Development of a Universal MOSFET Gate Impedance Model

Sripriya Das

Follow this and additional works at: http://scholarworks.rit.edu/theses

Recommended Citation

This Dissertation is brought to you for free and open access by the Thesis/Dissertation Collections at RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.
DEVELOPMENT OF A UNIVERSAL MOSFET GATE IMPEDANCE MODEL

by

SRIPRIYA DAS

A DISSERTATION

Submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy
in
Microsystems Engineering
at the
Rochester Institute of Technology

December 2006

Author: SriPriya Das

Certified by: P.R. Mukund

Certified by: Harvey J. Palmer

Microsystems Engineering Program

P.R. Mukund
Professor of Electrical Engineering

Harvey J. Palmer
Dean Kate Gleason College of Engineering

Approved by: Mustafa A.G. Abushagur

Mustafa A.G. Abushagur
Director of Microsystems Engineering Program
NOTICE OF COPYRIGHT

© 2006 All rights reserved

Sripriya Das

REPRODUCTION PERMISSION STATEMENT

Permission Granted

TITLE:

"Development of a universal MOSFET gate impedance model"

I, Sripriya Das, hereby grant permission to the Wallace Library of the Rochester Institute of Technology to reproduce my dissertation in whole or in part. Any reproduction will not be for commercial use or profit.

Signature of Author: Sripriya Das

Date:__________________________
RIT DIGITAL MEDIA LIBRARY NON-EXCLUSIVE DISTRIBUTION LICENSE

In order for the RIT DML to reproduce, translate and distribute your submission worldwide, your agreement to the following terms is necessary. For works with multiple authors, the submitting author bears responsibility for complying with this license. Please take a moment to read the terms of this license, fill in the information requested, and sign and submit this license to the community administrator.

By signing and submitting this license, you (the author(s) or copyright owner) grant to Rochester Institute of Technology (RIT) the non-exclusive and not for profit right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

- You agree that RIT may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

- You also agree that RIT may keep more than one copy of this submission for purposes of security, back-up and preservation.

- You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright.

- If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant RIT the rights required by this license, and that such third-party owned material is clearly identified and acknowledged within the text or content of the submission.

- IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN RIT, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT.

RIT will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

Title of item submitted to the RIT DML

_Sripriya Das_
Signature ___________________________ Date ___________________________

_Sripriya Das_
Print name
DEVELOPMENT OF A UNIVERSAL MOSFET GATE IMPEDANCE MODEL

By

Sripriya Das

Submitted by Sripriya Das in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Microsystems Engineering and accepted on behalf of the Rochester Institute of Technology by the dissertation committee.

We, the undersigned members of the Faculty of the Rochester Institute of Technology, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Microsystems Engineering.

Approved by:

Dr. P.R. Mukund
(Committee Chair and Dissertation Advisor)

Dr. J.E. Moon

Dr. R. Pearson

Dr. S. Rommel

Dr. L.G. Shantharama
(External member, Eastman Kodak Co.)

MICROSYSTEMS ENGINEERING PROGRAM
ROCHESTER INSTITUTE OF TECHNOLOGY
December 2006
Scaling of CMOS technology to 100 nm & below and the endless pursuit of higher operating frequencies drive the need to accurately model effects that dominate at those feature sizes and frequencies. Current modeling techniques are frequency limited and require different models for different frequency ranges in order to achieve accuracy goals. In the foundry world, high frequency models are typically empirical in nature and significantly lag their low frequency counterparts in terms of availability. This tends to slow the adoption of new foundry technologies for high performance applications such as extremely high data rate serializer/deserializer transceiver cores. However, design cycle time and time to market while transitioning between technology nodes can be reduced by incorporating a reusable, industry-standard model. This work proposes such a model for device gate impedance that is simulator-friendly, compact, frequency-independent, and relatively portable across technology nodes. This semi-empirical gate impedance model is based on depletion in the poly-silicon gate electrode. The effect of device length and single-leg width on the input impedance is studied with the aid of extensive measured data obtained from devices built in 110 nm and 180 nm technologies in the 1-20GHz frequency range. The measured data illustrates that the device input impedance has a non-linear frequency dependency. This variation in input impedance is the result of gate poly-silicon depletion, which can be modeled by an external RC network connected at the gate of the device. Excellent agreement between the simulation results and the measured data validates the model in the device active region for 1-20GHz frequency range. The gate impedance model is further modified by incorporating parasitic effects, extending its range to 200MHz-20GHz. This model performs accurately for 180 nm, 110 nm and 90 nm technologies at different bias conditions and dimensions. The model and model parameter behavior are consistent across technology nodes thereby enabling re-usability and portability. The accuracy of this new gate impedance model is demonstrated in various applications: to validate the model extraction techniques for different device configurations, to assess the input data run-length variations on CML buffer performance and to estimate the jitter in ring oscillators.
To my parents.....
ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr.P.R.Mukund. I am grateful to him for believing in me and giving me the opportunity to work with him. I could not have asked for a better advisor and without his supervision and guidance, I would have never finished this work. I learnt a lot from him, not just about technical aspects but about life too. I am glad that our paths crossed.

I would like to thank my committee members, Dr. James E.Moon, Dr. Robert Pearson, Dr. Sean Rommel and Dr. L.G.Shantharama, for taking time out, at a short notice and for their valuable inputs.

I would like to extend my gratitude to Clyde Washburn, with whom I worked closely on this project from day one. Thank you, Clyde for spending long hours and sharing your knowledge.

This work would not have been possible without the support of LSI Logic Corp., especially the Mixed-signal modeling group. I would like to thank Jeff for his time and patience during the course of the project and Jan Kolnik for his positive feedback. Other members of the group, Ken Paradis and Steve Howard – Thank you for all the help.

My colleagues, past and present – Anand, Tejasvi, Simo, Sharmila, Mark, Eric and Chris, thank you for making the lab a fun place to work.

Ajay, my coffee partner and a friend I could always count on – thank you for always being there and for the innumerable rides.

Thanks to Ajish for his stand-up comedy shows during the breaks from work.

I would like to thank my family – Latha and Anil, I would have never been able to come this far without your help and support. Srinu and Linda, for believing in me and in everything I did so far.

My deepest gratitude to Vanditha aunty and Anu, my foster kid 😊

Thanks to my husband – for lots of things.

There are others, apart from the ones listed above who helped me along the way. This section cannot be completed without acknowledging them – Thank you.
# Table of Contents

**List of Figures** .................................................. 3

**List of Tables** .................................................. 8

**Chapter 1: Introduction** ......................................... 9

**Chapter 2: Motivation and Background** .......................... 13

2.1. RF MOS model evolution ........................................ 13

2.2. Limitations of the present-day modeling techniques .......... 18

2.2.1. Sub-circuit parameter extraction ......................... 18

2.2.2. Other limitations ........................................... 20

2.3. Future of MOS models—need for poly depletion model ........ 21

**Chapter 3: MOS Gate Characterization** .......................... 23

3.1. Gate electrode modeling ........................................ 23

3.1.1. Low-medium frequency gate model ....................... 23

3.1.2. Gate model including scaling effects .................... 25

3.1.3. Gate model for RF ......................................... 28

3.1.4. Gate models for high frequency applications .......... 33

3.2. Effects of scaling and frequency on gate performance ........ 35

3.2.1. Variation of input impedance with width and length .... 35

3.2.2. Variation with frequency ................................... 37

3.2.3. Variation with gate bias ................................... 38

3.3. Poly-depletion based gate model ................................ 39

3.3.1. Poly depletion effect ...................................... 39

3.3.2. PDE impact on device performance ....................... 41

3.3.3. Analytical and numerical analysis of PDE ............... 43

3.4. Gate model with PDE ........................................... 46

3.4.1. Model description .......................................... 46

3.4.2. Calculation of Poly depletion capacitance .............. 50

**Chapter 4: Experimental Results** ................................. 58

4.1. Measurement and de-embedding ................................ 58

4.2. Poly depletion gate model ..................................... 60

4.3. Comparison of the PD model with other models ............. 63

4.3.1. BSIM3v3 model .............................................. 64

4.3.2. PD model with BSIM4.2.1 .................................. 65
4.4. Universal MOSFET gate model for wide frequency range 68
   4.4.1. Need for an unified wide-frequency gate model 68
   4.4.2. Wide frequency behavior of PD gate model 69
   4.4.3. Universal gate impedance model 70
   4.4.4. Effect of $C_{HF}$ on the input impedance 71
   4.4.5. Experimental Results: Wide frequency gate model 72
   4.4.6. PD gate model with BSIM3v3 81

CHAPTER 5: APPLICATIONS 83
5.1. Gate input impedance model for different device configurations 83
   5.1.1. CS and CD test structures 84
   5.1.2. BSIM3v3 Model 86
   5.1.3. BSIM4 model 90
5.2. Effect of input impedance on circuit performance 93
   5.2.1. Current mode logic: Brief introduction 93
   5.2.2. CML buffer 95
   5.2.3. Data run-length error 95
   5.2.4. CML buffer simulation results 96
5.3. Ring oscillator application 99
   5.3.1. Jitter in ring oscillators 99
   5.3.2. Ring Oscillator simulation results 101

CHAPTER 6: CONCLUSIONS 104

REFERENCES 109
LIST OF FIGURES

Figure 1 Design constraints for NG SerDes .......................................................... 10

Figure 2 MOS model development cycle.............................................................. 14

Figure 3 Progression of MOS models with time .................................................... 16

Figure 4 Increase in model complexity based on frequency of operation ............... 17

Figure 5 Simple gate electrode model comprising of oxide and channel capacitances .. 24

Figure 6 a) Quasi-static representation of the gate electrode b) Overlap and fringe capacitances associated with extrinsic parameters [7] ..................................................... 25

Figure 7 Distributed gate resistance in the gate electrode as a result of capacitive coupling between gate and the channel .......................................................... 28

Figure 8 a) Electrical equivalent of distributed gate resistance and  b) Lumped approximation of the gate resistance ........................................................................ 28

Figure 9 BSIM3v3 RF model [12] ........................................................................ 29

Figure 10 NQS model with R_{elid} and R_{ch} as given in [14] .................................... 30

Figure 11 Gate resistance model implemented in BSIM4.2.1 [17] ........................... 31

Figure 12 a) Gate leakage currents considered in BSIM4; b) Electrical equivalent of the gate leakage currents ................................................................. 33

Figure 13 High frequency input impedance model [23] ........................................ 34

Figure 14 Empirical gate impedance-high frequency model [23] ............................ 35

Figure 15 a) Real part of Z_m for 0.11um process for different single-finger W and with NF = 16; b) NF=8; c) for NF=16 in 180 nm process; and d) for NF=16 in 90 nm process .................................................................................................................. 36

Figure 16 a) Input impedance varying L for 110 nm; b) for 180 nm; and c) for 90 nm .. 37

Figure 17 Input impedance of a device with single-leg dimension of 20 μm/0.11 μm .... 38

Figure 18 Input impedance of a device with single-leg dimension 5 μm/0.11 μm ........ 39
Figure 19 Input impedance Vs. $V_{gs}$ for devices in a) 180 nm; b) 110 nm; and c) 90 nm technologies ................................................................. 40

Figure 20 Schematic representation of poly-depletion based gate model ............... 47

Figure 21 $R_{\text{dep}}$ distribution along the width and length of the gate electrode ........ 47

Figure 22 Carrier concentration profile in the gate electrode under poly-depletion..... 48

Figure 23 Distribution of $R_{\text{dep}}$ along the depletion thickness due to non-uniform variation of the carrier concentration .................................................. 49

Figure 24 Band bending in the gate and substrate regions for a device in inversion mode .................................................................................................................. 51

Figure 25 Process flow to consistently solve the Schrödinger and Poisson’s equations. 52

Figure 26 Variation of the poly-depletion width as a function of gate bias............... 54

Figure 27 Conduction band output from Schred with enhanced view of the depletion width $t_{\text{pd}}$ ............................................................................................................ 54

Figure 28 Calculated depletion widths for various gate bias conditions in a) 110 nm and b) 90 nm technologies ........................................................................... 55

Figure 29 Calculated poly-depletion capacitance as a function of $V_{gs}$ for a device in 110 nm process ........................................................................................................ 55

Figure 30 Real part of input impedance for various bias conditions in 1-20 GHz frequency range ........................................................................................................ 56

Figure 31 Typical layout of a test structure for a device in 90 nm process ............... 58

Figure 32 Chip micrograph of a test structure fabricated in 110 nm technology ....... 59

Figure 33 Standard de-embedding technique using short, open, and dummy test structures .................................................................................................................. 59

Figure 34 Schematic setup for converting S-parameter data ..................................... 60

Figure 35 BSIM4 gate resistance model with external gate impedance model ............ 60

Figure 36 Setup of the PD model with BSIM4 used in the simulations .................. 61

Figure 37 Measured and simulated curves for various bias conditions ................. 62
Figure 38 Simulated and measured input impedance for different widths at $V_{gs} = V_{th} + 0.2$.

