

coordinate is called the output IP3(OIP3). IIP3 can be obtained by making $IM3 = 1$, hence

$$IIP3 = \sqrt{\frac{4k_1}{3k_3}}. \tag{2.7}$$

IIP3 is more useful since it does not depend on the input signal level and can serve as a means of comparing linearity of different circuits. IIP3 can be expressed in terms of IM3 as

$$IIP3 = \frac{A^2}{IM3}. \tag{2.8}$$

On log scale they can be expressed as

$$10 \log IIP3 = 20 \log A - 10 \log IM3. \tag{2.9}$$

Equation 2.9 can be rewritten in terms of power as

$$P_{IIP3} = P_{in} + \frac{1}{2}(P_{o1st} - P_{o3rd}). \tag{2.10}$$

However, in practise if the input is increased to reach the intercept point, higher order IM products may become significant and in many circuits IP3 is beyond the allowable input range. Thus, the practical method for obtaining the IP3 is to measure the characteristics for small input amplitudes and use linear interpolation on a log scale to find the intercept point.

Chapter 3

HARMONIC BALANCE FUNDAMENTALS

Harmonic balance analysis is one of the most important techniques used for analyzing strongly or weakly nonlinear circuits that have single or multi tone excitation. This method is based on balancing currents between the linear and nonlinear sub circuits, hence it is named so. Harmonic balance is used to calculate steady state response of a circuit in frequency domain. The essential characteristic of this method is to implement circuit equations in the frequency domain. It can be summarized as the method where Kirchoff's current law is formulated in the frequency domain.

3.1 Transient vs Harmonic Analysis

Traditional time domain approaches often fall short when applied to simulating the steady state response of systems with long time constants or widely separated spectral components. For many high frequency (RF and microwave) applications, the solution of the state equations by standard transient methods can be prohibitively expensive. Harmonic balance solves the state equations in the frequency domain, and is almost completely insensitive to widely varying time constants, tone spacings and incommensurate frequencies. Transient analysis uses standard numeric integration, constructs a solution as a collection of time samples with an implied interpolating function. This interpolation is usually a polynomial

and polynomials fit sinusoids poorly, hence require more points to approximate sinusoidal solutions. Harmonic balance on the other hand uses a linear combination of sinusoids to build the solution. Thus, periodic and quasi periodic signals found in a steady state response can be approximated more accurately. It requires a small data set if the steady state response consists of only few dominant sinusoids.

3.2 Harmonic Balance Method and Example

One of the major difficulties with the harmonic balance approach is to compute the response of the nonlinear device. It is difficult to compute the coefficients of response directly from the coefficients of the stimulus. Hence the coefficients of the stimulus can be converted into a sampled data representation, which implies a frequency to time domain conversion. This can be done using Inverse Fourier Transform. This conversion helps in determining the response of the nonlinear devices accurately. The results are then back converted into coefficient form(frequency domain) using Forward Fourier Transform. The coefficients of the steady state response are now an algebraic function of the coefficients of the stimulus. Thus the nonlinear integro-differential equations that describe a circuit are converted by harmonic balance into a system of algebraic nonlinear equations , which when solved give the steady state response of the circuit [6]. These equations can be solved iteratively to get a steady state solution.

The basic implementation of the harmonic balance method has been explained in [2] using a simple example and is presented here for better understanding.

3.2.1 Example

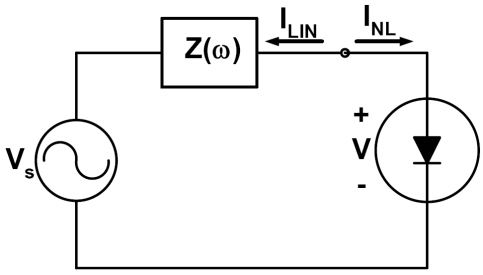

Figure 3.1: A Diode excited by an RF circuit.

Consider a simple circuit consisting of an RF source, a diode and an impedance $Z(\omega)$ as shown in Fig. 3.1. The diode when excited with an RF source at a frequency ω_p generates harmonics of current and voltage. Also, $Z(\omega)$ may vary with the harmonic frequency hence, can be written as $Z(k\omega_p)$ where k is the harmonic number. Assuming that the diode voltage consisting of its complex components at all harmonic frequencies $(k\omega_p)$ is known, the circuit can be sub-divided into linear and nonlinear circuits.

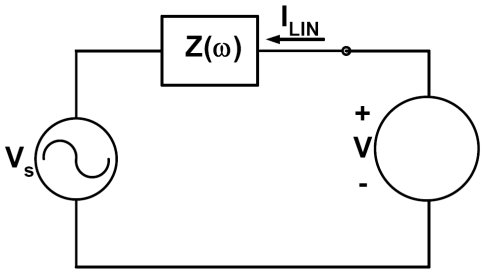

Figure 3.2: Equivalent Circuit describing the Linear part.

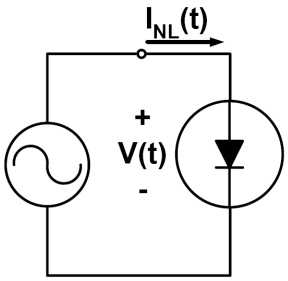

Figure 3.3: Equivalent Circuit describing the Nonlinear part.

The linear sub-circuit as shown in Fig. 3.2 can be analyzed in the frequency domain as

$$I_{lin}(k\omega_p) = \frac{V(k\omega_p) - V_s(k\omega_p)}{Z(k\omega_p)} \tag{3.1}$$

where V_s is periodic. Using Fourier theory $V(k\omega_p)$ is converted into time waveform $V(t)$. Considering the nonlinear sub circuit as shown in Fig. 3.3, diode current can be represented as

$$I_{nl}(t) = I_{sat}(\exp(\delta \cdot V(t)) - 1) \tag{3.2}$$

where $\delta = \frac{q}{\eta kT}$.

Using fourier transformations equation(3.2) can be converted to $I_{nl}(k\omega_p)$. To find out if $V(k\omega_p)$ is a solution, Kirchoff's Current Law is applied at all harmonics

$$I_{lin}(k\omega_p) + I_{nl}(k\omega_p) = 0, \tag{3.3}$$

and an error function is defined as

$$f_k = I_{lin}(k\omega_p) + I_{nl}(k\omega_p). \tag{3.4}$$

Equation(3.1)and equation(3.2) are substituted in equation(3.3) at each harmonic. If equation(3.3) is satisfied, then the solution is found else the assumed diode voltage is modified and the process is repeated. Appropriate numerical method is used till f_k becomes negligibly small.

As the number of ports increases, the above equations become complicated.

3.3 Application to Device Simulation

Large memory and CPU requirements have been the major obstacles for application of the HB method to semiconductor device simulation. With the recent use of iterative solution methods for solving large scale HB problems, this method is applicable to device simulation on modern computer workstations.

HB method which is commonly used in circuit simulation has been implemented into Taurus. This method when implemented in a device simulator has many advantages:

- Non-quasi static effects due to the distributed internal potentials and carrier densities may be observed
- Tradeoffs between the fabrication process and the resulting analog behavior are more readily observed.

- Internal properties of the semiconductor device can be visualized.
- Many RF figures of merit can be simulated efficiently in the frequency domain.

The HB method as implemented in Taurus is described in the following section. Since the work of this thesis concentrates more on the application of this HB method to device simulation, the detailed description of the solution methods are not included. Details of this method can be obtained from [2], [6], [3].

3.3.1 Implementation in TAURUS Tool

In Taurus, the frequency domain solution of the periodically varying potential and carrier densities known as periodic steady state can be solved using HB method.The details presented in the following sections have been extracted from the Taurus manual [7]. If a periodic signal is applied to a semiconductor device at a frequency f_1, the nonlinear relationship between the applied potential and terminal currents will result in harmonics at integer multiples of f_1. The HB method solves the periodic steady state at each of these frequencies simultaneously.

The total flux into the device at each node in a semiconductor device for each carier type is zero. Thus at each node the sum of the current densities i(t)and the accumulated number of carriers $q(t)$ must be zero. This is represented as

$$i(t) + \frac{\partial q(t)}{\partial t} = 0. \tag{3.5}$$

HB method ensures that this law is observed at each frequency, which implies

$$I(f) + j2\pi f Q(f) = 0 \tag{3.6}$$

where $I(f)$ and $Q(f)$ are the frequency components of current and charge at frequency f and $j2\pi f$ represents the frequency domain equivalent of the time derivative operation. This equation needs to be solved for all frequencies simultaneously. Using fourier analysis, we have

$$b = I + \Omega Q = \Gamma \begin{bmatrix} i(0) \\ i(1) \\ \cdots \\ i(s-1) \end{bmatrix} + \Omega\Gamma \begin{bmatrix} q(0) \\ q(1) \\ \cdots \\ q(s-1) \end{bmatrix} \tag{3.7}$$

where Γ is the fourier transform operator and Ω is a diagonal matrix containing the time derivative entries. i and q are evaluated over s time samples for each node in the device. I and Q are the resulting current and charge vectors in the frequency domain. When $b = 0$ HB solution is found.

To find such solution in frequency domain, an iterative technique such as Newton's method is used. To use this method a Jacobian matrix which relates the changes in flux to the changes in the solution variables is required. The solution variable is referred to as 'v' although there are atleast three solution variables per

node in the semiconductor device simulation. Once the Jacobian is known, the solution can be found.

3.3.2 Multi-tone Simulation

Taurus supports upto three simultaneous tones using the multi dimension fast fourier transform. Since the response at higher intermodulation orders is less significant,the simulator reduces the number of solution variables by removing any frequencies that have a intermodulation order greater than that specified for the simulation.

The HB simulation is carried out as described above though the solution method for multi tone simulation may differ. As the magnitude of input signal increases, the number of frequencies to be solved also increases thus making Newton's method impractical.

A complex variant of a GMRES(Generalized Minimum residual) linear solver is used in the HB simulation [3]. The problem to be solved for each nonlinear iteration is then

$$r = AP^{-1}X - b \tag{3.8}$$

where r is residual, P is the preconditioner matrix and X is solution update vector. Linear solver tries to minimize r. The new HB solution for nonlinear iteration k at the end of the linear solve is then

$$v_k = v_{k-1} + P^{-1}X. \tag{3.9}$$

Preconditioner is defined as

$$P = \Omega \begin{bmatrix} \bar{g} & & & \\ & \bar{g} & & \\ \cdots & & & \\ & & & \bar{g} \end{bmatrix} + \Gamma \begin{bmatrix} \bar{c} & & & \\ & \bar{c} & & \\ \cdots & & & \\ & & & \bar{c} \end{bmatrix}$$

where

$$\bar{g} = \frac{1}{S} \sum_{s=0}^{s-1} g(s)$$

$\bar{c}$ is defined similar to $\bar{g}$ and S is the number of time samples required. Preconditioner is equivalent to the HB Jacobian except for entries relating coupling between the frequencies is neglected. Further information about the solution methods can be obtained in [2] [7].

Chapter 4

RF CMOS LINEARITY

With the growth of digital mobile communications, many RF amplifiers such as low-noise amplifiers (LNA) operate in a region of weak linearity at RF front-end [8]. Linearity is one of the important issues in RFIC design as it limits system's dynamic range. Since scaling of CMOS has resulted in a strong improvement in the RF performance of MOS devices, various performance metrics like noise figure, linearity have been widely studied [9]. As the channel length is decreased, thin gate oxide is needed to maintain electrostatic integrity. RF distortion is shown to be worse with decreasing oxide thickness [9]. Devices with high linearity can minimize signal distortion. Hence, analysis and simulation of linearity helps to understand limiting factors for a given technology and to optimize transistor structure and circuit topology [8]. For linearity, figure of merit $IP3$ is used as a first order parameter. Larger $IP3$ is required for higher linearity [9].

Fig. 4.1 shows first and third order powers vs input power and the extrapolated IP3 point obtained using taurus simulations.

4.1 RF Distortion

$IP3$ of CMOS devices has been studied recently using either measured or simulated I-V data [8], [9], [10]. Experimental characterization of linearity has also been reported recently [11]. In analog MOS circuits, a purely sinusoidal input

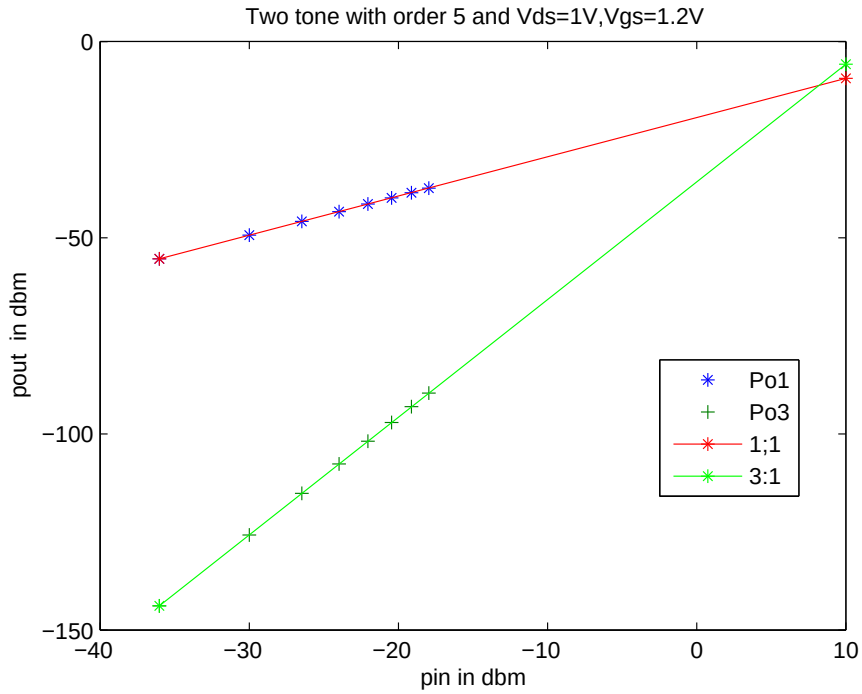

Figure 4.1: First and third order output powers vs input power.

can produce a distorted output signal with higher-order harmonics due to the nonlinearity of MOS transistors. These harmonics are mainly induced due to higer order derivatives of the current-voltage(I-V)characteristics [12]. Of particular interest is the third order derivative of the drain current I_{ds} with respect to the gate voltage V_{gs}, which needs to be minimized for low distortion application.

4.2 First Order Analysis

At first order, the drain current is simply a function of gate voltage and is represented as

$$I_{ds} = f(V_{gs}). \tag{4.1}$$

Hence the intermodulation is a function of ac V_{gs}. The transconductance g_m at a fixed V_{ds} can be defined as

$$g_m = \frac{\partial I_{ds}}{\partial V_{gs}}. \tag{4.2}$$

Increase of g_m in the weak inversion is faster because of the exponential nature of the I-V characteristics, and is slower in the strong inversion region. The second order derivative and third order derivatives of transconductance g_m at a fixed V_{ds} are given by

$$g_{m2} = \frac{\partial^2 I_{ds}}{\partial^2 V_{gs}} \tag{4.3}$$

and

$$g_{m3} = \frac{\partial^3 I_{ds}}{\partial^3 V_{gs}}. \tag{4.4}$$

Since the rate of increase of g_m is highest in the moderate inversion region, g_{m2} has it peak near the threshold voltage during the transition from sub-threshold to strong inversion. Also, the third order derivative g_{m3} is zero at this point. Since the third order derivative is zero, intermodulation product (IM3)is zero. Thus, IP3

has a peak value at this point. Once g_m, g_{m3} are known, IIP3 can be calculated as

$$IIP3 = \sqrt{\frac{4k_1}{3k_3}}. \tag{4.5}$$

k_1, k_3 are defined using power series

$$I_{ds} = g_m \cdot V_{gs} + k_2 \cdot V_{gs}^2 + k_3 \cdot V_{gs}^3 + \cdots \tag{4.6}$$

where $k_1 = g_m$ and

$$k_3 = \frac{1}{3!}\frac{\partial^3 I_{ds}}{\partial^3 V_{gs}}. \tag{4.7}$$

IIP3 is usually expressed in dBm as $10log(10^3 IIP3)$.

It is usually believed that IP3 has it peak value at the threshold point at which g_{m3} is zero. But recent studies have shown that the peak value may not be at the threshold point [11]. Further, IP3 value is observed to be higher in the strong inversion region for higher V_{gs} [11].

Fig. 4.2 shows the simulated g_m, g_{m2}, g_{m3} when plotted vs V_{gs} and the corresponding I-V characteristics for a 100nm MOSFET. It can be seen that the g_{m2} peak occurs at 0.33V. Hence, the IIP3 peak occurs at 0.33V. It can also be seen that g_m, g_{m3} are almost flat in the strong inversion region. Thus, from first order theory, the well known linearity sweet spot of gate bias can be easily found using the simulated or measured I-V data.

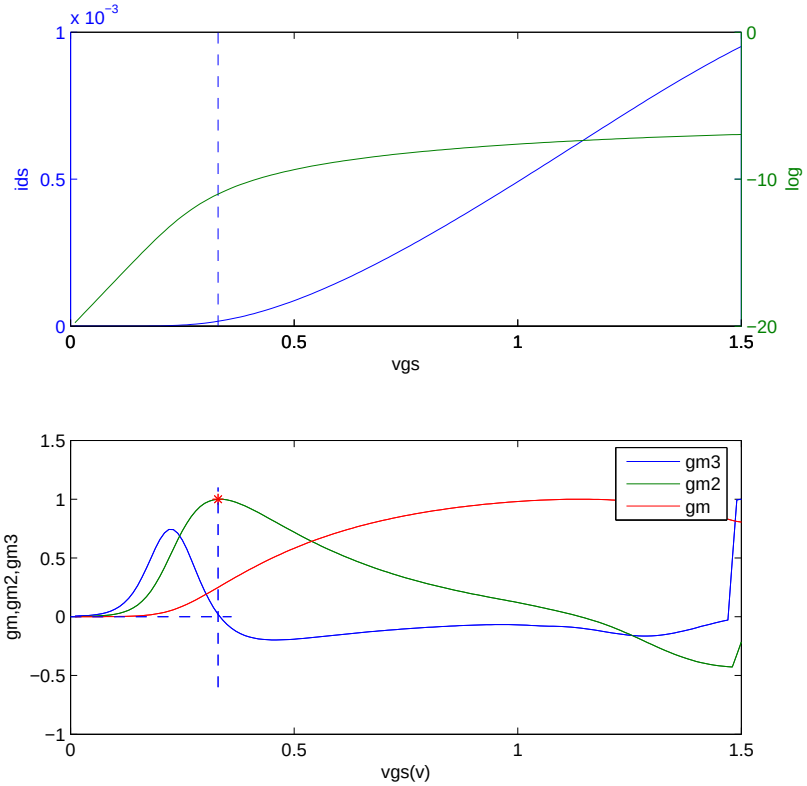

Figure 4.2: I-V characteristics and the corresponding gm, gm_2, gm_3 vs V_{gs}.

IIP3 at this sweet spot need not necessarily be the highest, since higher V_{gs} in strong inversion can lead to even higher IIP3 [11]. It is important to further analyze this deviation because it has significant implications in RFIC design. Further, the effect of CMOS scaling on RF distortion needs to be analyzed.

Chapter 5

$100nm$ MOSFET SIMULATION RESULTS

In this chapter, parametric characterization of intermodulation linearity is presented using 100nm technology. The impact of technology scaling on linearity is analyzed by varying the channel length and oxide thickness of a 100nm MOS device. IIP3 is also characterized as a function of drain and gate voltages.

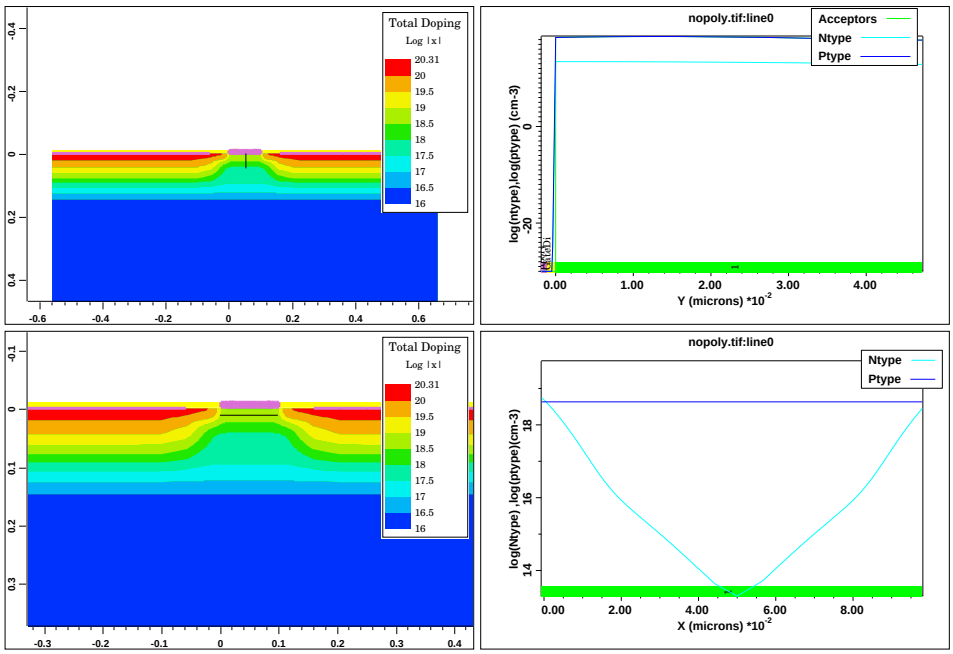

Figure 5.1: Doping profiles using a cut-line through the channel.

A simple 2-D MOS device with 100nm channel length, 1nm oxide thickness and uniform channel doping is used for linearity analysis.

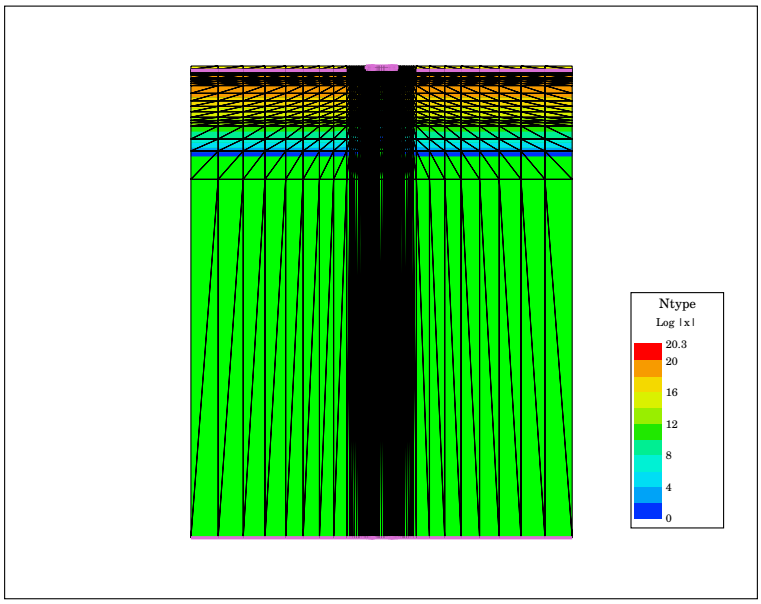

Figure 5.2: MOS 100nm mesh structure.

The device is built using Medici Device simulation tool and is visualized using Taurus. Input code for the device in Medici is included in Appendix A. Fig. 5.1 shows the typical doping profile when a cut-line is made through the channel.

5.1 DC Simulation

The device built in Medici is imported into Taurus and is further analyzed as described in the following sections. 100nm MOS structure is divided into 3135 grid points and the mesh structure is shown in Fig. 5.2.

At each of these grid points Taurus simulator solves Drift-Diffusion equations given by

$$\nabla \cdot (-\epsilon \nabla \psi) = q(p - n + N_D^+ + N_A^-) \tag{5.1}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla J_n + (G_n - R_n) \tag{5.2}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla J_p + (G_p - R_p) \tag{5.3}$$

where $J_n = q(n\mu_n\varepsilon + D_n\nabla n)$ and $J_p = q(p\mu_p\varepsilon - D_p\nabla p)$.

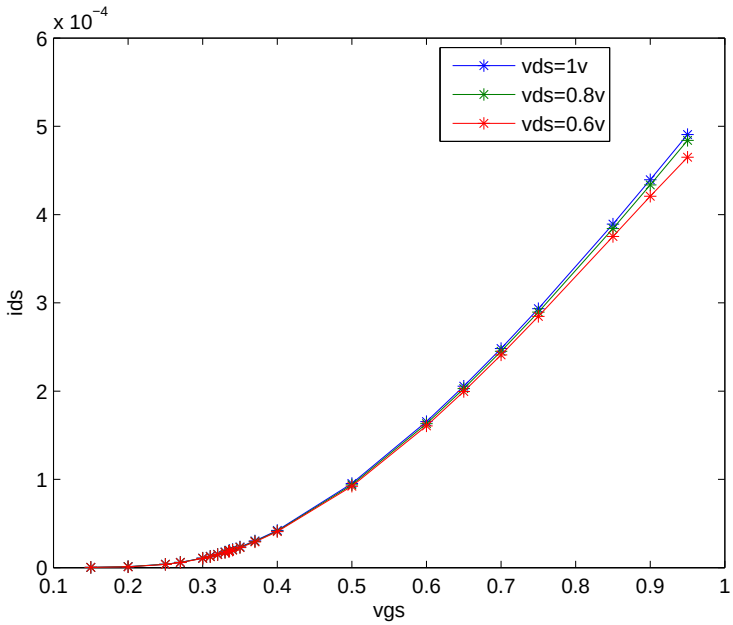

Figure 5.3: $I_d - V_g$ curves for three different values of V_{ds}.

Terminal currents are the most important characteristics of the device simulation and can be easily obtained by applying a simple DC bias. Current -Voltage output characteristics are always sought since they give an intuition about the device performance.

For a fixed drain voltage, the variation of drain current with gate voltage can be obtained from DC simulation. The $I_d - V_g$ curves for three different V_{ds} values are as shown in Fig. 5.3.

5.2 AC Simulation

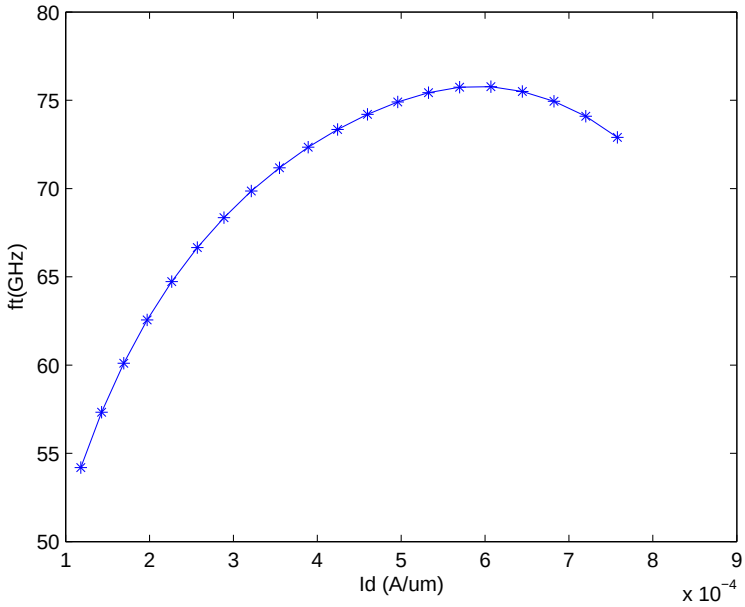

Figure 5.4: Cutoff frequency(f_T) vs Drain Current.