Figure 39 Simulated and measured input impedance for different device lengths.

Figure 40 Simulation setup of PD model with BSIM3v3.

Figure 41 Comparison of BSIM3v3 with and without gate impedance models.

Figure 42 Test setup to verify the PD model with BSIM4.2.1.

Figure 43 Performance of BSIM4.2.1 with and without gate PD model.

Figure 44 Simulation and measurement results for different $V_{gs}$ in a) 180 nm and b) 110 nm technologies.

Figure 45 Comparison of BSIM3v3 and BSIM4.2.1 model with the external gate PD model.

Figure 46 Wide frequency behavior of the PD based gate model.

Figure 47 Simulation setup of the wide frequency range gate model.

Figure 48 Comparison of $C_{HF}$ with $C_{dep,HF}$ used in for a device in 110 nm process.

Figure 49 Variations in the a) real and b) imaginary parts of the input impedance due to $C_{HF}$.

Figure 50 a) Gate model performance for device width $W=20 \mu m$; b) for $W=5 \mu m$; and c) for $W=2.5 \mu m$ in 110 nm technology.

Figure 51 a) Gate impedance model parameters $C_{dep,HF}$; b) $R_{dep}$; and c) $C_{HF}$ variation as a function of width for 110 nm technology.

Figure 52 Measured and simulated results at different gate bias for 200 M – 20 GHz frequency range in 110 nm technology.

Figure 53 a) Gate impedance model parameters $C_{dep,HF}$; b) $R_{dep}$; and c) $C_{HF}$ Vs. gate bias voltage for 110 nm technology.

Figure 54 a) Gate impedance model performance for device width $W=5 \mu m$; b) for $W=2.5 \mu m$; and...

Figure 55 a) Gate impedance model parameters $C_{dep}$; b) $R_{dep}$; and c) $C_{HF}$ Vs. device width for 90 nm technology.
Figure 56 Improved gate impedance model for BSIM3v3 ........................................ 81

Figure 57 a) Modified gate impedance model performance for device width W=22.5 μm; b) for W=5 μm; and c) for W=2.5 μm in 180 nm technology ........................................ 82

Figure 58 Test structures for a) CS and b) CD configurations [41].................................. 85

Figure 59 Measured Input Impedance for CS and CD Configurations [41]................. 86

Figure 60 BSIM3v3 with the external gate impedance model for CS [41]..................... 87

Figure 61 CS Measured and simulated input impedance for BSIM3v3 [41]................. 87

Figure 62 BSIM3v3 with an additional capacitance for CD [41]................................. 88

Figure 63 CD Measured and simulated input impedance with BSIM3v3 [41].......... 89

Figure 64 CS Input impedance for different NF [41]................................................. 90

Figure 65 CD input impedance for different NF [41]................................................. 90

Figure 66 CS Simulated and measured input impedance for BSIM4 [41]................. 91

Figure 67 CD Simulated and measured input impedance for BSIM4 [41]................. 92

Figure 68 Basic CML buffer....................................................................................... 95

Figure 69 Frequency variation with run-length.......................................................... 96

Figure 70 Buffer outputs using various models at a) 16 GHz; b) 4 GHz; and c) 500 MHz ............................................................................................................. 98

Figure 71 Buffer outputs for asymmetrical input data................................................. 98

Figure 72 A simple ring oscillator built using odd number of inverter stages ........... 99

Figure 73 Variations in the zero-crossing of the output from an ideal reference due to jitter .................................................................................................................. 100

Figure 74 Simulation schematic of CML buffer used in the ring oscillator a) without and b) with gate PD model ................................................................................. 101

Figure 75 Ring oscillator output with and without PD model are compared with an ideal output .............................................................................................................. 102

Figure 76 Oscillator output’s low-high transition times for the three outputs .......... 102
Figure 77 Oscillator (with and without PD) output time periods for each cycle compared with the ideal case.
LIST OF TABLES

Table 1 Poly depletion widths at various bias conditions .................................. 53

Table 2 Run-length used in simulations................................................................. 97
CHAPTER 1

INTRODUCTION

Scaling of CMOS technology to 100 nm and below, along with subsequent increase in the frequency of operation, demands novel circuit topologies for high-speed logic applications. Aggressive scaling of CMOS devices also results in effects such as gate leakage and deterioration of transport characteristics. Most of the circuit design methodologies for future communication applications are broadly restricted by such constraints, making the study of these effects imperative in sub-100 nm devices. Moreover, the semiconductor industry is in need of modeling methodologies that focus on accurately predicting the sub-100 nm device configurations for ultra high-speed logic and enabling portability across 90 nm, 65 nm and 35 nm technology nodes.

Owing to the continuous scaling and increase in the speed of operation conforming to technology trends, next generation (NG) high speed circuits such as Serializer/Deserializer (SerDes) have to be designed for sub-100 nm feature sizes performing in the Gb/s range, as shown in Fig. 1. While faster transistors are consistently being developed as predicted by Moore’s Law, high-speed circuits also have stringent requirements in linearity, noise and the quality of the device characterization aside from merely attaining higher operating speeds. These requirements were not taken into account in the conventional digital IC design domain, which has driven the CMOS technology evolution. Consequently, the active device characterization for high frequency circuit applications is a domain that has suffered from neglect and many shortcomings. With constant scaling of the devices and the advent of operation in the Gb/s range, the
boundary between device characterization for high-speed digital and RF applications is vanishing. Hence the device modeling field is going through a paradigm shift, where the circuit design requirements can no longer be ignored and device models for high-speed digital and RF applications are merging. Traditional modeling approaches cannot be applied directly to these devices, but provide the basic frame-work to develop new models.

One of the major barriers in migrating to the sub-100 nm high-speed design regime is lack of accurate device models. The existing SPICE/BSIM models characterize the devices extremely well at low frequencies (for which they were intended), but suffer from many shortcomings in representing the high frequency behavior of transistors, particularly in predicting noise and impedance properties at GHz frequencies. Accurate characterization of these devices demands models and tools specifically developed for this purpose. Design cycle time and time to market while moving along the technology nodes can be reduced by incorporating reusable, industry-standard models. To enable

\[\text{Figure 1 Design constraints for NG SerDes}\]
portability across different technology nodes simulator-friendly, compact, scalable logic level circuits and models are required. Hence novel modeling methodologies and extraction techniques that focus on accurately predicting the sub-100 nm behavior have to be devised.

To meet the above needs, this work presents a gate model that accurately predicts the input impedance for the 1-20 GHz range. The model is based on the poly-depletion effect in the gate electrode and offers a simulator-friendly, compact, frequency-independent, and relatively portable solution. Unlike its predecessors, this model explains the behavior of input impedance with respect to scaling, frequency of operation and bias in the 1-20 GHz frequency. The performance of the new gate model is verified by comparing the simulation results with the measured S-parameter data for devices fabricated in CMOS technology. The devices are fabricated in 180 nm, 110 nm, and 90 nm processes to verify the portability of the model. Extensive measured data is collected from devices by replicating them across three different wafers in the 180 nm process and two wafers in the 110 nm and 90 nm processes. Fabricated devices include various combinations of widths, lengths and number of fingers across bias conditions ranging from 0 V to supply voltage in increments of 0.2 V at 25°C. Finally, the model is further modified to operate at a wide frequency range of 200 MHz-20 GHz.

The new gate model is employed in a current mode logic (CML) buffer, where it is shown that the frequency of the input data degrades the performance of a CML buffer designed at 16 GHz due to inaccuracies in the gate model. Contribution of the gate model parameters to the jitter calculation for a ring oscillator application is also shown. The
input impedance obtained using the poly depletion gate model is used as a tool to verify the model extraction techniques for common source and common drain configurations.

This work was fully funded by LSI Logic Corporation at Fort Collins, CO. The financial support included fabrication of all the devices used in this study by LSI Logic Corporation. Due to the nature of the project, device information is protected under a Non-disclosure agreement (NDA). However, the range of values are provided wherever applicable. The tasks shared by LSI Logic and RF/Analog/Mixed signal Lab (RAMLAB) at RIT in developing this work are presented here. The problems faced by modeling community in accurately representing the MOS gate electrode at 1-20 GHz frequencies were identified by LSI Logic. These issues were validated by RAMLAB with the help of measured data at those frequencies. The solution and theoretical analysis were provided by RAMLAB. Relevant test structures along with specific testing conditions (bias, geometry, and frequency) were developed at RAMLAB. Fabrication and complete testing of the devices in 0.1 – 20 GHz was carried out at LSI Logic Corporation. The raw data was analysed at RAMLAB to validate the proposed theory. Finally, relevant applications were developed at RAMLAB to verify the working of the new gate model. However, during the course of the project there was a close interaction between RAMLAB and LSI Logic Corporation.
CHAPTER 2

MOTIVATION AND BACKGROUND

Device modeling is an integral and essential process for successful development and production of semiconductor chips. As CMOS processes and device performances become complicated with scaling, existing models fail to incorporate these changes effectively. The lack of accurate models is a matter of concern for the IC industry as it affects the quality of the finished product and can cost millions of dollars in design cycle iterations. A typical methodology of modeling CMOS devices from model selection to model development is shown in this chapter. Though the MOS models have advanced over time, the techniques used to populate these models have not changed much and the model parameters are generated by curve fitting the measured data. The ill-effects of such techniques and the need for physics-based models are presented in this chapter.

2.1. RF MOS model evolution

One of the most important steps in the production cycle is verification of circuit performance before fabrication. The only way of achieving this is by simulating the circuit. The simulation results provide the necessary performance information such as circuit functionality, conformance to specifications and its limitations. Accurate simulation results are essential to ensure successful designs and aid in achieving higher yields. However, accuracy of the simulation results is limited by the accuracy of the models used for the simulations. The MOS electrical properties are translated to a set of equations and the required parameters are provided based on device measurements. The
process of device modeling from simulator selection to development of device design kits can be shown as in Fig. 2 as adapted from [1].

MOS models take different forms depending on various factors, such as frequency of operation, region of operation, process technology, and simulator tools. For example, when a MOS device is used as a logic gate, the device functionality is to operate as a switch. In simulation, this can be represented as a simple first order approximation, but as
circuits become more complex precise models are needed to correctly represent the
device. As the device scales down to lower feature sizes, certain phenomena such as drain
induced barrier lowering, surface scattering, punchthrough that are otherwise not present
in the long-channel devices come into effect. These effects have to be considered as the
simple models are no longer applicable to high-speed digital circuits. Due to the above
reasons MOS modeling techniques constantly evolve to meet the needs of state of the art
circuits and applications. This section presents the evolution of MOS models from the
analytical equations created to understand the MOS behavior to the present day state-of-
the-art models and highlights various shortcomings of these models.

In [2], MOS models are classified chronologically. The author states that the
initial models were physics-based analytical descriptions, which later changed to set, of
mathematical equations solely used for simulations. Interestingly, due to recent
advancements of CMOS technology, the models need to be developed again based on
physical effects. The progression of the MOS models is shown in Fig. 3, where the first
generation or pre-BSIM models are divided into three categories based on the level of
complexity involved [2]. This is followed by the BSIM family of models, with the
introduction of BSIM1 in 1984 [3]. BSIM1 modeled some basic device behavior but was
mostly empirical in nature, developed by fitting the measured data. This was followed by
the BSIM2 model, including features such as hot-electron effects, inversion layer
capacitance, and source/drain capacitance. These inclusions made BSIM2 more suitable
for analog applications unlike BSIM1 which was sufficient for digital applications. This
is followed by the next generation of models, BSIM3 and BSIM4. Introduction of BSIM3
marked the beginning of device-physics-based modeling, unlike its predecessors.
BSIM3v3, especially, incorporates many deep sub-micron effects and was widely accepted, making it an industry standard model. However, BSIM3v3 lacks some of the high-frequency effects which make it inaccurate at RF frequencies. BSIM3v3 can be a more complete RF model by adding a “sub-circuit”, which adds functionality to the model and is limited by the accuracy of the sub-circuit itself. The next version, BSIM4, addresses some of the high frequency issues unanswered by BSIM3v3, by including more elaborate models for gate resistance, substrate network and accurate gate current models. The complexity of MOS models for various regions based on the frequency of operation.
is shown in Fig. 4. The simplest form of the MOS device is the static or DC model that works well for frequencies less than 1 KHz or at DC [2].

<table>
<thead>
<tr>
<th>Frequency of operation</th>
<th>MOS equivalent circuits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td><img src="image" alt="DC MOS Equivalent Circuit" /></td>
</tr>
<tr>
<td>Low frequency</td>
<td><img src="image" alt="Low Frequency MOS Equivalent Circuit" /></td>
</tr>
<tr>
<td>High frequency</td>
<td><img src="image" alt="High Frequency MOS Equivalent Circuit" /></td>
</tr>
<tr>
<td>RF</td>
<td><img src="image" alt="RF MOS Equivalent Circuit" /></td>
</tr>
</tbody>
</table>

**Figure 4 Increase in model complexity based on frequency of operation**
The model represents the modulation of the drain current with the gate voltage, shown as dependent current source and the threshold voltage for all regions of operation. The low frequency models for analog applications include additional effects such as transconductance, output conductance and backgate transconductance [4]. As operating frequencies are increased, the model is further modified to include parameters such as intrinsic gate capacitance, overlap capacitances, parasitic and junction capacitances to accurately model the device [4]. But as the device moves into the RF regime, the model acquires greater levels of complexity, as shown in Fig. 4. The RF model is built by adding sub-circuit models to the “high-frequency” analog models such as BSIM3v3 [5]. As shown in the figure, the sub-circuit models typically include components such as gate resistance \( R_G \), gate-source overlap capacitance \( C_{GSO} \), gate-drain overlap capacitance \( C_{GDO} \), gate-bulk overlap capacitance \( C_{GBO} \), source series resistance \( R_S \), drain series resistance \( R_D \), source-bulk junction diode \( D_{SB} \), drain-bulk junction diode \( D_{DB} \), and substrate resistances \( R_{SB} \), \( R_{DB} \) and \( R_{DSB} \). These parameters are usually extracted empirically from the measured data.

### 2.2. Limitations of the present-day modeling techniques

It has been shown that the BSIM versions 3.3 (with sub-circuits) and 4 are the industry standard models for GHz applications. Though being the models of choice, these models suffer from various shortcomings, and some critical inadequacies are highlighted in the subsequent sections.

#### 2.2.1. Sub-circuit parameter extraction

Advances in CMOS technology have increased transition frequency \( f_T \), and provided cost-effective solutions for RF applications by replacing GaAs and SiGe with
CMOS technology. On the other hand, sub-circuit models aided designers in implementing CMOS solutions by characterizing device behavior at RF. However, the process of extracting these models is primitive in nature and increases design-cycle time frames immensely as CMOS technology scales. It is well known that these sub-circuit models are built empirically by fitting measured data and adding components that accurately fit the data. Hence the device operates very well at the dimensions, frequency and bias conditions at which the sub-circuit model parameters are extracted. This poses various limitations on the use of these models as follows:

- Sub-circuit models are foundry dependent, limiting the migration of design across different processes. In other words, a circuit designed in a particular process using the sub-circuit models from that foundry is not applicable to other processes even though the technology remains the same. This results in redesigning the circuit for a new set of models whenever the foundries are changed.

- The sub-circuit models impose limits on the device dimensions that can be used in the design. This is due to the inclusion of the dimension-based sub-circuit parameters outside the intrinsic compact device models. These sub-circuits are usually extracted by fixing the width of the fingers and by varying the number of fingers (NF). Hence the device dimensions for simulations are predetermined with some flexibility provided by NF.

- Sub-circuit components are usually bias dependent, but they are extracted at specific bias conditions and represented as passive components, degrading the performance of these models at other bias conditions.
• The quality of the model depends on the measurement setup of the vendors. Due to lack of standard measurement techniques there is a wide variance in the performance of models provided by various vendors.

• Due to the empirical nature of these models, the components are added based on the best fit to measured data. At high frequencies, as sub-circuits get more complex, simulation speeds are affected due to the large number of model components.

• As the technologies scale, the time to develop models for the new technologies is subject to delays caused by fabrication and the model development. It has been reported in [6] that companies incur tremendous losses due to inadequate models provided by the vendors, affecting design-cycle times.

2.2.2. Other limitations

BSIM4 provides a better high-frequency solution than BSIM3v3 with the sub-circuit, but it fails to address certain effects that cannot be ignored at high frequencies, such as the poly depletion effect. Another drawback of BSIM4 is symmetry, which results in errors at zero drain bias conditions. Some of these issues are being addressed in the next version, BSIM5 model. However advancements of modeling techniques cannot be fully utilized if the parameter measurement and extraction are not modified accordingly. Even though an accurate model is available, it cannot be used if the vendors fail to provide the model parameters. The high frequency substrate network model is one such example, where no set standards or techniques are available to extract these parameters.
In summary, the MOS model has come a long way from conception of the MOS devices, incorporating various trends and advancements in the MOS technology. CMOS technology has revolutionized the IC industry by offering novel, cost effective solutions. However the models and modeling techniques did not develop at the same pace, affecting design time and time to market. Present day models fail to meet industry needs mainly because of the practice of traditional non-physics-based methods such as sub-circuits. BSIM3v3 and BSIM4 models appear promising but are not without faults when it comes to high frequency performance. Hence the CMOS IC industry is in need of physics-based, accurate, compact, portable models for next generation high-speed circuits.

**2.3. Future of MOS models-need for poly depletion model**

Use of polysilicon for gate electrode has several advantages such as, self-alignment during source and drain implantation making it self-aligned process, its ability to withstand heat treatments during activation of source and drain dopants and use of dual work function doped polysilicon gate to reduce short channel effects. However as technology advances, it is projected that polysilicon gate will be replaced by metal gate due to effects such as, polysilicon depletion, incompatibility of polysilicon with high-\(k\) dielectrics and boron penetration.

Even though it is projected that metal gates will replace polysilicon, CMOS devices used polysilicon gates up to 65 nm technology node. Research is still undergoing to develop new metals, processes and metal – high-\(k\) dielectric combinations. The standards to implement metal gate technology are still not developed. Moreover, industry lags advancements in the technology. It is still not apparent if the industry on the whole
will shift to metal gate process. Even if the shift occurs it will take time for the industry to adapt it completely. Till then the industry is faced by problems concerning polysilicon gates. This work addresses such issues and provides a solution to the current and near future problems.
CHAPTER 3

MOS Gate Characterization

Scaling below 100 nm along with simultaneous increase in the frequency of operation to the GHz region gives rise to various effects in MOS devices that were either not present or were insignificant in previous technologies. One such effect is depletion in the gate polysilicon electrode which dominates only at GHz frequencies. The gate electrode is thoroughly analyzed in this chapter and various techniques to model the low and high frequency effects reported in literature are presented. The inadequacy of these models in reporting the poly-depletion effect, and a technique to model it are presented.

3.1. Gate electrode modeling

Modeling the gate accurately is very important as it impacts various performance factors such as power gain, transconductance, thermal noise, input match, distortion at RF frequencies, etc. Further inaccuracies in the gate models introduce errors in estimating the effective gate-to-source voltage, which results in faulty estimation of device operation. The gate models take different forms based on the complexity required for the particular application, which in turn depends on device dimensions, frequency of operation, etc. The progression of the gate model from simple low frequency representation to complex RF model as reported in the literature is presented here.

3.1.1. Low-medium frequency gate model

In its simplest form, the gate electrode is represented by oxide capacitance, $C_{ox}$, in series with the capacitance in the channel region $C_C$ as shown in Fig. 5. The complexity of the gate electrode model increases for quasi-static conditions, where various intrinsic
and extrinsic capacitances are included in the model [7]. The quasi-static gate model is shown in Fig. 6, with intrinsic and extrinsic components.

![Figure 5 Simple gate electrode model comprising of oxide and channel capacitances](image)

The intrinsic components represent capacitive effects that arise due to variations in the gate charge, \( Q_G \) with respect to different terminal voltages. The dependences of each of the gate intrinsic capacitances are given by the following expressions:

\[
C_{gs} = \left. \frac{\partial Q_G}{\partial V_S} \right|_{V_G,V_D,V_S}, \quad C_{gd} = \left. \frac{\partial Q_G}{\partial V_D} \right|_{V_G,V_D,V_S}, \quad C_{gb} = \left. \frac{\partial Q_G}{\partial V_B} \right|_{V_G,V_D,V_S}
\]  

(1)

Parameters relevant to the gate terminal only are shown in Fig. 6 and all the other parameters between source, drain and body terminals have been omitted. The charge stored in the extrinsic part of the device is modeled by \( C_{gse} \), \( C_{gde} \) and \( C_{gbe} \). These parameters consist of two parts, overlap (\( C_{ov} \)) and fringe (\( C_f \)) capacitances as shown in Fig. 6. This quasi-static model is applicable over low-medium frequency range. At frequencies where non-quasi static effects become dominant, these models do not provide accurate results.
3.1.2. Gate model including scaling effects

The models considered so far do not include any short-channel effects that arise due to scaling. In this section further modifications to the gate model based on scaling are presented. One of the important effects that cannot be ignored for short channel devices is gate resistance. Before proceeding to the gate resistance model, the effect of gate resistance on various performance parameters is presented below [8].
3.1.2.1. Impact of gate resistance on device performance [8]

Gate resistance affects the device performance when the transistor width is increased (to increase current or transconductance), especially as device dimensions scale below 100 nm. The impact of gate resistance can be reduced by using silicidation or multiple finger structures of the same total width. These remedies come with limitations such as higher resistivity for short length devices or higher drain/source junction capacitances. Hence the impact of gate resistance on some of the important performance parameters of the MOS device is presented here.

I. Cut-off Frequency ($f_T$)

In the lumped gate resistance model, gate resistance appears as a resistor in series with ideal current source in the $f_T$ measurements and shows no effect on $f_T$. Even with the distributed gate resistance model, it has been shown that the current gain (and hence $f_T$) is independent of the gate resistance. Practically, the gate resistance still induces errors, leading to faulty measurement of $f_T$.

II. Maximum Frequency of oscillation ($f_{max}$)

The relation of $f_{max}$ with gate resistance is, $f_{max} \propto \frac{1}{\sqrt{R_g}}$ for the lumped model. For the distributed model, $R_g$ is replaced with $\frac{R_g}{3}$ for accurate predictions.

III. Thermal Noise:

The gate resistance contributes to the thermal noise of the device. For uniformly distributed MOS, the input-referred noise due to the gate resistance is modeled as a voltage source given by $4kTB\frac{R_g}{3}$, where $k$ is the Boltzmann’s constant, $T$ is the temperature, $B$ is the bandwidth and $R_g$ is the gate resistance. The gate resistance, along
with the gate to drain and gate to source capacitance, develops a time constant that becomes significant if the device is driven by a circuit with impedance comparable with \( \frac{R_s}{3} \).

### 3.1.2.2. Gate resistance model

Polysilicon has long been used as the gate electrode in CMOS technologies due to its excellent process compatibility. One of the drawbacks of using polysilicon is high resistivity compared with metals. The resistance in a poly gate electrode has a distributive effect, due to continuous transfer of charge from the gate electrode to the channel [9]. For poly gates, the charge entering at one end of the gate electrode does not come out at the other end due to capacitive coupling between the electrode and the channel. This effect varies the resistance along the width of the gate electrode, resulting in a distributive effect as shown in Fig. 7. Studies on the distributed gate resistance have shown that this resistance can be approximated to a lumped resistor [8] [10]. The electrical equivalent of the distributed gate resistance and the lumped approximation is shown in Fig. 8 [8]. Based on various techniques presented in [8][9][10], the distributed resistance can be approximated as:

\[
R_g = \frac{W \times R_s}{3L}
\]  

where \( W \) is width of the transistor, \( R_s \) is the sheet resistance of gate material, \( L \) is the length of the transistor and the factor 1/3 is attributed to the distributed nature of the gate resistance. [11] describes a technique to extract the gate resistance from measured S-parameters as:
\[ R_g = \text{Re} \left[ \frac{1}{y_{11}} \right] \]  

(3)

where \( y_{11} \) is deducted from the measured scattering parameter \( S_{11} \). In [11], a gate model considering only the gate resistance was used for an LNA and a mixer in 0.5 \( \mu \)m process at 2.4 GHz. It was shown that the gate resistance reduces the gain factor and increases the 1-dB compression point.