Cutoff frequency f_T, one of the important considerations in RFIC design can be obtained from AC simulation data. For a fixed drain voltage, AC analysis is applied keeping the frequency constant at 1MHZ and sweeping the gate voltage from 0.5 V to 1.2V. Cutoff frequency vs drain current is plotted and is as shown in Fig. 5.4. Peak f_T is found to be 75 GHz using drift-diffusion equations.

5.3 HB Simulation

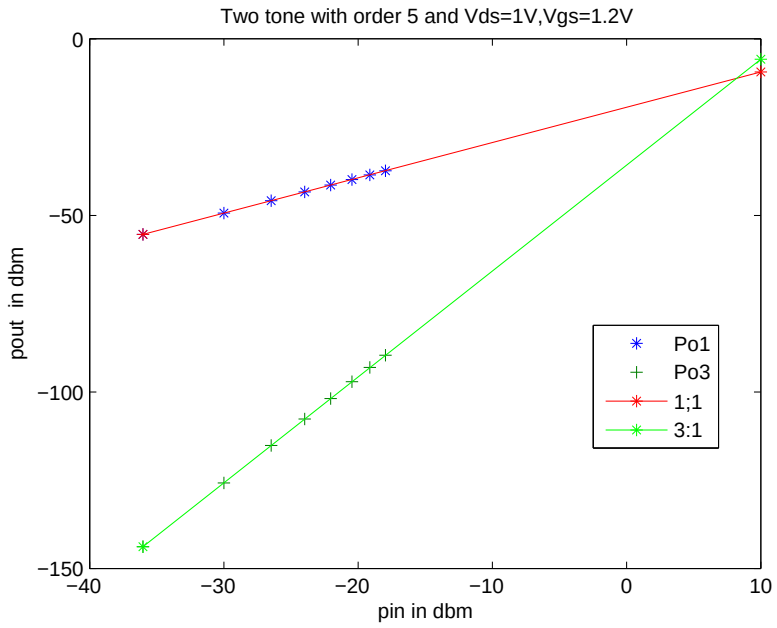

Figure 5.5: First and third order output powers vs input power.

HB analysis is done by applying a periodic signal to the 100nm MOS device at a fixed dc bias point of $V_{ds} = 1V$ and $V_{gs} = 1.2V$. Source voltage is swept from 10mV to 60mV and the corresponding output power for the harmonic balance simulation is plotted. Since a two-tone test is used for IIP3 calculations, two tones at frequencies 5.8 GHz and 5.9 GHz with a tone spacing of 1MHz are used. Source and load resistances are assumed to be 50 ohms. Truncation order for each tone is specified as five.

Input power is obtained as

$$P_{in} = \frac{V_{in}^2}{8R_s}. \tag{5.4}$$

First and third order powers are calculated as

$$P_{O1} = \frac{I_{w1}^2 R_l}{2} \tag{5.5}$$

$$P_{O3} = \frac{I_{2w2-w1}^2 R_l}{2} \tag{5.6}$$

where V_{in} is the source voltage, R_s and R_l are the source load resistances, I_{w1} and I_{2w2-w1} are the first and third order output currents. First and third order output powers obtained from the above equations are expressed in dBm and are plotted against input power as shown in Fig. 5.5. For higher input powers the simulation does not converge hence, the values are extrapolated. From Fig. 5.5, IIP3 value is found to be 8 dBm.

5.4 Characterization of IIP3

Characterization of IIP3 with respect to V_{gs}, V_{ds}, channel length and oxide thickness is presented in this section. 100nm MOS device with 1nm oxide thickness is used for the analysis, unless specified.

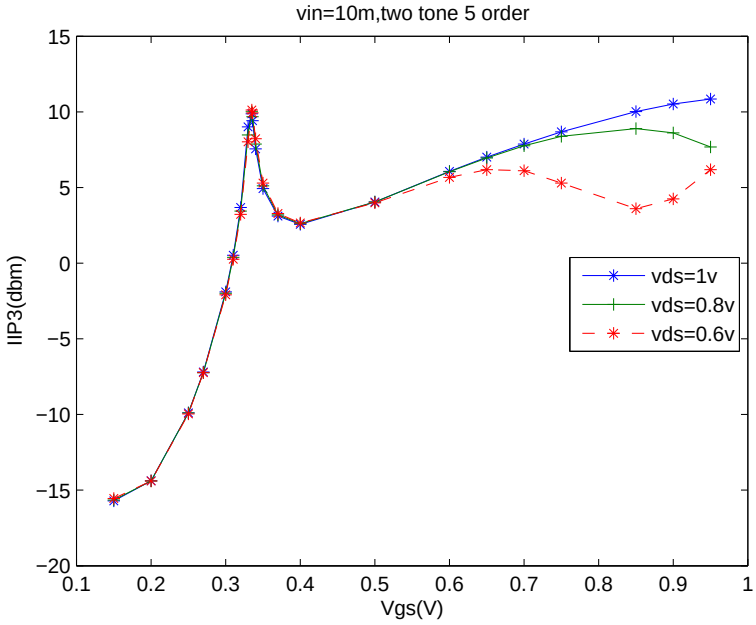

Figure 5.6: IIP3 vs V_{gs} for different V_{ds} values.

5.4.1 V_{gs} and V_{ds} Dependence

IIP3 values are plotted vs V_{gs} for three different values of V_{ds} as shown in Fig. 5.6. It can be seen that IIP3 has a sharp peak near threshold, during the transition between sub-threshold to strong inversion. At this point the second order g_m nonlinearity coefficient g_{m2} is highest, thus the third order nonlinearity coefficient g_{m3} is zero, leading to $IM3 = 0$. IIP3 value at this point was considered to be the highest but, recent studies have indicated that IIP3 values in the strong inversion may be much higher [11].

From Fig. 5.6, it can be observed that IIP3 peak does not change much in the weak inversion region as V_{ds} increases from $0.6V$ to $1.0V$, while it varies strongly

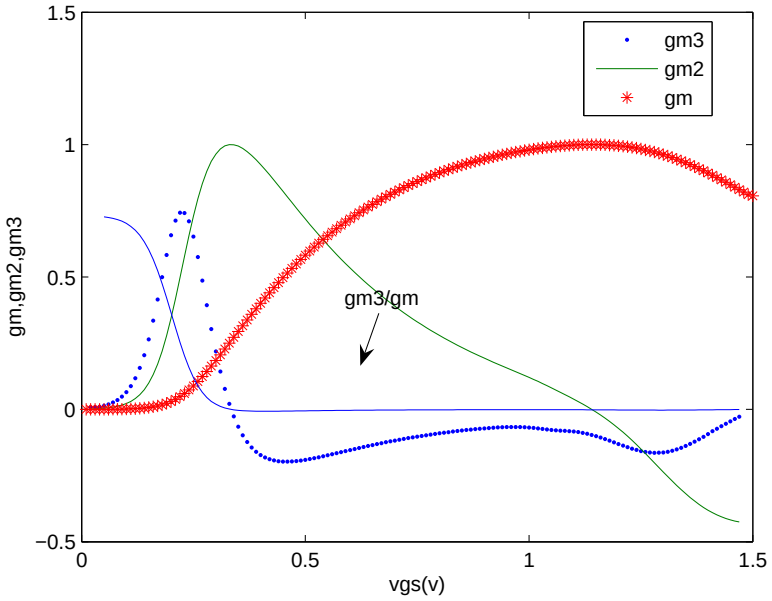

Figure 5.7: gm, gm_2, gm_3, gm_3/gm vs V_{gs} for 100nm MOS, W=1um.

with V_{ds} in strong inversion. Fig. 5.7 shows g_m, g_{m2} and g_{m3} plotted vs V_{gs} for a fixed $V_{ds} = 1$V. It can be seen that as V_{gs} increases from saturation to linear region, the ratio of g_{m3}/g_m remains constant. Hence, IIP3 remains constant at higher V_{gs}. IIP3 is also plotted versus drain current for different V_{ds} values as shown in Fig. 5.8. For same drain current, IIP3 increases with V_{ds} in strong inversion.

Fig. 5.9 shows Gain versus V_{gs} for 100nm MOS device with 1μm width. Gain also improves with V_{ds} but the increase is very weak. This weak increase in gain is due to increase in g_m with v_{ds} at higher V_{gs}. The negative gain values as seen from Fig. 5.9 are mainly because the device width is equal to 1μm. As the width

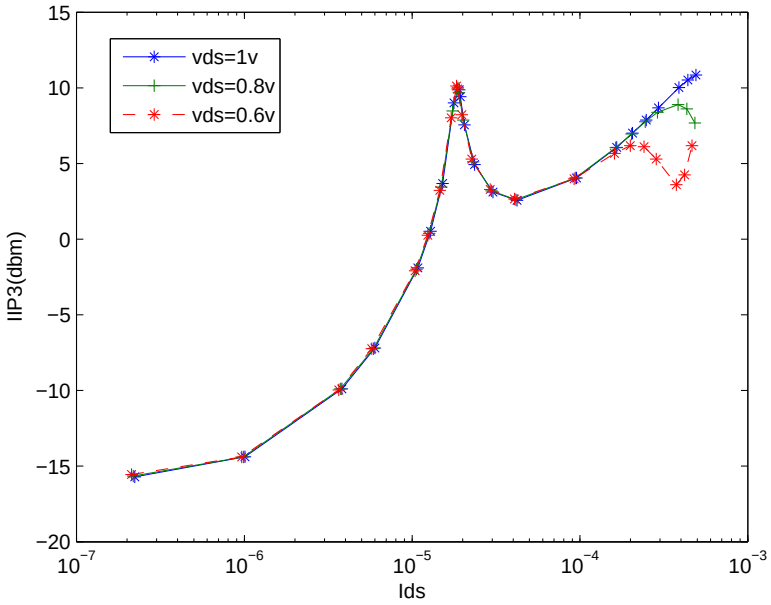

Figure 5.8: IIP3 vs I_{ds} for different V_{ds} values.

of the device increases, drain current increases and is given by

$$I_{ds} = I'_{ds} \cdot W. \tag{5.7}$$

First and third order output powers can then be expressed as

$$P_{o_3} = I'_{ds_3} \cdot W^2 \tag{5.8}$$

and

$$P_{o_1} = I'_{ds_1} \cdot W^2 \tag{5.9}$$

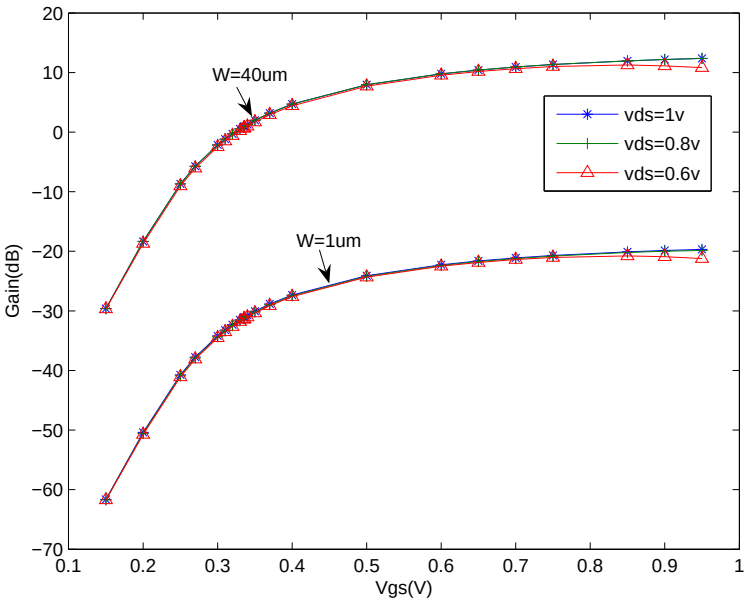

Figure 5.9: Gain vs V_{gs} for different V_{ds} values.

where $I'_{ds}, I'_{ds_3}, I'_{ds_1}$ are the drain currents obtained from the DC and HB simulations using Taurus and W is the width of the device.

Gain and IIP3 can be calculated as

$$G = \frac{P_{o_1}}{P_{in}} \cdot W^2 \tag{5.10}$$

$$IIP3 = P_{in} - \frac{1}{2}(P_{o_1} - P_{o_3}) \tag{5.11}$$

where P_{in} is the input power. Fig. 5.9 shows the increase in gain as the device width is scaled by 40. Thus, scaling the device width improves gain but does not

affect IIP3 since both first and third order powers are scaled by the same number. All the results presented in this chapter have $1\mu m$ device width.