![Figure 7](image)

**Figure 7** Distributed gate resistance in the gate electrode as a result of capacitive coupling between gate and the channel

![Figure 8](image)

(a)

(b)

**Figure 8** a) Electrical equivalent of distributed gate resistance and b) Lumped approximation of the gate resistance

3.1.3. Gate model for RF

The gate models seen so far work very well up to several hundred MHz, but they fail at higher frequencies due to exclusion of components that are needed to model device
behavior at these frequencies. This section presents the modifications required in the gate model for high frequencies.

3.1.3.1. BSIM3v3 extension

In [12] and [13] a simple RF gate model is realized by adding a parasitic resistor to BSIM3v3 model gate, as shown in Fig. 9. In this model, the gate resistance \( R_g \) accounts for both physical gate resistance and non-quasi static (NQS) effect. To accurately model both the effects, measured input impedance data for device in 0.35 \( \mu \)m and equation (3) were used. The extension of BSIM3v3 model was evaluated at circuit level in [12] and it was shown that it can predict the device behavior accurately up to 10 GHz.

![Figure 9 BSIM3v3 RF model [12]](image)

3.1.3.2. Non Quasi Static gate model

The NQS effect arises at high frequency operations, where the rise and fall times of signals are comparable to or smaller than the channel transit time. It has been shown that QS models cannot be used to predict the transadmittance at high frequencies [14]. It has also been shown from measurement results that NQS effects are present in devices with gate lengths in the range of 0.35 \( \mu \)m and has strong frequency dependence [15]. The BSIMv3 RF model in Fig. 9 incorporates the NQS effect using the lumped gate resistor \( R_g \). The value of \( R_g \) is obtained from curve fitting of measured impedance profiles.
Consequently it is not easy to model bias and geometry dependence of \( R_g \) by this method. Hence a bias-dependent \( R_g \) model was proposed in [16] by splitting \( R_g \) into two parts as shown in Fig. 10.

![Figure 10 NQS model with \( R_{eltd} \) and \( R_{ch} \) as given in [14]](image)

The gate resistance equation is given by:

\[
R_g = R_{eltd} + R_{gch}
\]  

(4)

where \( R_{eltd} \) is the physical resistance, independent of bias and frequency and its lumped approximation is:

\[
R_{eltd} = \frac{\alpha W}{L} + \beta
\]  

(5)

\( R_{eltd} \) is the gate material sheet resistance, \( \alpha \) is equal to 1/3 and 1/12 when gate terminal is brought from one side and two sides, respectively. Here, \( \beta \) is the external gate resistance factor. The channel resistance \( R_{gch} \) in equation (4) is calculated as:

\[
\frac{1}{R_{gch}} = \gamma \left( \frac{1}{R_{st}} + \frac{1}{R_{ed}} \right)
\]

(6)

where \( R_{gch} \) is the channel resistance seen from gate, \( R_{st} \) is the static channel resistance, \( R_{ed} \) is the gate voltage dependent excess channel charge distribution and \( \gamma \) accounts for
the distributive nature of the channel charge and $C_{ox}$. $R_{st}$ and $R_{cd}$ both determine the time constant of the NQS effect.

3.1.3.3. BSIM4 gate model

The gate impedance was improved considerably in BSIM4, to include more accurate high frequency and scaling effects. The high frequency gate resistance model in BSIM4 includes the NQS effects described in the previous section. The new gate model with a non-linear resistance along with the gate electrode resistance is shown in Fig. 11 [17]. The resistance $R_{gelh}$ is the gate electrode resistance calculated using the lumped gate resistance approximation, the resistance $R_{ii}$ is the non-linear resistance obtained by using drift and diffusion current expressions. The capacitors $C_{gs0}$ and $C_{gd0}$ are the gate to source and gate to drain capacitors, respectively.

![Gate resistance model implemented in BSIM4.2.1](image)

The resistance $R_{ii}$ is not a physical resistance but an equivalent resistance used to represent the first order NQS effects in the channel. Hence there is no thermal noise source associated with this resistance, but it is a critical parameter for matching the noise data. The resistance $R_{ii}$ can be calculated using,
\[
R_i = \frac{1}{XRCRG1 \times \left( \frac{I_{DS}}{V_{DS,eff}} + XRCRG2 \frac{kT}{q} \mu_{eff} \frac{W_{eff}}{L_{eff}} C'_{ox} NF \right)}
\]

where XRCRG1 and XRCRG2 are fitting parameters, \(I_{DS}\) is the drain current, \(V_{DS,eff}\) is the effective drain to source bias, \(kT/q\) is the thermal voltage, \(\mu_{eff}\) is the effective voltage, \(W_{eff}\) is the effective width, \(L_{eff}\) is the effective length, \(C'_{ox}\) is the per-area oxide capacitance and NF is number of fingers. BSIM4 provides the flexibility in including these parameters through a model parameter RGATEMOD. When this parameter is set to ‘0’, no gate impedance is added. For RGATEMOD =1 just the lumped gate resistance \(R_{geltd}\) is added to the intrinsic gate terminal, for RGATEMOD=2, the resistance \(R_{ii}\) is added in series with \(R_{geltd}\), and for RGATEMOD=3, the position of \(R_{ii}\) is moved such that it is in parallel with the two capacitances (as shown in Fig. 11).

Another effect that is direct result of device scaling is gate tunneling, which is very important for sub-100 nm devices. As the channel length is scaled for the device to perform as a transistor, the gate oxide is scaled correspondingly to exert control on the channel. The oxide layer is roughly a few nm for sub-100 nm technologies, and such thin oxide layers yield to leakage currents due to several tunneling effects. The main contributors to the leakage current are hot carrier effects, resonant tunneling, and direct tunneling [18] [19] [20]. Various gate leakage components considered in BSIM4 are shown in Fig. 12a and the corresponding dc equivalent circuit is shown in Fig. 12b [9]. The currents \(I_{gd,tunnel}\) and \(I_{gs,tunnel}\) are tunneling currents between the gate terminal to source and drain terminals, \(I_{gcd,tunnel}\) and \(I_{gcs,tunnel}\) denotes tunneling current from gate to the respective terminals through the channel, and \(I_{gb,tunnel}\) is the current between gate and the bulk.
Figure 12 a) Gate leakage currents considered in BSIM4 and b) Electrical equivalent of the gate leakage currents

### 3.1.4. Gate models for high frequency applications

The BSIM4 model, although being a standard model for high frequencies, does not include all the high frequency effects in the gate. To compensate for this shortcoming, additional sub-circuit models are required to model the gate at high frequencies. These models are added to the intrinsic device model at the gate terminal. One such model to improve the lumped approximation to accurately predict the y-parameters and noise in the frequency range of 1-5 GHz is presented in [21] and is shown in Fig. 13.
Recent developments in MOSFET gate modeling [22] [23] show that the effect of gate impedance is significant up to the 90 nm process node. The studies at RF frequencies show a resistive droop in $S_{11}$ which is due to complex behavior of gate impedance at those frequencies [23]. This characteristic is modeled empirically by modifying the lumped gate resistance in [23]. The new gate model corrects for both the real and imaginary parts of the input impedance with very good accuracy. It was also proved with measured data in [23] that the droop is absent when the device is biased below its threshold voltage, proving the bias dependence of the gate model parameters. The improved model for a grounded source transistor is shown in Fig. 14. However, $R_g$, $R_{gc}$ and $C_g$ in Fig. 14 are obtained empirically and hence cannot be scaled with technology. Even though the gate resistance dependence on bias voltage is mentioned in [24] [9], the gate model parameters are characterized for typical bias voltages. Another aspect that is lacking in most of the present models is the effect of scaling on the gate impedance. Due to the empirical nature of the models, they are confined to particular technology and device dimensions. Hence there is a need to develop a physics-based, bias-dependent characterization of gate impedance of MOSFETs for high-speed applications.
The evolution of gate modeling techniques based on frequency of operation and scaling were presented in this section. It was seen that BSIM4 captures most of the RF and scaling effects. However, for next generation devices that operate below 100 nm and frequencies in 1-20 GHz range, these models are not sufficient [25]. As these models do not include effects that are dominant only at those frequencies (such as poly depletion) they have to be included in next-generation gate models [26]. The gate model is further developed in this work, to include this effect. The following sections explain this development in detail.

3.2. Effects of scaling and frequency on gate performance

Before proceeding to develop the gate model in the 1-20 GHz range based on depletion in the poly region, the behavior of the gate at these frequencies is studied for short-channel devices. Measured device input impedances from test structures designed and fabricated in 180 nm, 110 nm and 90 nm processes are used as the performance measure in all regions of operation.

3.2.1. Variation of input impedance with width and length

The effect of device dimensions (width & length) on input impedance is presented here. The real part of measured input impedance ($Z_m$) with respect to total width is
plotted varying single-finger width, keeping the number of fingers (NF) constant in Fig. 15. The real part of $Z_m$ for devices in 110 nm process, for different single-finger widths, is shown in Fig. 15a and Fig. 15b for NF=16 and 8, respectively. It can be seen from Fig. 15 that the input impedance has strong frequency dependence. The variation in $\text{Re}\{Z_m\}$ with frequency is linear for larger widths compared to smaller widths, making width dependency of $Z_m$ non-linear with frequency. This behavior is consistent with devices in 180 nm and for 90 nm processes shown in Fig. 15c, and 15d.

![Figure 15](image)

Figure 15 a) Real part of $Z_m$ for 0.11um process for different single-finger W and with NF = 16; b) NF=8; c) for NF=16 in 180 nm process; and d) for NF=16 in 90 nm process

The variation in the input impedance with device length, keeping the width and NF constant is shown in Fig. 16a, Fig. 16b and Fig. 16c for 110 nm, 180 nm and 90 nm processes, respectively. From these figures, it can be seen that, as the length decreases or
as the number of the squares in the gate region increases, the input resistance increases. However, in contrast to the complex behavior observed in the 'constant L' cases, the shape of the length-dependent curves remains constant for various L values. Hence, the single-finger width and length parameters have different effects on the input impedance. This observation holds for devices in 180 nm, 110 nm and 90 nm processes.

![Graphs showing input impedance variation with frequency for different L values](image)

**Figure 16** a) Input impedance varying L for 110 nm; b) for 180 nm; and c) for 90 nm

### 3.2.2. Variation with frequency

As seen above, the input impedance of the device has a strong frequency dependency, and this dependency is studied in detail here. It has been observed that the variation in input impedance is a complex mechanism, with a non-linear frequency dependence. This is evident from $Z_m$ of a device, with single-leg dimension of
20μm/0.11μm shown in Fig. 17. The input impedance curve in Fig. 17 has different slope with respect to frequency as marked by dashed and solid circles. This indicates that the resistive part of the device input impedance is a combination of different topologies, each acting at a different frequency range.

![Graph showing input impedance](image)

**Figure 17 Input impedance of a device with single-leg dimension of 20 μm/0.11 μm**

It has been noted that this shelving effect—i.e., the variation of $Z_m$ between two very different low and high frequency value asymptotes—is absent (or diminishing) in devices with lower single-leg W/L dimension. This can be seen from Fig. 18 for a device with (single-leg) W/L dimension of 5 μm/0.11μm in the same process. Hence it can be concluded that the behavior of the input impedance is a strong function of the device width or number of squares in the gate region, and has to be modeled by an appropriate gate impedance model.

### 3.2.3. Variation with gate bias

The bias dependence of the input impedance is shown in Fig. 19a for 180 nm, Fig. 19b for 110 nm and Fig. 19c for 90 nm processes. The data shows that the amplitude
change in input impedance reduces with the increase in gate bias across all technology nodes.