5.4.2 Channel Length Dependence

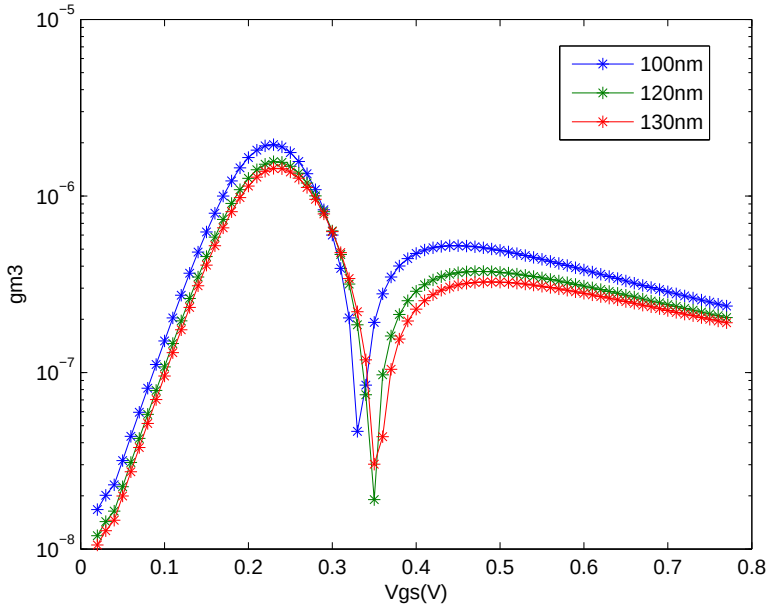

Figure 5.10: g_{m3} vs V_{gs} for different channel lengths.

Channel length is one of the important considerations for RFIC design. Scaling channel length though improves f_T, trades other performance metrics. Longer channel length reduces output conductance g_d and its nonlinearities in both weak and strong inversion. This is due to reduced drain-induced-barrier-lowering (DIBL) and weaker channel length modulation effects. In short channel devices, both these effects are significant and hence it is difficult to describe how channel length affects

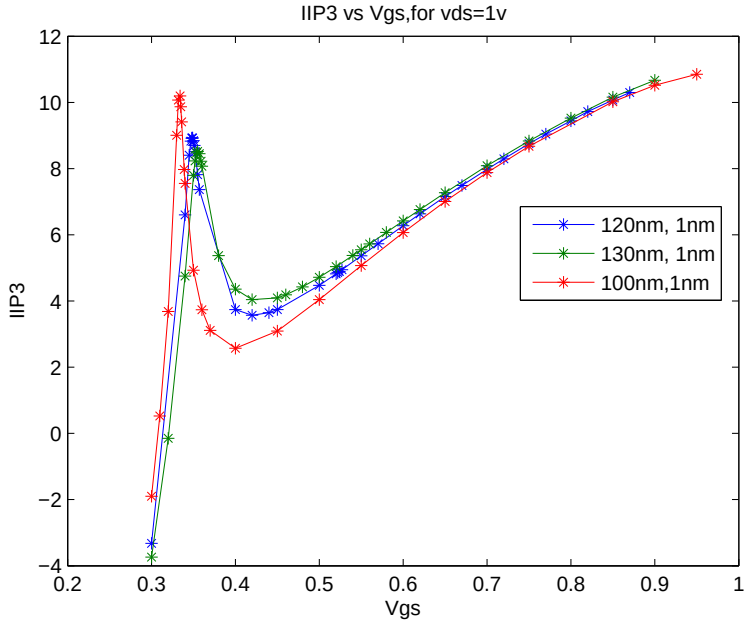

Figure 5.11: IIP3 vs V_{gs} for different channel lengths.

IIP3. However, g_{m3} can be used to explain the channel length dependence. Fig. 5.10 shows g_{m3} versus V_{gs} for different channel lengths. It can be clearly seen that RF distortion increases with scaling which in turn decreases IIP3.

Fig. 5.11 shows simulated IIP3 versus V_{gs} at 5.8 GHz frequency for three channel lengths 100nm, 120nm, 130nm. IIP3 is almost same for the devices in strong inversion.

As channel length decreases, threshold voltage reduces. Hence, IIP3 peak shifts to lower V_{gs} values. Since devices with different channel lengths have different I_{ds} values, it is important to analyze the effect of channel length on IIP3 at the same I_{ds} value. Fig. 5.12 shows IIP3 vs I_{ds} for different channel lengths. It can be

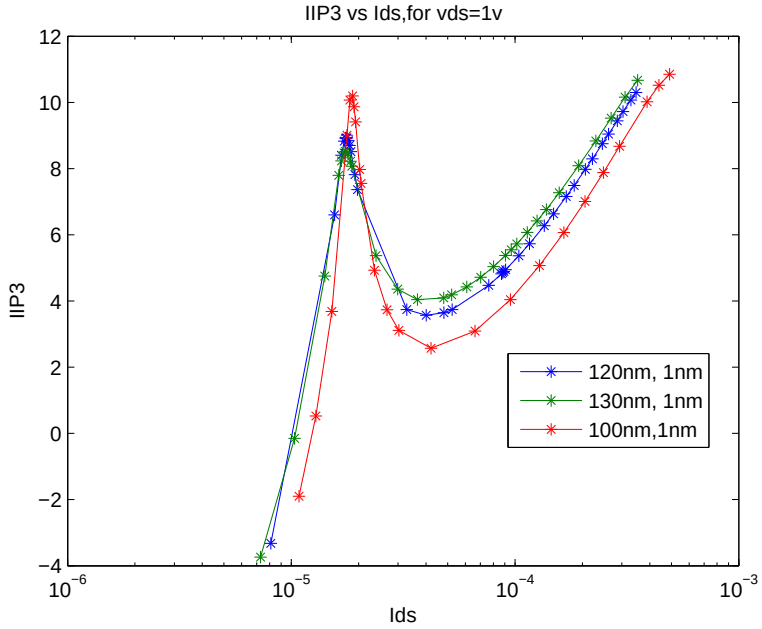

Figure 5.12: IIP3 vs I_{ds} for different channel lengths.

seen that IIP3 peak occurs at a higher current in a small channel device. However, at higher V_{gs}, IIP3 is higher for a longer channel device in strong inversion. Thus, longer channel length improves linearity provided gain requirement is satisfied.

5.4.3 Oxide Thickness

In scaled CMOS processes, multiple oxide thickness devices are provided to facilitate circuit designs requiring a higher voltage swing [8]. Linearity dependance on oxide thickness needs to be analyzed, since oxide scaling may be traded if g_m requirements are not high. The effect of oxide scaling on IIP3 is analyzed in this section by varying the oxide thickness of a 130nm MOS device. Fig. 5.13 shows

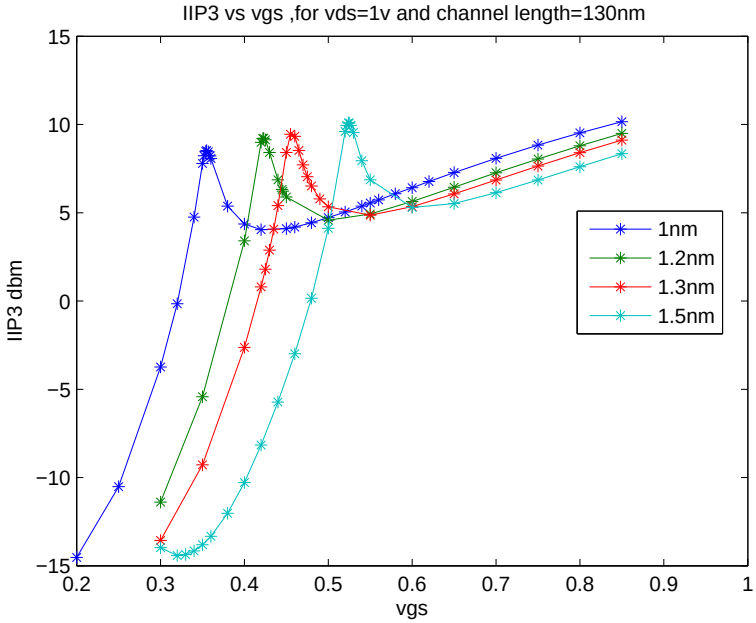

Figure 5.13: IIP3 vs V_{gs} for different oxide thickness.

IIP3 vs V_{gs} for oxide thicknesses 1nm, 1.2nm, 1.3nm and 1.5nm. All the devices are biased at same V_{gs} and have $V_{ds} = 1$V. A shift in threshold voltage is observed as the oxide thickness is increased. It can be seen that the IIP3 peak occurs at higher V_{gs} as the oxide thickness is increased. Thus, devices with thin oxide have smaller threshold values.

Fig. 5.14 shows IIP3 vs I_{ds} for different oxide thicknesses. For the same drain current, IIP3 is found to be higher for thicker gate oxide. Thus, oxide scaling trades linearity and hence increases RF distortion.

From Fig. 5.15, it can be seen that Gain improves with oxide scaling. Thus, short channel devices trade linearity to gain and threshold voltage.

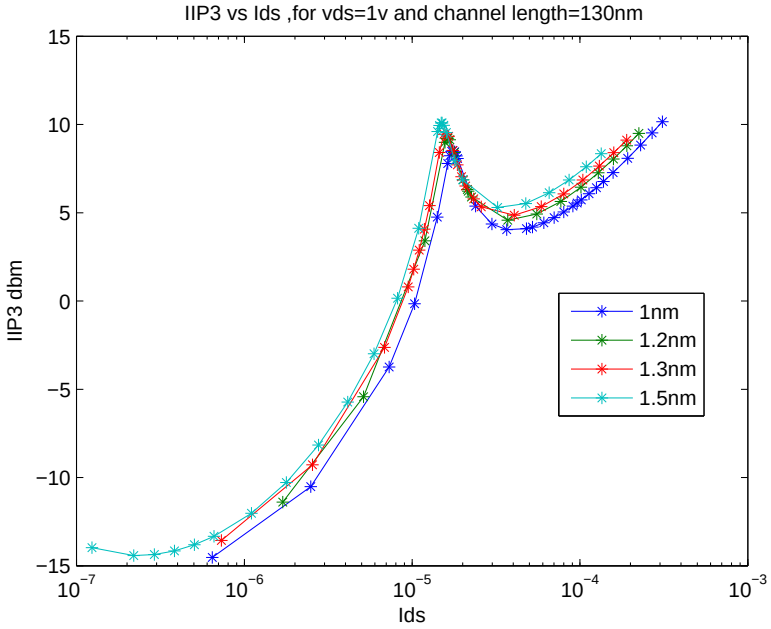

Figure 5.14: IIP3 vs I_{ds} for different oxide thickness.

5.5 Simulation and Theoretical Analysis

In this section, the simulation results obtained from Harmonic Balance are analyzed using power series. Power series is a simple mathematical representation used to obtain the direct response of a nonlinear system in frequency domain. This series when restricted to be Taylor's series around a predetermined quiescent point(usually the dc bias point), inherently represents the device's small signal behavior [6]. Thus, it can be used to predict a circuits weakly nonlinear behavior [13]. The nonlinear relation between drain current and gate potential can be expressed using power series by defining a number of nonlinear coefficients as described in

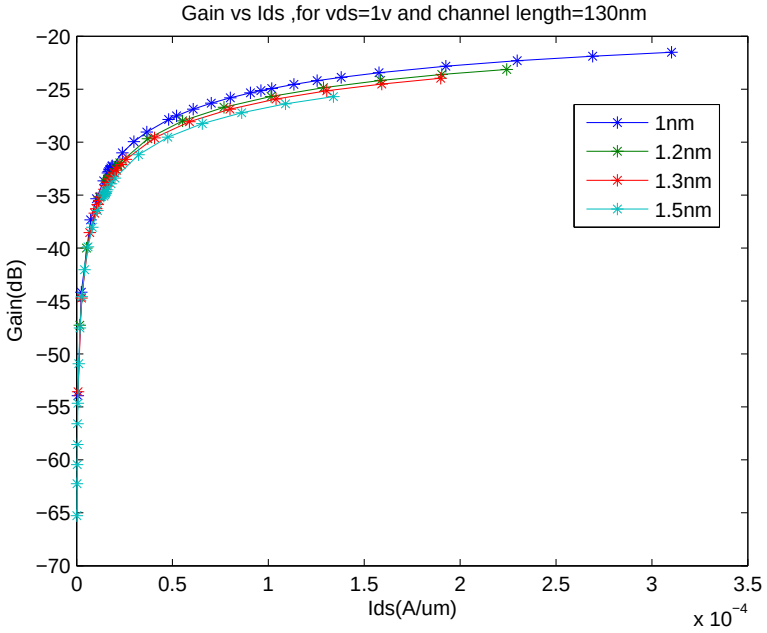

Figure 5.15: Gain vs I_{ds} for different oxide thickness, L=130nm, W=1um.

chapter 4. These coefficients help to characterize I-V nonlinearity. From DC simulation, g_m, g_{m_2}, g_{m_3} can be calculated and IIP3 as described in chapter 5 can be calculated theoretically as

$$IIP3 = \sqrt{\frac{4k_1}{3k_3}} \tag{5.12}$$

where $k_1 = g_m$ and $k_3 = \frac{1}{3!}\frac{\partial^3 I_{ds}}{\partial^3 V_{gs}}$. IIP3 is usually expressed in dBm as $10log(10^3 IIP3)$.