![Graph](image)

**Figure 18 Input impedance of a device with single-leg dimension 5 μm/0.11 μm**

### 3.3. Poly-depletion based gate model

In the previous sections, the shortcomings of present day models in the high frequency regime and the impact of gate impedance on device performance in the 1-20 GHz frequency range have been shown. In this section, the gate is modeled by incorporating depletion in the polysilicon gate and also explaining the physical phenomena behind the behavior of input impedance with respect to scaling, frequency and bias conditions. A semi-empirical poly-depletion gate model is developed for the 1-20 GHz range. The model is verified with the help of measured input impedance data for three different technology nodes.

#### 3.3.1. Poly depletion effect

Polysilicon depletion or poly depletion effect (PDE) results due to depletion of carriers at the poly-SiO₂ interface and can also be attributed to insufficient redistribution and/or insufficient activation of dopant near the interface [27]. When the device is biased in strong inversion, the electric field in the oxide penetrates into the gate electrode and
depletes the polysilicon. PDE introduces an additional capacitance in series with the gate oxide capacitance, reducing the overall gate capacitance. This leads to reduction in inversion channel charge density and transconductance. Also as the channel length shrinks, the gate oxide is made thinner, which in turn makes $C_G/C_{OX}$ in the inversion region sensitive to poly-silicon doping concentration [28]. Hence as CMOS technology scales down, PDE contributes significantly to the total gate capacitance. PDE is also a major concern when the gate is not heavily doped or counter doped as in dual gate processes.

Figure 19 Input impedance Vs. $V_{gs}$ for devices in a) 180 nm; b) 110 nm; and c) 90 nm technologies
3.3.2. PDE impact on device performance

Apart from varying the gate capacitance, PDE affects various aspects of device and circuit performance parameters. As devices scale down, the gate oxide thickness reduces relatively, to reduce the short channel effects. It has been shown in [29] that thinning of oxide along with the high doping gate electrode causes mobility degradation in the channel due to Remote Charge Scattering (RCS). It has also been shown that RCS reduces the peak effective mobility by 20% for a MOSFET with oxide thickness of 1 nm. Another factor that is affected by PDE is threshold voltage shift, which is shown in [30]. The shift in the threshold voltage due to poly-gate depletion can be as large as 68 mV for poly doping of \( N_D = 10^{19} \text{ cm}^{-3} \), and drops down to about 18 mV for \( N_D = 2 \times 10^{20} \text{ cm}^{-3} \). This observation suggests that poly-gate depletion must be accounted for if accurate results for the threshold voltage are desired. Moreover, low doping of poly gates causes rapid reduction in linear current due to hot carriers [31]. Under stress conditions used in [31], the effect seen on linear current due to hot carriers is larger when poly is biased into depletion than accumulation over the full length of the channel. This is due to the fact that for the stress conditions in [31], the damage due to stress is more at the drain side where poly is in accumulation \( (V_g < V_d) \) rather than depletion and does not contribute to the overall damage. Hence poly doping level does not have a large impact on the amount of damage measured when the gate biased into inversion or accumulation.

Scaling of devices increases the cut-off frequency \( (f_t) \), as \( f_t \) increases proportionally to transconductance. Hence effects such as PDE become inevitable for modern, thin oxide RF MOSFETs due to series connection of the poly depletion capacitance with the oxide capacitance, leading to degradation of inversion gate
capacitance and transconductance. In analog MOS circuits, a purely sinusoidal input signal can produce a distorted output signal with higher-order harmonics due to the nonlinearity of MOS transistors. These harmonics are mainly induced by higher-order derivatives of the CV characteristics of the transistors. In particular, a lower limit on the distortion of MOS transistors is determined by the third-order derivative of the drain current with respect to gate bias \( \left( g_{m3} = \left( \frac{\partial^3 I_{ds}}{\partial V_{gs}^3} \right) \right) \). Therefore, the amplitude of the third-order derivative of \( I_{ds} \) should be minimized for low-distortion applications. The measured data in [32] shows that the amplitude of \( g_{m3} \) for 70 nm device is greater than that for 0.25 \( \mu \)m device with a thicker oxide. Also, there are multiple zero-crossing points for the device with 70 nm. These results imply that the linearity of MOS transistors can be worse for scaled devices with very-thin oxide thickness. It is also shown that the third derivative of \( I_{ds} \) which is the measure of distortion is actually dependent on the 2\(^{nd}\)-order derivative of the gate capacitance with respect to \( V_{gs} \). Hence the total gate capacitance in strong inversion including PDE is given by:

\[
C(V_{gs}) = \frac{\varepsilon_{ox}}{t_{ox} + t_{dp}(V_{gs})} = \frac{\varepsilon_{ox}}{t_{ox}^2 + \frac{2\varepsilon_{si}}{9qN_p} \left( V_{gs} - V_{fb} - \phi_s(V_{gs}) \right)}
\]

where \( t_{dp} \) is the poly depletion thickness, \( N_p \) is the poly doping, \( V_{fb} \) is the flat band voltage and \( \phi_s \) is the surface potential. From this equation it is apparent that a nonlinear relationship exists between the gate terminal voltage and the gate capacitance in the presence of PDE. When the PDE is severe, MOSFET distortion behavior becomes worse, since the term \( \left( \frac{\partial^2 C(V_{gs})}{\partial V_{gs}^2} \right) \) becomes substantial. These observations imply that the
RF distortion will be degraded in scaled MOS with ultra-thin oxides, which results from the nonlinear gate capacitance behavior under the influence of PDE.

It has been shown above that PDE shifts the device threshold voltage, reducing device current driving capability. This has an adverse effect on transistor DC I-V characteristics. The DC component of the PDE degrades circuit speed performance. However, PDE also reduces transistor gate and overlap capacitances and speeds up the circuits. The impact of PDE on circuit performance was studied on an 11-stage ring oscillator circuit constructed in [33]. With the help of simulations it was shown that the delay time under DC PDE is longer than that under the DC+ac PDE. It was also shown that the percentage increase in delay time under DC+ac PDE behaves similarly as the DC PDE case, except that it has stronger $t_{ox}$ dependence. PDE on circuit performance at different power supply voltages was also studied in [33]. The results show that circuit performance degradation due to PDE becomes more significant during low-power operation. This is due to the fact that circuit performance becomes more sensitive to threshold voltage shift caused by PDE at low power supply voltage. The result clearly suggests that by scaling down gate oxide thickness, DC part of PDE can be significantly alleviated at low power operation. From the above observations it can be concluded that thinner gate oxide reduces the dominance of PDE at DC and also degrades the performance due to low power operation.

3.3.3. Analytical and numerical analysis of PDE

Various techniques are reported in the literature to analyze PDE and calculate the gate depletion capacitance for different applications. In this section some of these techniques are presented. A quantum C-V mechanism to correctly predict effective oxide
thickness including PDE is presented in [34]. The poly depletion capacitance is incorporated by approximating it as, \( C_{GD} \equiv qN_D F_S \), where \( q \) is electron charge, \( N_D \) is poly doping concentration and \( F_S \) is surface electric field. A charge-sheet-approximation-based model for tunneling capacitance that includes PDE capacitance, is described in [35] for a floating-gate EEPROM cell. The floating gate-depletion effect was calculated as:

\[
C_p = \sqrt{\frac{qN_s \varepsilon_0 \varepsilon_{si}}{2\psi_p}} \tag{9}
\]

where \( C_p \) is the floating gate depletion capacitance, \( N_g \) is gate polysilicon doping and \( \psi_p \) is the surface potential. \( \psi_p \) is calculated using charge neutrality as:

\[
\psi_p = \frac{(Q_{sc})^2}{2qN_g \varepsilon_{si}} \tag{10}
\]

where \( Q_{sc} \) is substrate charge and is a function of \( N_g \) and surface potential.

A physics-based PDE model to accurately predict the performance of a MOS capacitor in accumulation is given in [36]. \( C_{dep}' \), the capacitance of the gate depletion region per unit area in [36] is calculated as:

\[
C_{dep}' = \frac{dQ_G'}{d\psi_{dep}} = \frac{\gamma_p^2 C_{ox}'}{2(V_{GB} - \phi_{MS} - \psi_{S})} \tag{11}
\]

where \( V_{GB} \) is the applied gate voltage, \( \gamma_p \) is body effect coefficient for poly, \( \phi_{MS} \) is the work function of the bulk and gate materials. From the above equation the capacitance in the depletion region can be calculated for a given \( V_{GB} \), and the surface potential can be calculated as:

\[
\psi_{S} \approx 2\phi \left( \frac{V_{GB} - V_{FB} + 3\phi}{V_{GB} - V_{FB} + 6\phi} \right) \log \left( 1 + \frac{V_{GB} - V_{FB}}{\gamma \sqrt{\phi}} \right) \tag{12}
\]
where $\phi_t$ is the thermal voltage, $V_{FB}$ is the flat band voltage and $\gamma$ is well body effect coefficient.

The analytical model presented in [30] obtains the poly-gate capacitance by solving the oxide field $E_{ox}$ expression given by:

$$E_{ox} = \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \sqrt{\frac{2qN_D}{\varepsilon_{Si}}} G(\phi_p)$$

where

$$G(\phi_p) = \left[ \frac{\phi_p}{N_D} \left[ d \right] \right]$$

$$G(\phi_p) = \left[ \phi_p + N_C \frac{-\phi_p}{N_D} \left[ d \right] \right]$$

Poly-capacitance is then calculated as:

$$C_{poly} = \left[ \frac{\sqrt{qN_D}\varepsilon_{SC}}{2G(\phi_p)} \right] \left[ 1 - \frac{N_C}{N_D} F_{1/2} \left( -\frac{q\phi_p + \Delta E_{CP}}{k_BT} \right) \right] + \frac{N_C}{N_D} F_{1/2} \left( \frac{\Delta E_{CP} - E_G + q\phi_p}{k_BT} \right)$$

The gate voltage is given by:

$$V_G = \phi_{SC} + \phi_p + \frac{(\Delta E_{CP} - \Delta E_{CB})}{q} + V_{ox}$$

where $\phi_{SC}$ is the surface potential in the semiconductor side of the semiconductor/oxide interface, $\Delta E_{CB} = (E_C - E_F)_{sc}$ from the charge neutrality condition in bulk silicon and $V_{ox} = E_{ox}t_{ox}$ is the voltage drop in the oxide. The total gate capacitance per unit area is then evaluated as a series combination of $C_{poly}$, $C_{ox}$ and $C_{inv}$. 
3.4. Gate model with PDE

It has been shown in the above sections that PDE has impact on various performance factors of the device such as mobility, threshold voltage and even distortion at RF. Recently it has also been reported [26] that next generation modeling techniques have to include PDE for accurate representation of MOS behavior. Due to the above reasons, it is imperative to include PDE while modeling the gate. The proposed technique focuses on modeling gate behavior based on the input impedance and its dependence on PDE for the 1 – 20 GHz frequency range.

3.4.1. Model description

PDE effect in the device gate electrode is a well studied phenomenon as presented in the previous sections. There are mechanisms reported in literature to model this effect. As extensive as the prior work has been, the focus has primarily been on calculating and numerically modeling the capacitive portion of the gate impedance which results from the PDE. The effect of PDE on gate impedance and overall device and circuit performance in the GHz regime is an area that needs to be explored further. This work develops a simulator-friendly input impedance model which includes PDE and can be used in conjunction with BSIM (3 or 4) models.

The main component that has been associated with PDE in earlier analyses is the capacitance resulting due to depletion in the poly gate electrode. Unlike the earlier PDE modeling techniques, this work reports a combined RC effect on the input impedance. An illustration of PDE and the equivalent electrical model is shown in Fig. 20. The polysilicon gate electrode can be treated as a capacitor, $C_{dep}$. This capacitance is comprised of a (poor) semiconductor dielectric and low carrier density in the depleted region, and gives
rise to a resistive effect, modeled by $R_{\text{dep}}$ in parallel with $C_{\text{dep}}$. $R_{\text{gate}}$ is the resistance of poly-silicon electrode, and $C_{\text{ox}}$ is the oxide capacitance.

![Schematic representation of poly-depletion based gate model](image)

**Figure 20** Schematic representation of poly-depletion based gate model

The parameters $C_{\text{dep}}$ and $R_{\text{dep}}$ are lumped approximations of the resistance and capacitance in the depleted region. The distribution of resistance $R_{\text{dep}}$ is complex and varies along the length and width of the depletion region. Moreover, this resistance has a 2-D distribution along the width and length of the gate electrode as shown in Fig. 21.