Simulated values obtained from Harmonic Balance and theoretical values computed using power series, for three devices with different channel lengths are plotted vs gate voltage as shown in Fig. 5.16. Fig. 5.17 shows the results when plotted versus the drain current on log scale. It can be seen that the values are in well

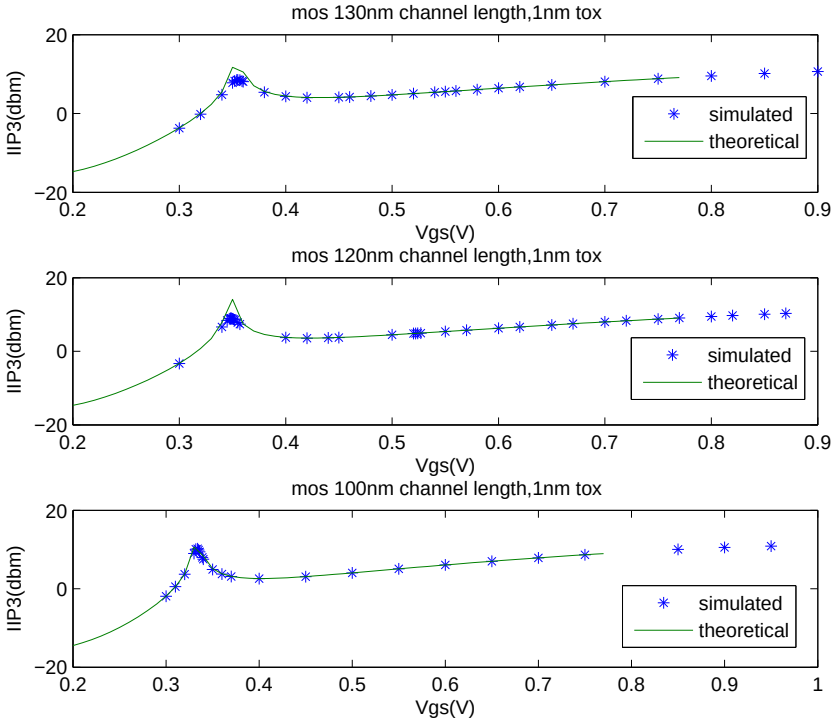

Figure 5.16: Simulated and theoretical values of IIP3 vs gate voltage V_{gs}.

agreement hence, the third order g_m nonlinearity can be used for analyzing linearity at the first order level.

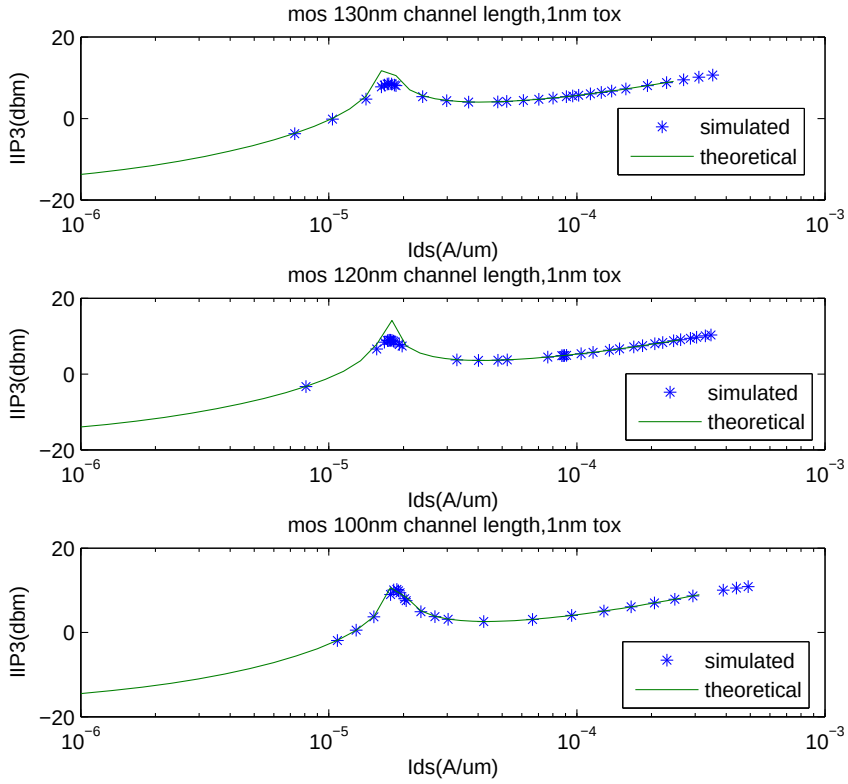

Figure 5.17: Simulated and theoretical values of IIP3 vs drain current I_{ds}.

Chapter 6

POLYGATE DEPLETION EFFECT

In this chapter, polysilicon technology and the effect of polysilicon gate depletion on RF distortion is analyzed using a 100nm MOS device with 1nm oxide thickness, at bias conditions relevant for RF design (i.e., saturation conditions and gate bias near threshold). Linearity dependence on the doping concentration of polysilicon gate is also presented.

6.1 Background of Polysilicon Gate

The use of polysilicon gate is a key advance in modern CMOS technology [17]. Polysilicon gate is used as a mask during the ion implantation so that the source and drain regions are self-aligned with respect to the gate. This self-aligned structure reduces the device size and also eliminates the large overlap capacitances between gate and drain while maintaining a continuous inversion layer between source and drain. Scaling of MOS devices results in depletion of polysilicon gate at higher gate bias. This effect is further analyzed in the following sections.

6.2 Scaling of MOS devices

Scaling of MOS devices requires thinner gate oxide to maintain electrostatic integrity [14]. Oxide scaling results in poly depletion effect which in turn affects RF distortion [15]. Thus, the impact of poly depletion effects on RF distortion

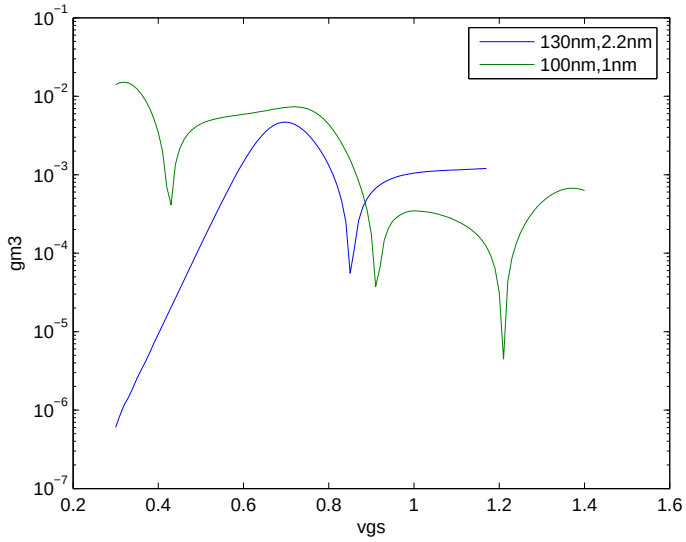

Figure 6.1: g_{m3} vs gate voltage(V_{gs})for two different technologies.

needs to be studied in order to project the gate oxide scaling in MOS RF circuits. Fig. 6.1 shows g_{m3} at $V_{ds} = 1V$ for two different technologies, 130nm gate length device with 2.2nm oxide thickness and 100nm gate length device with 1nm oxide thickness. It can been seen that the linearity can be worse for scaled devices with thin gate oxide.

6.3 Poly Depletion Effect

If the polysilicon gate is not heavily doped, depletion of the n^+ poly gate may occur at higher gate bias. In scaled devices, where t_{ox} is less then 5nm, thickness of this depletion layer cannot be neglected. The depletion layer acts as an extension

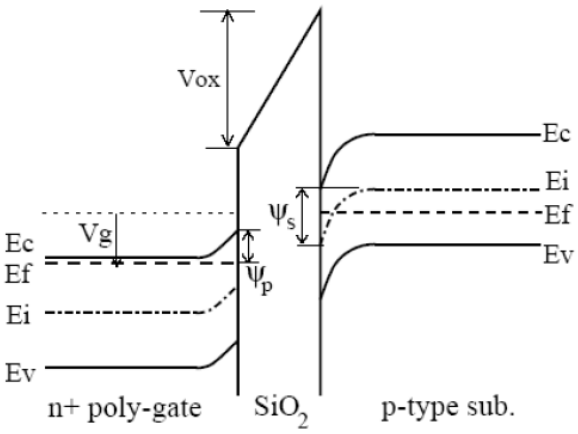

Figure 6.2: Band diagram of n^+ poly gate MOS structure.

of the gate oxide insulator thus, increasing the effective gate oxide thickness and decreasing the effective gate oxide capacitance.

This can be further explained using a band diagram of an n^+ polysilicon gated n-channel MOS structure from [17]. From Fig. 6.2 it can seen that as the gate bias is increased, the oxide field is in the direction of accelerating a negative charge towards the gate. Thus, the bands in the n^+ polysilicon bend slightly upward towards the oxide interface. This depletes the surface of electrons and forms a thin space-charge region in the polysilicon layer which lowers the total gate capacitance.

Poly depletion has a significant impact on RF distortion and can be analyzed using g_{m3} values. Fig. 6.3 shows the simulated g_{m3} curves in the presence of polydoping effects for dopant concentration $N_p = 2.5 \times 10^{19} cm^{-3}$. The solid line represents g_{m3} curve when the poly depletion effect is not taken into account and hence, gate electrode is defined by using a work function. As V_{gs} increases it can

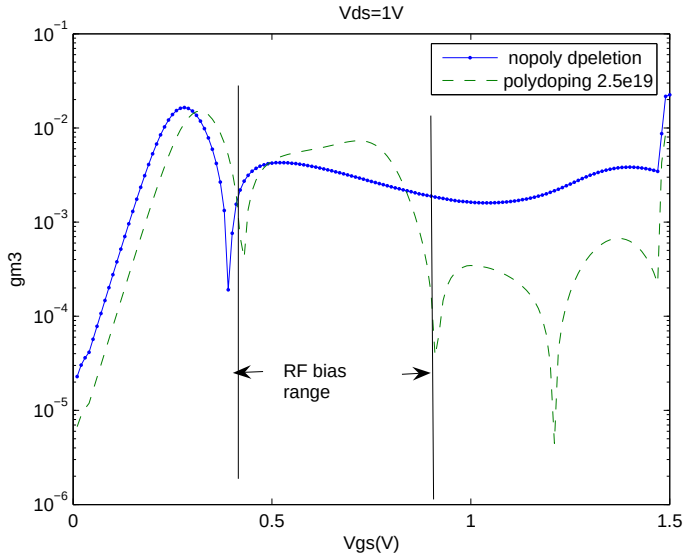

Figure 6.3: g_{m3} vs gate voltage(V_{gs})with and without poly depletion.

be seen that in the RF bias range, g_{m3} increases in the absence of poly depletion. Thus, at higher V_{gs} RF distortion increases due to poly depletion. This depletion effect can be decreased by increasing the dopant concentration in the polysilicon gate which in turn decreases the depletion layer thickness.

Fig. 6.4 shows g_{m3} values for different dopant concentrations of poly-gate for $V_{ds} = 1V$. Thus, as the doping concentration in poly-gate increases, RF distortion is reduced. Further g_{m3} is also plotted for different V_{ds} values. It can be seen from Fig. 6.5 that as V_{ds} increases the device is in saturation for higher V_{gs} values.

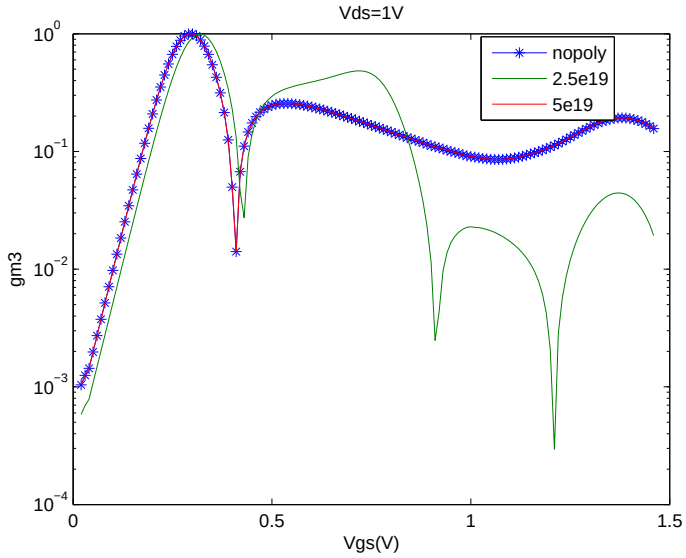

Figure 6.4: g_{m3} vs gate voltage(V_{gs})for two different polydoping concentration.

6.3.1 Effect on Gate Capacitance

Drain current in MOSFET can be expressed as

$$I_{ds} = WQv \tag{6.1}$$

where Q is the channel charge density along the current direction, W is the width of the device, v is the velocity of the carriers. For sufficiently high fields, the carrier velocity approaches saturation value v_{sat}. Hence, equation 6.1 can be written as

$$I_{ds} \simeq WQv_{sat}. \tag{6.2}$$

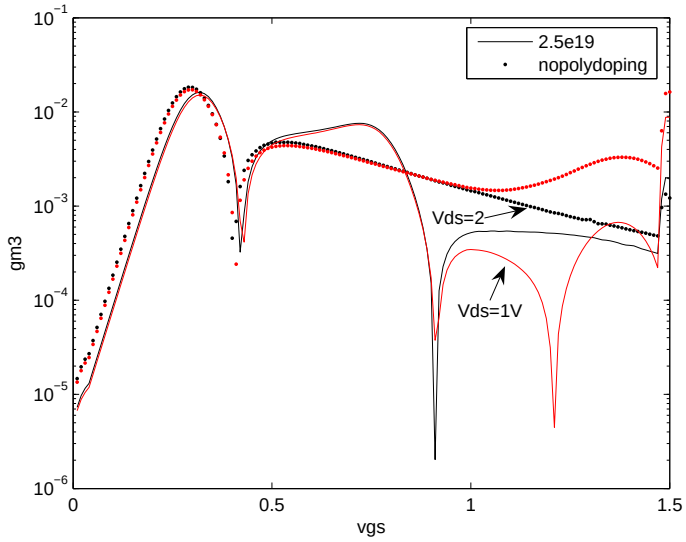

Figure 6.5: g_{m3} versus gate voltage(V_{gs})for two different for V_{ds} values.