![$R_{\text{dep}}$ distribution along the width and length of the gate electrode](image)

**Figure 21** $R_{\text{dep}}$ distribution along the width and length of the gate electrode
Another factor that contributes to the spatial variation of this resistive effect is the carrier distribution profile along the depletion thickness, \( t_{pd} \). The carrier distribution profile along different regions is shown in Fig. 22, where it can be seen that the carrier concentration varies from \( N_{poly} \) in the un-depleted gate electrode region to zero in the oxide, through a parabolic distribution in the depleted region. This results in the unequal distribution of the vertical component of the poly resistance, \( R_{vertical} \) along the depletion thickness shown clearly in Fig. 23. The resistance has a distributed effect along the length of the channel and varies according to the charge distribution in the depletion region. In addition, gate resistance is also distributed along the width of the device. Hence modeling the resistive effect is not straight forward, while the capacitive effect can be easily obtained by calculating the depletion thickness \( t_{pd} \), as explained in the next section.

The new gate model representation also successfully explains the dependence of input impedance on \( W, L, \) and frequency. It was observed in Section 3.2 that the input impedance has a non-linear relationship with the device width. This can be explained

![Figure 22 Carrier concentration profile in the gate electrode under poly-depletion](image-url)
with the help of the parallel RC model. As the width of the device decreases, \( C_{\text{dep}} \) increases while \( R_{\text{dep}} \) decreases resulting in a net increase in the input impedance. Due to

![Diagram](image)

*Figure 23 Distribution of \( R_{\text{dep}} \) along the depletion thickness due to non-uniform variation of the carrier concentration*

the domination of \( C_{\text{dep}} \) at low frequencies, the impedance is higher for small devices at low frequencies, as observed with the measured input impedance. The variation of the input impedance with respect to length was observed to be linear i.e, the input impedance decreases as the length is increased. This is due to the linear dependence of \( C_{\text{dep}} \) and \( R_{\text{dep}} \) on the device length. As the device length increases both \( C_{\text{dep}} \) and \( R_{\text{dep}} \) decrease, resulting in a net decrease of the input impedance. \( C_{\text{dep}} \) and \( R_{\text{dep}} \) collectively model the variation in input impedance with frequency. It was observed that the \( \text{Re}\{Z_m\} \) varies between two very different low and high frequency value asymptotes. This can be explained by the combination of different RC networks each dominant over a different frequency range.

The depletion region RC network models the input impedance behavior in the low frequency range (1-12 GHz), beyond which \( C_{\text{dep}} \) shorts \( R_{\text{dep}} \) and a high frequency RC network dominates. It can be recalled that this high frequency RC network was
incorporated by intrinsic gate resistance model for BSIM4 and by an additional RC network for BSIM3. Hence the depletion effect explains the input impedance behavior with respect to physical dimensions and frequency of operation.

3.4.2. Calculation of Poly depletion capacitance

When the device is biased in the inversion region, due to the formation of space-charge regions both in the substrate and polysilicon gate near the interface, the bands bend as shown in Fig. 24 (substrate inverted, poly-gate depleted). To obtain accurate results, numerical analysis based on quantum mechanics [30] is used to calculate the space-quantization in the inversion layer and poly depletion effects. In [30], charge in the inversion layer of deep sub-micron devices is calculated using the self-consistent solution of 1D Poisson equation,

\[ \frac{d}{dz} \left[ \varepsilon(z) \frac{d\varphi}{dz} \right] = -q \left[ N_D^+ (z) - N_A^- (z) + p(z) - n(z) \right] \]  \hspace{1cm} (17)

and 1D Schrödinger equation

\[ \left[ -\frac{\hbar^2}{2m_i^\perp} \frac{d^2}{dz^2} + V_{\text{eff}} (z) \right] \psi_{ij} (z) = E_{ij} \psi_{ij} (z) \]  \hspace{1cm} (18)

In the above equations \( \varphi(z) \) is the electrostatic potential, \( \varepsilon(z) \) is the spatially dependent dielectric constant, \( N_D^+ (z) \) and \( N_A^- (z) \) are the ionized donor and acceptor concentrations, \( n(z) \) and \( p(z) \) are the electron and hole densities, \( V_{\text{eff}}(z) \) is the effective potential energy term, \( m_i^\perp \) is the effective mass normal to the semiconductor-oxide interface of the \( i \)-th valley and \( E_{ij} \) and \( \psi_{ij}(z) \) are the corresponding energy level and corresponding wave-function of the electron residing in the \( j \)-th sub-band from the \( i \)-th valley. The self-
consistent solution provides the inversion layer electron density appearing in the 1D Poisson equation by summing over all sub-bands, resulting in:

\begin{equation}
    n(z) = \sum_{i,j} g_i \frac{m_i^* k_B T}{\pi \hbar^2} \ln \left[ 1 + \exp \left( \frac{E_F - E_{ij}}{k_B T} \right) \right] \psi_{ij}^2(z) \tag{19}
\end{equation}

The charge density in the poly depletion is calculated by classical approach. For electrons, the charge density is given by:

\begin{equation}
    n(z) = N_c F_{1/2} \left[ \frac{E_F - E_C(z)}{k_B T} \right] \tag{20}
\end{equation}

where \( N_c \) is the effective density of states of the conduction band. For holes:

\begin{equation}
    p(z) = N_v F_{1/2} \left[ \frac{E_C(z) - E_G - E_F}{k_B T} \right] \tag{21}
\end{equation}

where \( N_v \) is the effective densities of states of the valence band and \( E_G \) is the semiconductor bandgap. Fermi-Dirac integrals, \( F_{1/2} \) are evaluated using analytical approximations as mentioned in [30]. Once the self-consistent results for the charge densities...
variation of the charge distribution as a function of the gate voltage is obtained, the total
gate capacitance is calculated by differentiating the total induced charge density in the
channel with respect to $V_G$.

The above effects need to be included for an accurate characterization of
depletion in poly region. However, since the gate model is added as a sub-circuit to the
intrinsic model, it has to be compact and simple, without involving complex
computations that will increase the simulation time. Due to the above reasons, the
depletion width is calculated with the help of a numerical self-consistent solution to the
1D Schrödinger and Poisson equation solver Schred [37]. Schred is programmed in
FORTRAN and considers poly-depletion while calculating total capacitance. The
methodology used in Schred to solve the Schrödinger and Poisson equations is shown in
Fig. 25. This process is continued until the carrier concentration, $n(x)$ and the conduction
well $E_c(x)$ simultaneously satisfy both the Schrödinger and Poisson equations.

Figure 25 Process flow to consistently solve the Schrödinger and Poisson's equations
The inputs to the program to compute the charge density in the gate electrode are the gate doping concentration, substrate doping and gate bias for both p and n type devices. Once the depletion width is computed in Schred, the depletion capacitance is calculated based on the physical dimensions of the device. The depletions widths for devices in the 180 nm, 110 nm and 90 nm process nodes are calculated. For simplification, two assumptions that are used in the calculation are uniform doping profiles for the gate and substrate regions. The gate doping values used (three technologies) were in the range of $0.5 \times 10^{20}/\text{cm}^3 - 3 \times 10^{20}/\text{cm}^3$. The oxide thickness values were in the range of 2-5 nm. The depletion widths calculated using Schred for different bias conditions for the 110 nm process NMOS transistor are shown in Table 1. The conduction band output for the same device, for a gate bias of $V_t + 0.6 \text{ V}$ obtained through Schred is shown in Fig. 27.

<table>
<thead>
<tr>
<th>Gate bias(V)</th>
<th>Depletion width (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.475</td>
</tr>
<tr>
<td>$V_t$</td>
<td>2.916</td>
</tr>
<tr>
<td>$V_t + 0.2$</td>
<td>3.068</td>
</tr>
<tr>
<td>$V_t + 0.4$</td>
<td>3.175</td>
</tr>
</tbody>
</table>

**Table 1 Poly depletion widths at various bias conditions**

In the figure, the substrate region is depicted by the positive x-axis while the oxide and poly-gate is the negative x-axis. It is difficult to notice the depletion effect in the figure due to its small value. Hence it is enhanced in the inset to show the depletion width clearly. The depletion widths as a function of gate bias $V_{gs}$ for devices in two different technologies, 110 nm and 90 nm is shown in Fig. 28a and Fig. 28b, respectively.
Figure 26 Variation of the poly-depletion width as a function of gate bias

Figure 27 Conduction band output from Schred with enhanced view of the depletion width $t_{pd}$
Once the depletion width is calculated, $C_{dep}$ values are then obtained using the expression:

$$C_{dep} = \frac{\varepsilon_{poly}}{t_{pd}} \times W \times L \times NF$$  \hspace{1cm} (22)$$

where $\varepsilon_{poly}$ is the permittivity of poly-silicon, $t_{pd}$ is the poly depletion width, $W$ is the
device width, $L$ is the length, and $NF$ is the number of fingers. The depletion capacitance calculated using the above expression for the depletion widths of the Table 1 is plotted in Fig. 29. This behavior of the depletion capacitance is consistent with the measured input impedance with respect to gate bias and is shown in Fig. 30.

![Figure 29 Calculated poly-depletion capacitance as a function of $V_{gs}$ for a device in 110 nm process](image)

![Figure 30 Real part of input impedance for various bias conditions in 1-20 GHz frequency range](image)
This proves that the gate depletion model conforms to the measured data and accurately predicts the input impedance behavior over three different technologies. Further verification of the model with the help of simulation results is presented in the next chapter.
CHAPTER 4
EXPERIMENTAL RESULTS

In the previous chapter, a new gate model including poly depletion effect has been developed for 1-20 GHz frequency range. The capacitance resulting due to PDE is calculated based on quantum mechanical analysis. The accuracy of the model in the desired frequency range is verified with the help of measured data in 180 nm, 110 nm and 90 nm technologies. Further modifications of the PDE based gate model to fit a wide frequency range of 200 M-20 GHz is presented. The accuracy of the new model is validated with the help of measured data.

4.1 Measurement and de-embedding

The measurement and de-embedding of the data was carried out by the high-speed modeling group at LSI Logic Corporation. A typical device layout (90 nm) and test structure (110 nm) under study are shown in Fig. 31 and Fig. 32, respectively.

![Figure 31 Typical layout of a test structure for a device in 90 nm process]
To ensure accuracy of measured data, standard calibration and de-embedding techniques are used. Calibration is handled with a standard short, open, load, through (SOLT) technique using an Impedance Standard Substrate (ISS) from Cascade Microtech. The de-embedding procedure is carried out using open, short and dummy test structures and includes the steps shown in Fig. 33 [38].

***Figure 32 Chip micrograph of a test structure fabricated in 110 nm technology***

To compare simulated results with measured data, the data was imported to the simulation tool as the input impedance parameter $Z_m$. The measured S-parameter data
was converted to $Z_m$ parameters using Spectre™. This was achieved with the help of Cadence’s intrinsic function that can be used as “port instance” and accepts the S-parameter information as input. This instance, upon simulation in Spectre™, generates the desired parameters (Z, Y, H etc.) from S-parameter input. The schematic setup that achieves this is shown in Fig. 34.

![Figure 34 Schematic setup for converting S-parameter data](image)

### 4.2. Poly depletion gate model

For simulations, the poly depletion (PD) model can be added as an external RC network to the device’s intrinsic gate as shown in Fig. 35. The intrinsic BSIM4 gate resistance model was turned “ON” by setting the BSIM4 model parameter RGATEMOD to 3 to model the gate impedance for high frequency range. An additional resistance, $R_h$, is added to account for the head resistance associated with the contact landing pads of the

![Figure 35 BSIM4 gate resistance model with external gate impedance model](image)
unit gate fingers. While the depletion capacitance is calculated as described in the previous chapter, all the other parameters in the external impedance model are obtained empirically from the measured data. The figure also shows two parasitic capacitances indicated by dotted lines.

The model was simulated in the 110 nm process and the simulation setup is shown in Fig. 36. In the figure, the capacitors $C_{dc}$ are the DC blocking capacitors and the chokes $L_C$ have high inductance values to avoid shorting of the device input and output ports.