First order derivative of drain current with respect to gate bias, g_m is given by [15]

$$g_m \simeq \frac{\partial I_{ds}}{\partial V_{gs}} = WC(V_{gs})v_{sat}. \tag{6.3}$$

An ideal gate capacitance in strong inversion must be equal to the total oxide capacitance and is given by

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}. \tag{6.4}$$

However, the real gate capacitance in strong inversion depends on gate bias when poly effects are taken into effect and can be expressed as [15].

$$C(V_{gs}) = \frac{\epsilon_{ox}}{t_{ox} + t_{dp}(V_{gs})} \tag{6.5}$$

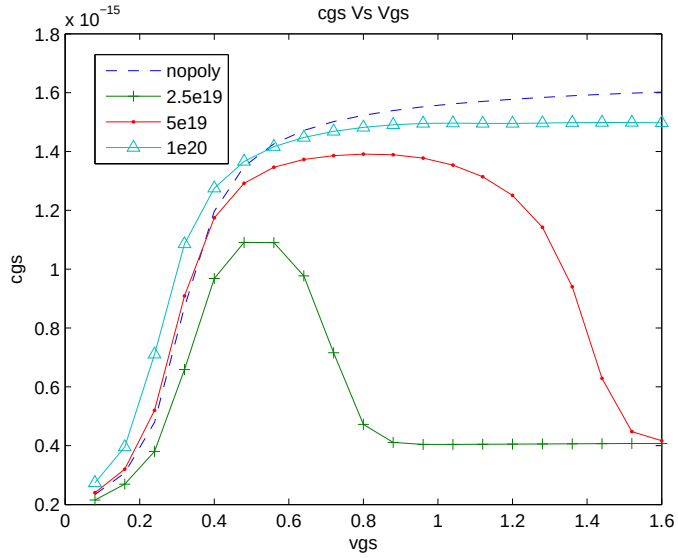

Figure 6.6: Gate capacitance vs gate voltage for different poly gate doping concentration.

where t_{dp} represents the depleted thickness of poly-gate. Thus, decrease in gate capacitance adversely effects g_m.

Fig. 6.6 shows C-V plots at 1MHz frequency. It can be seen that gate capacitance decreases with the decrease in dopant concentration of polysilicon gate and then remains constant at higher V_{gs} values. This is because, as V_{gs} increases, the depletion layer thickness increases and reaches a saturation point beyond which the surface gets inverted. Since there is no p-type semiconductor available to supply holes, inversion holes cannot be generated fast enough to follow the ac signal (assuming there is no Generation/Recombination process). Thus, at higher V_{gs} the gate capacitance remains constant since the depletion layer thickness remains constant.

6.3.2 Effect on IIP3

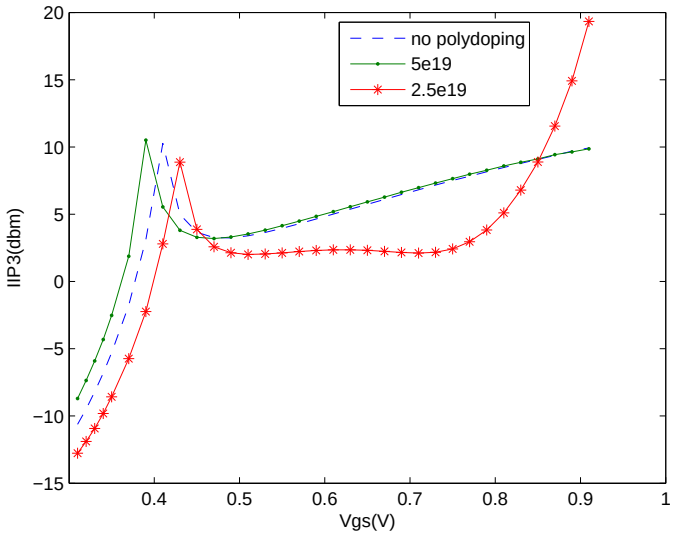

Figure 6.7: IIP3 vs gate voltage for different poly gate doping concentrations.

The effect of poly depletion on linearity is presented in this section. Harmonic Balance simulations are done using Taurus, by changing the dopant concentration in the polysilicon gate. Fig. 6.7 shows variation of IIP3 with the dopant concentration in polysilicon gate, as the gate voltage is increased. It can be seen that as the polydoping is decreased, the IIP3 decreases because of the poly depletion effect in the strong inversion region. Fig. 6.8 shows the IIP3 vs drain current for different doping concentration. It can be seen that linearity decreases with drain current in the strong inversion region.

The simulated values of the IIP3 for different poly-gate concentrations are compared to the theoretical values obtained using the power series and are in good

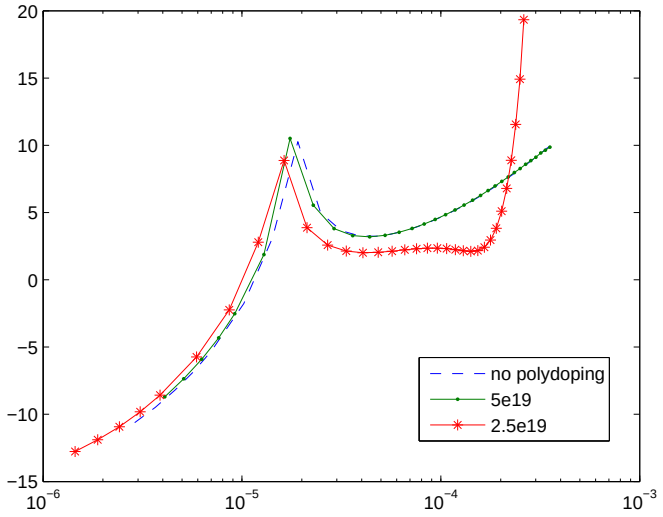

Figure 6.8: IIP3 vs drain current for different poly gate doping concentrations.

agreement as shown in Fig. 6.9. Thus, from the simulation results it can be stated that the RF distortion increases with scaling.

The only way to overcome polysilicon gate depletion effect is to go back to the use of metal gate. As the critical MOS dimensions shrink, conventional polysilicon gates are being replaced by the metal gates. Research is being done to introduce new gate materials which can suppress the poly depletion effect.

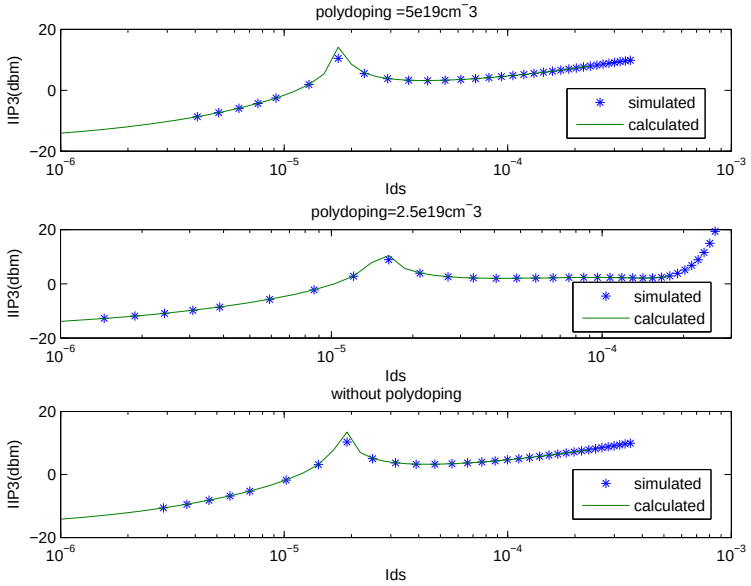

Figure 6.9: Simulated and theoretical IIP3 values for different poly gate doping concentrations.

Chapter 7

HALO DOPING

In this chapter, linearity analysis is extended to 90nm technology. With scaling of CMOS, uniform channel doping has been slowly replaced by super retrograde doping and source/drain halo. Halo design is a critical element in CMOS scaling. Super Halo doping consists of highly nonuniform profile in both vertical and lateral directions. These high doping regions are self-aligned to the gate and source-drain, which help shield the gate controlled depletion region from penetration of the source and drain fields [20].

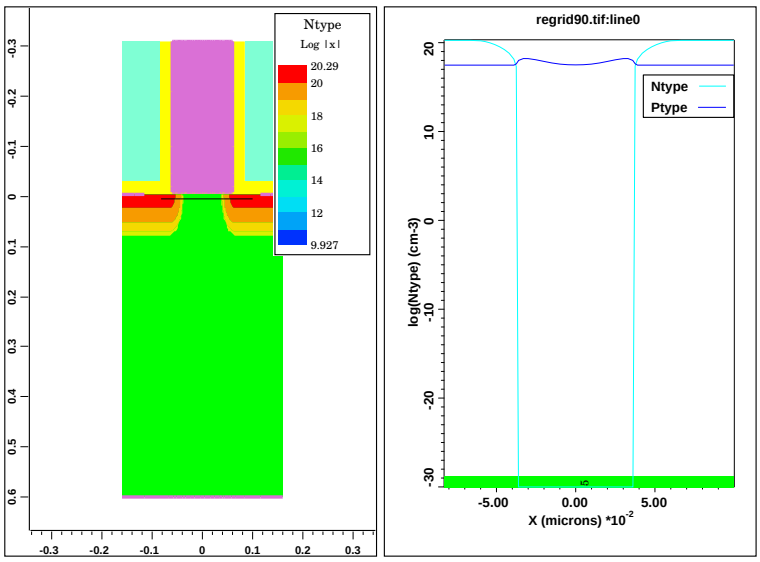

Figure 7.1: Doping Profile of 90nm NMOSFET along the channel.

7.1 MOS 90nm

A hypothetical NMOSFET device with 90nm L_{eff}, 4.5nm oxide thickness and super retrograde channel doping and source/drain halo has been used to extend the linearity characterization to short channel devices. The input code in Medici for this device is obtained from [16]. The detailed description of the device topology is described in [16]. This structure is imported into Taurus and Harmonic Balance analysis is performed.

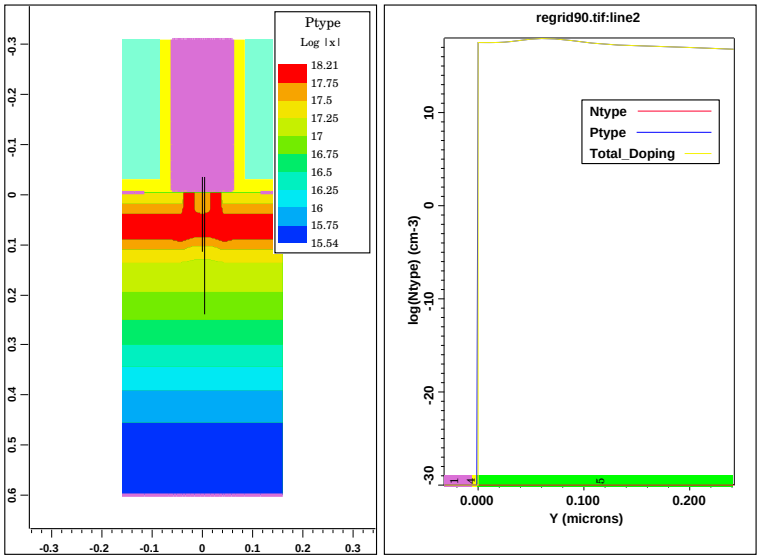

Figure 7.2: Doping profile of 90nm NMOSFET across the channel.

The doping profiles through the channel are shown in Fig. 7.1 and Fig. 7.2.

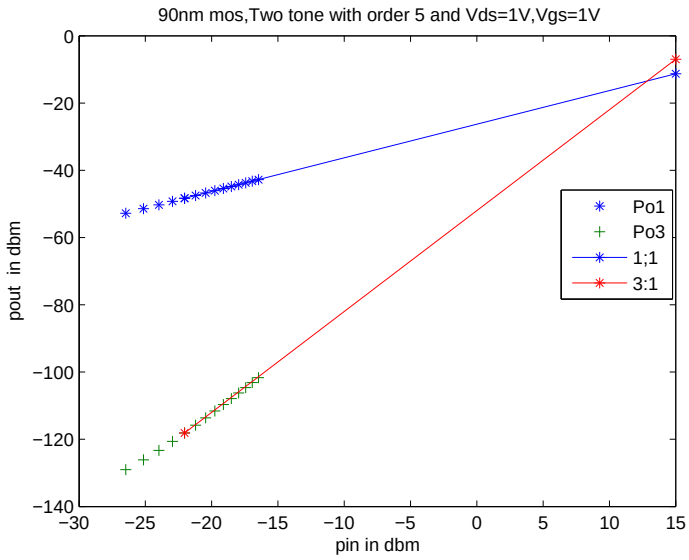

Figure 7.3: First and third order output powers vs input power.

7.2 Harmonic Analysis

Applying HB simulation for the 90nm NMOS device, the impact of halo doping on linearity is analyzed. First and third order output powers are plotted against the input power and the extrapolated IP3 point is as shown in the Fig. 7.3. Fig. 7.4 shows g_m, g_{m2}, g_{m3} vs V_{gs} for $V_{ds} = 1V$. Further, IP3 dependence on V_{ds} and V_{gs} is analyzed using an input drive of 10mV and is plotted as shown in Fig. 7.4. It can be seen that the IIP3 peak as seen for the 100nm MOS device is not very pronounced in this device, which may be attributed to the halo doping. It can also be observed that the IP3 value at threshold point where $gm_3 = 0$ may not be a peak value and its value is higher in strong inversion.

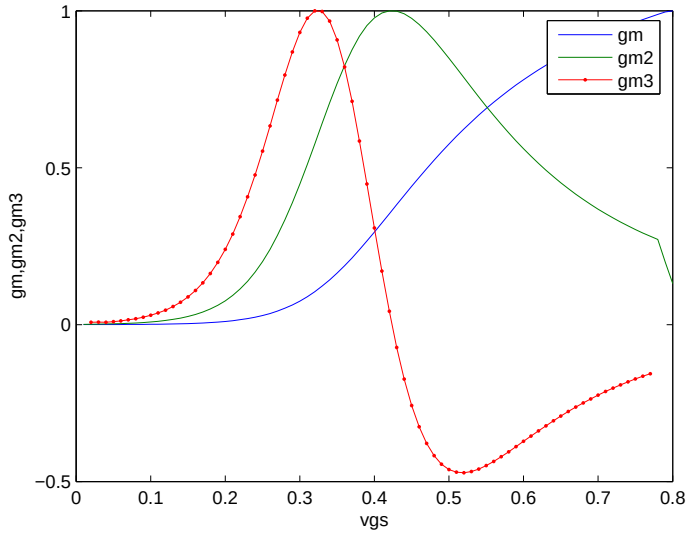

Figure 7.4: g_m, g_{m2}, g_{m3} vs V_{gs}.

7.3 Discrepancy

Unlike previous results, there is a discrepancy between simulated and calculated results when plotted versus V_{gs} and is shown in Fig.7.6. This deviation may be attributed to the presence of halo doping or the way halo doping is handled in the simulator. The effect of halo doping on linearity could not be clearly established and hence requires further research.

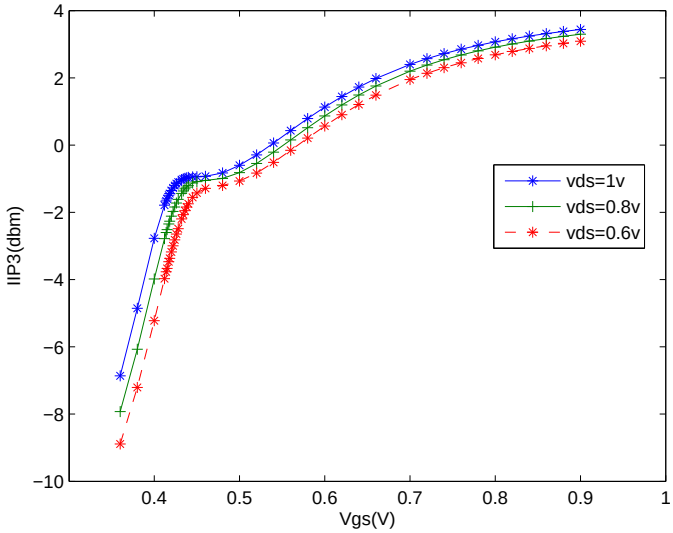

Figure 7.5: IIP3 vs V_{gs} for three different V_{ds} values.

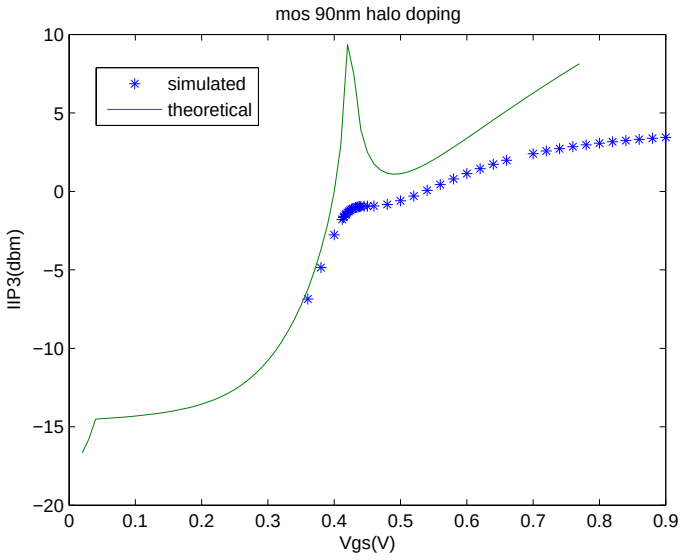

Figure 7.6: Simulated and theoretical values.

Chapter 8

CONCLUSIONS AND FUTURE WORK

In this thesis, Intermodulation Linearity characteristics of CMOS devices is analyzed using hypothetical 130nm, 100nm, 90nm NMOS devices. The effect of technology scaling on linearity is examined by varying the channel length and oxide thickness of the NMOS devices. Polysilicon gate depletion and its impact on linearity is also presented for various doping concentrations of the poly-gate.

HB Method is used to analyze harmonic distortion at semiconductor device level using Taurus simulation tool. The IIP3 values obtained from HB simulation are compared using power series.

Although it is not possible to represent any dynamic system by power series, it can still be used in cases where the system can be represented by several non-interacting subsystems, and where nonlinearities are memoryless [13].

It can be concluded from this work that devices with longer channel and thicker gate oxide can be used to achieve improved linearity. Further, it can also be concluded that RF distortion increases with scaling. In short channel devices, linearity can be improved by increasing the dopant concentration of the poly-gate. Thus, CMOS scaling trades linearity to cutoff frequency and gain.

8.0.1 Future Work

It is observed that the IIP3 does not have a sharp peak at the threshold point for the 90nm MOS device, which may be attributed to the presence of Halo doping. The effects of Halo doping on linearity are yet to be analyzed. The impact of polysilicon depletion on gate capacitance need to be further analyzed.

Bibliography

[1] Guofu Niu, John D. Cressler, *SiGe Heterojunction Bipolar Transistor*, Artech House Publishers Jan. 2003

[2] Stephen Mass, *RF and microwave steady state analysis*,Artech House publishers,2nd edition Feb. 2003.

[3] Boris Troyanovsky "Frequency domain algorithms for simulating large signal distortion in semiconductor devices", *Dissertation* Standford University November 1997.

[4] Behzad Razavi, *RF Microelectronics*, Prentice Hall PTR, 1st edition, Nov.1997

[5] John Rodgers, Calvin Plett, *RF Integrated circuit design*, Artech House Publishers,May 2003

[6] kenneth S. kundert, Jacob K. White, Alberto De Sangiovanni-Vincentelli, *Steady state methods for simulating Analog and Microwave Circuits*, Springer, 1 edition, March 1990

[7] Taurus manual from *Synopsys*

[8] Guofu Niu, Qingqing Liang, John D. Cressler, Charles S.Webster, David L.Harame "RF Linearity Characeristics of SiGe HBTs" *IEEE Transactions On Microwave Theory And Tech.*, Vol.49, No.9, Sep. 2001

[9] Pierre H.Woerlee, Mathijis J. Knitel, Ronald van Langevelde, Dirk B. M. Klaassen, Luuk F. Tiemeijer, Andries J.Scholten, Adrie T.A.Zegers-van Duijnhoven,"RF-CMOS performance trends", *IEEE Transactions on Electron Devices*, Vol.48, No.8, Aug. 2001

[10] Sanghoon Kang, Byounggi Choi, Bumman Kim,"Linearity Analysis of CMOS for RF Application", *IEEE Transactions on Microwave Theory And Tech.*, Vol.51, No.3, March 2003

[11] Guofu Niu, Jun Pan, Xiaoyun Wei, Stewart S.Taylor and David Sheridan "Intermodulation Linearity Characterististics of CMOS transistors in a 0.13um process" *RFIC Conference*, 2005

[12] R.Van Langevelde, L.Tiemeijer, R.Havens, M.Knitel, R.Ores, P.Woerlee,and D.Klaassen,"RF-distortion in deep-submicron CMOS technologies," *IEDM Technical Digest*, p807-810, Dec.2000

[13] Jose Carlos Pedro, Nuno Borges Carvalho, *Intermodulation distortion in Microwave and Wireless Circuits*, Artech House Publishers, Dec.2002

[14] Yannis Tsividis *Operation and Modeling of The Mos Transistor* Oxford University press, 2nd edition, May 2003

[15] Chang-Hoon Choi, Zhiping Yu, Robert W.Dutton "Impact of Poly-Gate Depletion on MOS RF Linearity" *IEEE Electron Device Letters*, Vol.24, No.5, May 2003

[16] http://www-mtl.mit.edu/researchgroups/Well/device1/topology1.html

[17] Yuan Taur, Tak H.Ning *Fundamentals of Modern VLSI Devices*, Cambridge University Press, 1998

[18] Ulrich L.Rohde, David P.Newkirk *RF/Microwave circuit design for wireless applications*, Wiley-Interscience, 1 edition ,March 2000

[19] Y.Taur and E.J.Nowak,"CMOS Devices below 0.1um:How will performance Go?" *IEDM Technical Digest*,p.215(1997)

[20] Y.Taur, C.H.Wann and D.J.Frank,"25nm CMOS Design Considerations" *IEDM Technical Digest*, p.789, 1998

[21] Hang Hu, Jarvis B.Jacobs, Lisa T.Su, Dimitri A.Antoniadis,"A study of Deep-SubMicron MOSFET scaling based on Experiment and simulation", *IEEE Transactions on Electron Devices*, Vol.42,No., April 1995

[22] Bart Van Zeghbroeck, *Principles of Semiconductor Devices*

[23] Boris Troyanovsky et al.,"Efficient Multi-Tone Harmonic balance Simulation of Semiconductor Devices in the presence of Linear High-Q Circuitry" *Proc.SASIMI*, Dec.1997

[24] Boris Troyanovsky et al.,"Large signal analysis of RF/microwave devices with parasitics using harmonic balance device simulation, *Proc.SASIMI*, Nov.1996

[25] Peng Li, Lawrence T.Pileggi,"Efficient Per-Nonlinearity Distortion Analysis for Analog and RF Circuits", *IEEE Transactions on computer-Aided design of intergrated circuits and systems* Vol.22,No.10,Oct.2003

[26] F Yuan and A.Opal,"An efficient transient analysis algorithm for midly nonlinear circuits", *IEEE Transactions on Computer Aided design of Integrated Circuits and Systems*, Vol.21, no.6, pp 662-673, 2002

[27] P. Feldmann, R. C. Melville, and D. Long, Efficient frequency domain analysis of large nonlinear analog circuits, *Proc.IEEE CICC*, pp 461464, May 1996

[28] Medici manual, *Synopsys*

[29] P.L.Heron and M.B Steer,"Jacobian calculation using the multidimensional Fast Fourier Transform in the harmonic balance analysis of nolinear circuits," *IEEE Transactions on Microwave Theory and Techniques*, Vol.38, pp 429-431, April 1990

[30] V. Rizzoli, C. Cecchetti, A. Lipparini and F.Mastru,"General-purpose harmonic balance analysis of nonlinear microwave circuits under multitone excitation," *IEEE Transactions on Microwave Theory and Techniques*, pp 1650-1659, Dec.1988

[31] Paulo J. C. Rodrigues, *Computer-aided analysis of nonlinear microwave circuits* Artech House Inc, 1998

[32] Ulrich L. Rohde, David P. Newkirk,*RF/Microwave Circuit Design for wireless Applications* A Wiley-Interscience Publication