![Figure 36 Setup of the PD model with BSIM4 used in the simulations](image)

The PD model with BSIM4 is simulated for different device widths, lengths, bias conditions in the 1-20 GHz range. The simulation results are presented below with $S_{11}$ as a measure of input impedance. Fig. 37 shows parts of the Smith charts of $S_{11}$, with the fit between simulated and measured data across various bias conditions for a device with a single finger dimension of 20 $\mu$m/0.11 $\mu$m. Note that the model deviates from the measured data for $V_{gs} = 0$ and $V_{gs} = V_{th}$. This can be seen from Fig. 37 (a) and (b), where symbols representing measured (triangles) and simulated (squares) results do not overlap. However, for the saturation regime in Fig. 37, (c) and (d), the model agrees very well
with the measured data. Below the cutoff conditions and at $V_{dd}$, the value of $C_{dep}$ required to fit the measured data is less than the value predicted by equation (22). At these extreme bias conditions, the simple external gate model does not fully capture all the complex effects of gate impedance. However, for most RF applications, devices are biased in the active region and the model is applicable and relevant for these devices.

![Graphs showing measured and simulated curves for various bias conditions](image)

**Figure 37 Measured and simulated curves for various bias conditions**

The measured and simulated results for devices biased at $V_{th} + 0.2$ V (active region) with different widths and lengths are shown in Fig. 38 and Fig. 39, respectively, where the symbols represent the measured data. For both cases, the model is in good agreement
with the measured data. In essence the model captures the input impedance behavior very well in the active region of operation over different device geometries.

Figure 38 Simulated and measured input impedance for different widths at $V_{gs} = V_{th} + 0.2$

Figure 39 Simulated and measured input impedance for different device lengths

4.3. Comparison of the PD model with other models

In Chapter 3, prior work in modeling gate for high frequencies was discussed. In this section, the performance of these models is evaluated, especially BSIM3v3, BSIM4 with respective external and intrinsic gate models. The performance of the high
frequency gate model of [19] shown in Fig. 13 and the PD gate model is also verified by comparing the simulation results with the measured data. These models are added externally to the gate terminal of the BSIM3v3 and BSIM4 models. The simulations were carried out in the 1 G – 20 GHz frequency range for devices in 110 nm and 180 nm processes in all regions of operation. The real part of the $Z_m$ parameter (that best describes the input impedance) is used for comparisons.

### 4.3.1. BSIM3v3 model

The limitations of BSIM3v3 model to accurately predict the input impedance of a MOSFET are verified in this section. To achieve this, a simulation setup was devised to compare the performance of the BSIM3v3 model with and without the external high frequency gate models. The simulation setup with the external gate model is shown in Fig. 40, which is simulated for devices in 110 nm technology in the 1 G – 20 GHz frequency range.

![Figure 40 Simulation setup of PD model with BSIM3v3](image)

The simulation results of the three models viz. BSIM3v3, PD and the high frequency gate models are compared in Fig. 41. The symbols in Fig. 41 represent the measured data, the dashed line represents the bare BSIM3v3 model, the dotted line represents BSIM3v3 with the approximate gate impedance model of [19] and the solid
line represents the BSIM3v3 with the PD gate model. From the figure, it can be seen that both the intrinsic BSIM3v3 and BSIM3v3 with the gate model of [19] do not predict the frequency dependence of input impedance very well. However the semi-empirical PD gate model tracks the measured data very well across the whole frequency range and hence is more accurate. The parallel combination of $R_{dep}$ and $C_{dep}$ in this model explains the strong frequency behavior of the input impedance in conjunction with the head resistance, $R_H$ and the gate electrode resistance and capacitance, $R_g$ and $C_g$, respectively.

![Comparison of BSIM3v3 with and without gate impedance models](image)

**Figure 41 Comparison of BSIM3v3 with and without gate impedance models**

### 4.3.2. PD model with BSIM4.2.1

The external gate model for BSIM4 is simpler than BSIM3v3 and has a smaller number of parameters. This is due to the inclusion of the gate electrode resistance and gate-to-source and gate-to-drain capacitors (Fig. 11) in BSIM4’s gate model. This will eliminate the parameters $R_g$ and $C_g$ of Fig. 40 from the external gate impedance model for BSIM4. The simulation setup used is shown in Fig. 42 along with the external PD gate
model. Intrinsic BSIM4 gate resistance model was turned “ON” by setting the model parameter RGATEMOD to 3.

![External PD gate model diagram](image)

**Figure 42 Test setup to verify the PD model with BSIM4.2.1**

The comparison of BSIM4.2.1 model with and without the external gate model and the measured data is shown in Fig. 43. The dashed line in Fig. 43 is for BSIM4.2.1 model with the intrinsic gate resistance model, the solid line is for BSIM4.2.1 with the intrinsic and external gate impedance models and the symbols represent measured data. It can be seen from the Fig. 43 that BSIM4.2.1 with its gate resistance model turned on fits the data well above 15 GHz. This is due to the BSIM4 gate resistance model parameters $R_{ij}$, $C_{gd0}$ and $C_{gs0}$. Due to relatively small values, these parameters start to contribute to the overall impedance at high frequencies. Hence BSIM4 model is useful only at the high-end of the frequency range, whereas the external gate model works very well with both BSIM3v3 and BSIM4 models in the desired frequency range. The performance of the gate models is analyzed with respect to bias conditions, by simulating the PD model with BSIM4.2.1 for devices in 180 nm and 110 nm technologies. The comparison of the simulated results with the measured data for above mentioned technologies is shown in Fig. 44. The results show that this model works well at the high frequency end but deviates from the measured data at lower frequencies. Finally, to assess the adaptability
of the PD gate model, it was simulated with BSIM3v3 and BSIM4.2.1 models. The models were then simulated using a device in 110 nm technology with gate bias, $V_{gs} = V_{th} + 0.1$ V and drain bias, $V_{ds} = V_{gs} + 0.2$ V. The curves comparing the simulation results with the measured data for the above conditions are shown in Fig. 45. From the results it can be seen that the external PD gate model can be used with both BSIM3v3 and BSIM4.2.1 models.

![Graph showing comparison of simulation and measured data](image1)

**Figure 43 Performance of BSIM4.2.1 with and without gate PD model**

![Graph showing simulation and measurement results](image2)

**Figure 44 Simulation and measurement results for different $V_{gs}$ in a) 180 nm and b) 110 nm technologies**
Figure 45 Comparison of BSIM3v3 and BSIM4.2.1 model with the external gate PD model

The above results prove that the PD gate model captures the frequency dependency effectively compared to other models in the 1-20 GHz frequency range. The adaptability of this model with BSIM3v3 and BSIM4.2.1 provides flexibility.

4.4. Universal MOSFET gate model for wide frequency range

The scaling of CMOS technology to 100 nm and below and the endless pursuit of higher operating frequencies drives the need to accurately model effects such as gate leakage and the deterioration of transport characteristics that dominate at those feature sizes and frequencies. Current modeling techniques are frequency limited and require different models for different frequency ranges in order to achieve accuracy goals. This section presents an extension of the model in Chapter 3 for device gate that can operate for a wide frequency range making it frequency independent.

4.4.1. Need for an unified wide-frequency gate model

It has been shown in [25] that as the input data run-length and hence frequency of the incoming data to a CML buffer varies, the output degrades due to inaccurate gate
impedance modeling. The gate impedance in [25] was modeled at a particular frequency but did not work accurately at other frequencies. For production design work, it is essential to have a gate model that can predict the performance of a circuit under any frequency variations which would occur under normal operating conditions. To meet these requirements the PD gate model developed in the previous chapter is further modified to be applicable for a frequency range of 200 M-20 GHz and is explained in this section. The new empirical model is simulator-friendly, compact, and relatively portable across technology nodes. It also performs accurately over 200 M-20 GHz at different bias conditions and widths and has been verified by measured data in three technology nodes. The model and model parameter behavior are consistent across technology nodes thereby enabling re-usability and portability.

4.4.2. Wide frequency behavior of PD gate model

Before proceeding to develop a wide-frequency model, the performance of the PDE based model developed for 1-20 GHz is studied for a wider frequency range of 200 M-20 GHz. The external gate impedance model of Section 4.2 is simulated for the required 200 M-20 GHz range. The simulation result for a device in 110 nm technology is compared with measured data is shown in Fig. 46. It can be observed, as already shown in Section 4.2 that this model works very well in the high frequency (1-20 GHz) range. However, it deviates from the measured data (symbols) at lower frequencies (200 M-1 GHz) as shown by the curves of input reflection coefficient, $S_{11}$, in Fig. 46. Hence the PD gate model is modified to fit the entire frequency range making it one of its kind that can operate for such a wide range of frequencies.
4.4.3. **Universal gate impedance model**

To further improve the performance of the PD model below 1 GHz, the model was modified by splitting it into two parts. In the improved model, the parallel RC combination now models gate impedance at low frequencies (below 800 MHz-1 GHz) while the gate impedance at higher frequencies is represented with a shunt capacitance shown as $C_{HF}$ in Fig. 47 [39]. The values of all parameters in the improved model in Fig. 47 are derived empirically from the measured data. This model predicts the gate impedance accurately from 200 MHz to 20 GHz and was verified by comparing simulation results and measured data in 110 nm technology. In addition to supporting a wide frequency range of operation, the model is simple enough to be incorporated in existing simulation setups. The new model corrects for both the real and imaginary parts of the input impedance with very good accuracy. However, to enable portability across technology nodes, the physical effects behind such behavior must be understood. The following section details the investigation and analysis of these effects.
4.4.4. Effect of $C_{HF}$ on the input impedance

The shunt capacitance, $C_{HF}$, which models the high frequency behavior, was also empirically derived from the measured data. The inclusion of $C_{HF}$ between gate and source terminals minimally alters the gate-source impedance [39]. For the devices under study it has been observed that $C_{HF}$ values range from 5f-100f F depending on the bias and device dimensions. The comparison of $C_{HF}$ with the depletion capacitance $C_{dep,HF}$ of the wide frequency range model is shown in Fig. 48 for a device in 110 nm technology. Being in parallel with a comparatively large capacitance $C_{gs}$, $C_{HF}$ nominally impacts the gate-source impedance. To show the effect of $C_{HF}$ on the gate-source impedance, $Z_{11}$ plots of a test device in 110 nm technology are shown in Fig. 49. The plots show the real (Fig. 49a) and imaginary (Fig. 49b) parts of the $Z_{11}$ parameter that best represents the gate-source impedance, with and without $C_{HF}$. The average variations in $Z_m$ are 14% in the real part and 4% in the imaginary part. The new model was verified using simulations and measurement results, which are presented in the next section. The portability was validated by applying the gate impedance model to devices in 180 nm, 110 nm, and 90 nm technologies. However, determination of the physical phenomena that result in the
high frequency capacitive effect ($C_{HF}$) at the gate needs to be investigated. The possible candidates are overlap, fringing capacitances and the head capacitance of the gate electrode that cannot be ignored at these frequencies.

![Graph showing comparison of $C_{HF}$ with $C_{dep,HF}$](image)

**Figure 48** Comparison of $C_{HF}$ with $C_{dep,HF}$ used in for a device in 110 nm process

![Graphs showing variations in real and imaginary parts of input impedance](image)

**Figure 49** Variations in the a) real and b) imaginary parts of the input impedance due to $C_{HF}$

### 4.4.5. Experimental Results: Wide frequency gate model

In order to validate the new gate model, simulations were used to generate input reflection coefficient ($S_{11}$) data over different widths and bias conditions for the 110 nm and 90 nm technologies [39]. BSIM4.2.1 models were used in the simulations with
BSIM4's intrinsic gate resistance model set to “RGATEMOD=3”. The simulated results were then compared with the measured input reflection coefficient data to determine the performance of the model. For the 110 nm technology, the frequency range spanned from 200 MHz to 20 GHz. However, for the model analysis in 90 nm and 180 nm technologies, the validation frequency ranged from 1-20 GHz. Future work includes collection and analysis of the data below 1 GHz in these technologies.

4.4.5.1. 110 nm Technology

The PD capacitance was calculated based on uniform poly and substrate doping and these assumptions may induce discrepancies between the simulated and measured data. To verify the extent of errors induced due to the above assumptions, two simulation setups were used one with the intrinsic value obtained from calculations, called “empirical model”, and the other “PD simulated model” which uses a $C_{dep}$ value that better fits the measured data. It was seen that major difference between these two models occurs in sub-threshold region which is further discussed below (Fig. 53 & Fig. 54a).