[33] http://www.intel4004.com/sgate.htm

[34] http://en.wikipedia.org/wiki/MOSFET

Appendix A

Input code for 100nm MOS structure in Medici

```
  1... $input code for 100nm mos device
  2... mesh smooth=1
  3... $ X-COORDINATES
  4... $ -------------
  5... x.mesh n=1 l=-0.56
  6... x.mesh n=9 l=-0.06 r=0.9
  7... x.mesh n=20 l=0 r=0.9
  8... x.mesh n=37 l=0.02 r=0.9
  9... x.mesh n=48 l=0.05 r=1.2
 10... x.mesh n=59 l=0.08 r=0.8
 11... x.mesh n=76 l=0.10 r=1.1
 12... x.mesh n=87 l=0.16 r=1.1
 13... x.mesh n=95 l=0.66 r=1.1
 14... $ Y-COORDINATES
 15... $ ------------
 16... y.mesh n=1 l=-0.0110
 17... y.mesh n=2 l=-0.0010 r=1
 18... y.mesh n=4 l=0.0000 r=1
 19... y.mesh n=5 l=0.0002 r=1
```

```
20... y.mesh n=8 l=0.0015 r=1
21... y.mesh n=9 l=0.0025 r=1
22... y.mesh n=10 l=0.0035 r=1
23... y.mesh n=11 l=0.0050 r=1
24... y.mesh n=12 l=0.0070 r=1
25... y.mesh n=13 l=0.0090 r=1
26... y.mesh n=14 l=0.0120 r=1
27... y.mesh n=15 l=0.0150 r=1
28... y.mesh n=16 l=0.0160 r=1
29... y.mesh n=17 l=0.0270 r=1
30... y.mesh n=18 l=0.0300 r=1
31... y.mesh n=21 l=0.0500 r=1
32... y.mesh n=23 l=0.1000 r=1
33... y.mesh n=26 l=0.1500 r=1
34... y.mesh n=29 l=0.1800 r=1
35... y.mesh n=30 l=0.2200 r=1
36... y.mesh n=31 l=0.2600 r=1
37... y.mesh n=32 l=0.3500 r=1
38... y.mesh n=33 l=1.5000
39... $ REGION AND ELECTROD
40... $ -------------------
41... region num=1 silicon ix.min=1 ix.max=95 iy.min=4 iy.max=33
```

```
42... region num=2 oxide  ix.min=1 ix.max=95 iy.min=1 iy.max=4
43... region name="GateDi" oxide ix.min=9 ix.max=87 iy.min=1 iy.max=4
44... region num=5 oxide ix.min=1 ix.max=20 iy.min=1 iy.max=2
45... region num=4  oxide ix.min=76 ix.max=95 iy.min=1 iy.max=2
46... $polysilicon gate
47... region name="npoly" silicon ix.min=20 ix.max=76 iy.min=1 iy.max=2
48... electrod name=Gate ix.min=20 ix.max=76 iy.min=1 iy.max=2
49... electrod name=Substr ix.min=1 ix.max=95 iy.min=33 iy.max=33
50... electrod name=Source ix.min=1 ix.max=9 iy.min=4 iy.max=4
51... electrod name=Drain ix.min=87 ix.max=95 iy.min=4 iy.max=4
52... $ DOPING
53... $ ------
54... $poly doping
55... $profile n-type unif n.peak=1e20 region=npoly out.file=profile.dop
56... profile unif conc=1e16 p.type x.right=10 x.left=-10 y.t=-10 y.b=10
57... profile p.type conc=4e18 x.right=10 x.left=-10 char=0.019 y.min=0.0130
58... profile p.type conc=6e17 x.right=10 x.left=-10 char=0.050 y.min=0.0550
59... $D/S:
60... profile conc=1e20 n.type x.left=-10 x.right=-0.01 junc=0.03  erfc.lat
  ... + x.char=0.0103
61... profile conc=1e20 n.type x.left=-10 x.right=-0.06 junc=0.10  erfc.lat
  ... + x.char=0.03
```

```
62... profile conc=1e20 n.type x.left=0.11 x.right=10 junc=0.03  erfc.lat
  ... + x.char=0.0103
63... profile conc=1e20 n.type x.left=0.16 x.right=10 junc=0.10  erfc.lat
  ... + x.char=0.03
64... $ SAVE MESH
65... $ ---------
66... save mesh out.file=nopoly.tif tif
67... end.
```

This structure is imported into Taurus device and the following code describes DC, AC and HB analysis in Taurus device.

```
 1: taurus{device}
 2:
 3: DefineDevice (name=mos,meshfile=regrid.tif)
 4:#models
 5: physics
 6: (
 7: Silicon
 8:   (
 9:   Electroncontinuity
10:   (
```

```
11:       Mobility
12:       (
13:           Lowfieldmobility
14:           (
15:           SurfModelActive=True
16:           ,SurfModel=LombardiSurfaceModel
17:           )
18            highfieldmobility
19              (
20              highefieldmodel=caugheythomas
21              )
22:       ),
23:
24:   )
25:   HoleContinuity
26:   (
27:       Mobility
28:       (
29:           Lowfieldmobility
30:           (
31:           SurfModelActive=True
32:           ,SurfModel=LombardiSurfaceModel
```

```
33:           )
34:       ),
35:
36:   )
37:    )
38: )
39    Contact definitions
40: contact(name=gate,workfunction=4.17)
41: setbias(value=-0.5){contact(name=Gate, type=contactvoltage){mos}}
42: setbias(value=0.0){contact(name=back, type=contactvoltage){mos}}
43: setbias(value=0.0){contact(name=Source, type=contactvoltage){mos}}
44: setbias(value=0.0){contact(name=Drain, type=contactvoltage){mos}}
45:
46:
47:
48:
49: symbolic(carriers=0,newton,direct)
50: solve{}
51: symbolic(newton, carriers=2, direct)

    #-------------DC Bias-------------------------------
52: solve{
```

```
53:    ramp(logfile=mosfet100nm_dc.data,
54:    RampSpecification(endValue=1.0, nsteps=20)
55:        {BiasObject(name=drain,type=contactVoltage)}
56:    )
57:
58: }

59:#----------   AC analysis-----------------------------
60
61: solve
62: {
63: acanalysis
64: (
65: logfile=ac_mos100nm.data,acxfile=mosfet100nm_ac,frequency=2e9,
66: terminal(gate)
67: rampspecification(endvalue=1.2,nsteps=20)
68: {biasobject(name=gate,type=contactvoltage)}
69: extract(cutofffrequency(basecontact=gate,collectorcontact=drain))
70: )
71: }
72: save(meshfile=mos100nm_Ac.tdf)
73:
```

```
74: # --------------  harmonic balance--------------------------------------

75: # one tone HB simulation
     frequency= 5.8GHZ frequency,
     truncation order=5 and is specified by nharm,
     periodic source specified using hb_src has 10mV amplitude
76: harmonicbalance(
77: hb_numerics(iterations=20 maxiiter=100 maxbackvector=10)
78: tone1=5.8e9 nharm1=5 hb_src(name=gate m1=0.010)
79: logfile=mos100nm_5o_hb_010.data
80: display_solution=false
81: )

82: #two tone test used to find IIP3,
      f1=5.8GHz and f2=5.9GHz are the two frequencies applied
      Input voltage is sweeped from 20mV to 40mv

83: harmonicbalance(
84: tone1=5.8e9 nharm1=5
85: tone2=5.9e9 nharm2=3
86: hb_src(name=gate m1=0.020 m2=0.020)
87: logfile=mos100nm_5o_hb_020_020.data
```

```
88: display_solution=false
89: hb_continue=false
90: )
91:
92: harmonicbalance(
93: tone1=5.8e9 nharm1=5
94: tone2=5.9e9 nharm2=3
95: hb_src(name=gate m1=0.030 m2=0.030)
96: logfile=mos100_5o_hb_030_030.data
97: display_solution=false
98: hb_continue=false
99: )
100:
101: harmonicbalance(
102: tone1=5.8e9 nharm1=5
103: tone2=5.9e9 nharm2=3
104: hb_src(name=gate m1=0.040 m2=0.040)
105: logfile=mos100nm_5o_hb_040_040.data
106: display_solution=false
107: hb_continue=false
108: )
109:
```

```
110: save(meshfile=mos100_hb.tdf)
```

```
# The logfiles give the first and third order terminal currents
and hence first and third order powers can be calculated.
```

```
#Input code for plotting IIP3 dependence on Vgs and Vds,
#At a fixed drain voltage(1V), HB analysis is applied at
different gate voltages.

  1: taurus{device}
  2:
  3: DefineDevice (name=mos,meshfile=nopoly.tif)
  4:
  5:
  6: physics
  7: (
  8: Silicon
```

```
 9:   (
10:   Global
11:       (
12:       fermistatisticsActive=true
13:       )
14:   Electroncontinuity
15:   (
16:       Mobility
17:       (
18:           Lowfieldmobility
19:           (
20:           SurfModelActive=True
21:           ,SurfModel=LombardiSurfaceModel
22:           )
23:           Highfieldmobility(
24:           highfieldmodel=caugheythomasmodel
25:           )
26:       ),
27:   )
28:   HoleContinuity
29:   (
30:       Mobility
```

```
31:       (
32:           Lowfieldmobility
33:           (
34:           SurfModelActive=True
35:           ,SurfModel=LombardiSurfaceModel
36:           )
37:           Highfieldmobility(
38:                           highfieldmodel=caugheythomasmodel
39:                           )
40:
41:       ),
42:   )
43:    )
44: )
45: contact(name=gate,workfunction=4.17)
46: setbias(value=0.0){contact(name=Gate, type=contactvoltage){mos}}
47: setbias(value=0.0){contact(name=substr, type=contactvoltage){mos}}
48: setbias(value=0.0){contact(name=Source, type=contactvoltage){mos}}
49: setbias(value=0.0){contact(name=Drain, type=contactvoltage){mos}}
50:

51: symbolic(carriers=0,newton,direct)
```

```
52: solve{}
53: symbolic(newton, carriers=1,electron, direct)
  # drain voltage is ramped to 1V
54: solve{
55:    ramp(logfile=poly_vds_1p0.data,
56:
57:    RampSpecification(endValue=1.0, nsteps=20)
58:        {BiasObject(name=drain,type=contactVoltage)}
59:    )
60:
61: }
62:
63: solve{
64:    ramp(logfile=poly_vds_1p0_vgs_p40.data,
65:    RampSpecification(endValue=0.40 nsteps=20)
66:        {BiasObject(name=gate,type=contactVoltage)}
67:    )
68:
69: }
70:#harmonic balance analysis at gate voltage=0.4v
   #f1=5.8GHz,f2=5.9GHz are the two fundamental frequencies with
   1MHZ tone spacing, trucation order is 5, input voltage=10mV
```

```
 71: harmonicbalance(
 72: hb_numerics(iterations=20 maxiiter=100 maxbackvector=10)
 73: tone1=5.8e9 nharm1=5
 74: tone2=5.9e9 nharm2=3
 75: hb_src(name=gate m1=0.010 m2=0.010)
 76: logfile=hb_vds_1p0_vgs_p40_010_010_1carrier.data
 77: display_solution=false
 78: hb_continue=false
 79: )
 80:
81:
82: solve{
83:    ramp(logfile=poly_vds_1p0_vgs_p45.data,
84:    RampSpecification(endValue=0.45, nsteps=20)
85:        {BiasObject(name=gate,type=contactVoltage)}
86:    )
87:
88: }
89:#HB at Vgs=0.45V
90: harmonicbalance(
91: tone1=5.8e9 nharm1=5
92: tone2=5.9e9 nharm2=3
```

```
93: hb_src(name=gate m1=0.010 m2=0.010)
94: logfile=hb_vds_1p0_vgs_p45_010_010_1carrier.data
95: display_solution=false
96: hb_continue=false
97: )


98: solve{
99:    ramp(logfile=poly_vds_1p0_vgs_p5.data,
100:    RampSpecification(endValue=0.5, nsteps=20)
101:        {BiasObject(name=gate,type=contactVoltage)}
102:    )
103:
104: }
105:
106:#HB at vgs=0.5V
107: harmonicbalance(
108: tone1=5.8e9 nharm1=5
109: tone2=5.9e9 nharm2=3
110: hb_src(name=gate m1=0.010 m2=0.010)
111: logfile=hb_vds_1p0_vgs_p5_010_010_1carrier.data
112: display_solution=false
113: hb_continue=false
```

```
114: )
115:


116: solve{
117:    ramp(logfile=poly_vds_1p0_vgs_p55.data,
118:    RampSpecification(endValue=0.55, nsteps=20)
119:        {BiasObject(name=gate,type=contactVoltage)}
120:    )
121:
122: }
123:
124:#HB at Vgs=0.55V
125: harmonicbalance(
127: tone1=5.8e9 nharm1=5
128: tone2=5.9e9 nharm2=3
129: hb_src(name=gate m1=0.010 m2=0.010)
130: logfile=hb_vds_1p0_vgs_p55_010_010_1carrier.data
131: display_solution=false
132: hb_continue=false
133: )
134: solve{
135:    ramp(logfile=poly_vds_1p0_vgs_p6.data,
```

```
136:    RampSpecification(endValue=0.6, nsteps=20)
137:        {BiasObject(name=gate,type=contactVoltage)}
138:    )
139:
140: }
141:
142:#HB at Vgs=0.6V
143: harmonicbalance(
144: tone1=5.8e9 nharm1=5
145: tone2=5.9e9 nharm2=3
146: hb_src(name=gate m1=0.010 m2=0.010)
147: logfile=hb_vds_1p0_vgs_p6_010_010_1carrier.data
148: display_solution=false
149: hb_continue=false
150: )
151:
152:
153: solve{
154:    ramp(logfile=poly_vds_1p0_vgs_p65.data,
155:    RampSpecification(endValue=0.65, nsteps=20)
156:        {BiasObject(name=gate,type=contactVoltage)}
157:    )
```

```
158:
159: }
160:
161:#HB at Vgs=0.65V
162: harmonicbalance(
163: tone1=5.8e9 nharm1=5
164: tone2=5.9e9 nharm2=3
165: hb_src(name=gate m1=0.010 m2=0.010)
166: logfile=hb_vds_1p0_vgs_p65_010_010_1carrier.data
167: display_solution=false
168: hb_continue=false
169: )
170:
171:
172: solve{
173:    ramp(logfile=poly_vds_1p0_vgs_p7.data,
174:    RampSpecification(endValue=0.7, nsteps=20)
175:        {BiasObject(name=gate,type=contactVoltage)}
176:    )
177:
178: }
179:
```

```
200:#HB at Vgs=0.7V
201: harmonicbalance(
202: tone1=5.8e9 nharm1=5
203: tone2=5.9e9 nharm2=3
204: hb_src(name=gate m1=0.010 m2=0.010)
205: logfile=hb_vds_1p0_vgs_p7_010_010_1carrier.data
206: display_solution=false
207: hb_continue=false
208: )
209:
210: solve{
211:    ramp(logfile=poly_vds_1p0_vgs_p75.data,
212:    RampSpecification(endValue=0.75, nsteps=20)
213:        {BiasObject(name=gate,type=contactVoltage)}
214:    )
215:
216: }
217:
218:#HB at vgs=0.75V
219: harmonicbalance(
220: tone1=5.8e9 nharm1=5
221: tone2=5.9e9 nharm2=3
```

```
222: hb_src(name=gate m1=0.010 m2=0.010)
223: logfile=hb_vds_1p0_vgs_p75_010_010_1carrier.data
224: display_solution=false
225: hb_continue=false
226: )
227:save(meshfile=hb.tdf)

   #first and third order terminal currents can be obtained from the
   logfiles and hence IIP3 value for each bias can be calculated.
```