1. Width dependence of the model

The simulated and measured input reflection coefficient ($S_{11}$) curves for an NMOS transistor for different widths from 200 MHz – 20 GHz are shown on a Smith chart in Fig. 50. All the devices in Fig. 50 were biased in saturation region with gate voltage, $V_{gs} = V_{th} + 0.2$ V and $V_{ds} = V_{gs} + 0.2$ V, where $V_{th}$ is the threshold voltage. The symbols in Fig. 50 represent the measured data, solid line represents the “simulated PD model”, and the dashed line represents the “empirical model”. From these figures, it can be seen that the model fits the measured data very well over the entire frequency range for each of the different widths. The variation of the model parameters $C_{dep}$, $R_{dep}$, and $C_{HF}$ for three
device widths is shown in Fig. 51. In Fig. 51a, the depletion capacitance calculated from self-consistent 1-D Schrödinger and Poisson equations is represented by the dashed line (PD based model), and the solid line represents the values used in the simulations (empirical).

![Smith Chart](image)

**Figure 50** a) Gate model performance for device width \( W = 20 \, \mu m \); b) for \( W = 5 \, \mu m \); and c) for \( W = 2.5 \, \mu m \) in 110 nm technology
The discrepancy between these two values can be attributed to the approximations made while calculating the PD capacitance, such as uniform poly doping in the gate, and the inadequacy of the solver, Schred to capture the quantum effects efficiently in short channel devices. Even with these error sources, the difference between the calculated and simulated capacitance values is within 10%. The PD and the HF shunt capacitance values increase with device width while the resistance in the depletion region decreases with width, as expected.
II. Bias dependence of the model

To verify the applicability of the model to devices operating over various bias conditions, simulations were executed for $V_{gs} = 0 \text{ V}, V_{th}, V_{th} + 0.2, \text{ and } V_{th} + 0.4$. The simulation results and the measured data for various gate bias voltages are shown in Fig. 52. In Fig. 52, the representation of model performance at lower frequencies is enhanced by plotting the results in terms of the real and imaginary parts of the device input admittance. The symbols in Fig. 52 represent the measured data, the solid line represents the PD based simulated results, and the dashed line represents the simulated empirical model. As seen from Fig. 52, the model fits the measured data accurately over the entire frequency range of interest for bias conditions greater than threshold voltage. However, in sub-threshold region the (real part) model does not accurately capture the behavior at lower frequencies. This shows a strong bias dependency of the model and is consistent with the $C_{dep}$ behavior with respect to bias voltage as described in Fig. 53a. Variation of the model parameters with the gate bias is shown in Fig. 53.

The curves in Fig. 53a represent the variation in calculated (dashed line) and simulated (solid line) PD capacitance vs. gate voltage. The curve in Fig. 53b shows the variation of the depletion region resistance with respect to the gate bias. Fig. 53c depicts the variation in the HF shunt capacitance over gate voltage. It is seen from Fig. 53a that the $C_{dep}$ values (calculated and simulated) are in good agreement in the saturation region (within 10%). However, in the sub-threshold region the calculated value is much higher than the simulated value. In other words $C_{dep}$ is over-estimated in this region of operation. This can be attributed to the assumptions used in the calculation of $C_{dep}$. This will impact correlation for device performance in the sub-threshold region at low frequencies.
Figure 52 Measured and simulated results at different gate bias for 200 M - 20 GHz frequency range in 110 nm technology
However, for most applications, especially those requiring high frequency operation, devices will be biased in the saturation or active regions where the model is accurate and therefore applicable. Hence, this model is accurate for predicting the input impedance behavior of MOS devices over a wide frequency range and bias range, except the sub-threshold region. As the gate bias increases, the depletion region increases making the carrier density in that region lower and thereby increasing the resistance in the poly-depletion region as shown in Fig. 53b.
Figure 2.4(a) (c) shows the input impedance model performance for the device with frequency range of 1-20 GHz over various device technologies. Simulation results for a frequency range of 1-20 GHz over various device technologies verify the portability of the model. The model was applied to devices in 90nm technology.
Figure 55 (a) Gate Impedance model parameters ($C_g$, $R_g$) and (c) CHF, $V_s$, device width for 90 nm technology.

Hence, the new Pd gate Impedance model is universally applicable across different devices. 10% of the calculated values providing a consistent behavior across different processes. Extracted values of $C_g$ fall within this deviation is less than 5%. The model parameters also show the same qualitative at some frequency range (HP in Fig. 55b and 55c and mid frequency range in Figs. 54, 55a).

From the curves it is seen that although the model devalues slightly from the data.
4.4.6. PD gate model with BSIM3v3

The performance of the gate impedance model with BSIM4 was presented in 110 nm and 90 nm technology. This section presents the application of the model with BSIM3v3 model in 180 nm technology for 1-20 GHz range. BSIM4 has an intrinsic gate impedance model which consists of the lumped gate resistance as well as a non-linear resistance with the gate capacitances across it [17]. The gate impedance model has to be modified to accommodate for the lack of intrinsic gate impedance model in BSIM3v3. The simulation setup of the modified gate impedance model is shown in Fig. 56 [39]. The resistance $R_{g3}$, and the capacitances $C_{gs3}$ and $C_{gd3}$ are empirically derived from the measured data and the external gate impedance parameters are obtained as presented in Section 4.4.2.

![External gate model](image)

**Figure 56 Improved gate impedance model for BSIM3v3**

The setup was simulated for 180 nm technology for 1-20 GHz frequency range and different widths with the measured data is shown by input reflection $S_{11}$ curves in Fig. 57. The symbols represent measured data and the solid line represents the simulated input impedance. Fig. 57 shows the simulated results for a device with $W=22.5 \, \mu m$, Fig. 57b for $W=5 \, \mu m$ and Fig. 57c for $W=2.5 \, \mu m$. As seen from the simulated results,
the model fits the measured data very well for the entire frequency range for all device widths. It has been successfully demonstrated that the model is accurate and applicable not only with latest versions of BSIM such as BSIM4 but also with older versions such as BSIM3v3. Hence the proposed gate impedance model is universal, with wide frequency range, versatile and backward compatible making it an extremely useful tool for designers.

![Smith Chart](image)

**Figure 57** a) Modified gate impedance model performance for device width $W=22.5 \, \mu\text{m}$; b) for $W=5 \, \mu\text{m}$; and c) for $W=2.5 \, \mu\text{m}$ in 180 nm technology
CHAPTER 5
APPLICATIONS

The high frequency gate models including PD and wide frequency-range developed in this work were verified with the help of measured data in the previous chapter. The PD model is utilized in this Chapter, to study its importance and impact in evaluating high frequency behavior both at the device and circuit level. Device level application is presented in Section 5.1 by using the PD based gate model to verify device model extraction techniques in the GHz frequency range. At the circuit level, the model is used in high frequency circuit applications such as CML buffers and ring oscillators, and the impact of the model on their performance is studied in Section 5.2. It will be shown that the model correctly evaluates the degradation in the performance of a CML buffer as the input data frequency changes. The inaccuracies in jitter calculation of a ring oscillator are highlighted by comparing the ring oscillator outputs with and without the PD gate model.

5.1. Gate input impedance model for different device configurations

Extraction plays an important role in the performance of device models, especially in the high frequency regime. The present day extraction techniques mostly use a grounded-source or common source (CS) device configuration. For high frequency applications, sub-circuit components are extracted for a given process and device dimensions and are added to the intrinsic model [40]. The sub-circuit parameter extraction is usually carried out in CS configuration where the source of the device is grounded. Due to the complexity of the sub-circuit models, most of the parameters are
approximated as a single element. It is not established if this approximation and the parameters extracted using the CS configuration are valid for other configurations. There is no work reported in literature which addresses whether the models extracted for grounded-source configurations are valid for non-grounded-source configurations. In many circuit applications, MOS devices are stacked, and are not necessarily in CS configuration. Using the models extracted from the grounded source devices in other configurations may lead to discrepancies in predicting the circuit performance. Hence it is important to investigate these models and study their validity for other configurations.

The need for enhancement in present day models for transistors other than grounded-source configuration is presented based in [41]. The applicability of the models extracted using CS for other device configuration such as common drain (CD) is verified. This is carried out by studying the gate impedance behavior for both CS and CD configurations. An existing PD based gate model was employed to fit both the configurations, resulting in a single gate model for devices in CS and CD configurations. This model was then applied to BSIM3v3 for 1 to 20 GHz frequency range and later extended to BSIM4 for both CS and CD configurations. The models for BSIM3v3 and BSIM4 are verified using simulations and compared with measured data.

5.1.1. CS and CD test structures

The MOSFET is a four terminal device, but normally measurements are made using only three terminals, which could result in discrepancies of the extracted model. In order to make sure that the components of the high frequency model are connected to the correct terminals, both CS and CD structures are studied. This is achieved by comparing measured data from the CS and CD configurations. Common source and common drain
N-channel RF-MOSFETS with different widths, lengths and number of fingers (NF) were fabricated using LSI Logic’s 110 nm process technology. Test structures for the same device dimensions in CS and CD configurations were developed and fabricated.

Test structure schematics used for CS and CD configurations are as shown in Fig. 58. To characterize the devices, S-parameter measurements were made from 1 to 20 GHz at different bias conditions. For measurements, $V_{gs}$ was varied from 0 to $V_{dd}$ in increments of 0.2 V and $V_{ds}$ was varied as $V_{gs} + 0.2$ V for both CS and CD structures. A comparison of the measured data values of the input impedance for CS and CD test structures with dimensions, $W = 20 \, \mu\text{m}$, $L = 0.13 \, \mu\text{m}$, biased at $V_{gs} = 1 \, \text{V}$ and $V_{ds} = V_{dd}$ is shown in Fig. 59. The symbol “+” represents the measured CD input impedance and the solid line represents the measured CS input impedance.

![Test structures for a) CS and b) CD configurations](image)

**Figure 58 Test structures for a) CS and b) CD configurations [41]**

It can be seen from the figure that there is a difference in the input impedance behavior with respect to frequency for the two configurations. This establishes the need to study the behavior of the MOS models extracted from CS structures, as they may not predict the behavior of the devices in other configurations accurately.
5.1.2. BSIM3v3 Model

The input impedance behavior of CS and CD devices with BSIM3v3 model is presented in this section. The modifications in the gate impedance model required to make it compatible for both CS and CD configurations is discussed. Measured data from the test structures in the frequency range of 1-20 GHz are used to verify the model performance. These structures are simulated with the PD gate model and the simulation results are compared with the measured data. This study uses the real part of the input impedance denoted by Spectre™ parameter \( Z_m \) as the metric.

I. CS Configuration

The PD gate model that better predicts the impedance behavior over different dimensions, bias and technology nodes as explained in the previous chapters is used for this study. The gate impedance model was enhanced with the inclusion of a capacitance \( C_{gs} \), to model the input impedance for the CS configuration with BSIM3v3. The capacitance component was included in order to get an accurate high frequency performance (up to 20 GHz) as BSIM3v3 lacks a high frequency gate model. The
simulation setup for the gate impedance model is shown in Fig. 60. This model is used to empirically fit the measured data for the CS case.

![Figure 60](image)

**Figure 60 BSIM3v3 with the external gate impedance model for CS [41]**

The CS setup as shown in Fig. 60 was simulated and the comparison between the measured (symbols) and simulated (solid line) real part of the input impedance is shown in Fig. 61. It can be seen from the figure that the simulated curve tracks the measured data very well for the entire frequency range. This proves that the gate impedance model shown in Fig. 60 is ideal for CS device configurations. The applicability of this model for CD configuration is discussed in the next section.

![Figure 61](image)

**Figure 61 CS Measured and simulated input impedance for BSIM3v3 [41]**
II. CD Configuration

When the gate model was applied to a device in CD configuration, it was observed that there was a significant discrepancy between the measured and simulation data. It was shown with the help of measured data that the simple gate impedance model does not predict the input impedance performance for CD configuration. This is evident from the simulation results in Fig. 63 where the dashed line represents the CD input impedance curve with the CS gate impedance model, while the symbols represent the measured data. These results stress the need to investigate improvements in the existing model, which would also describe the behavior of CD configuration.

For this purpose, a capacitance $C_{gs2}$ as shown in Fig. 62 (within the dotted lines) was placed between the gate and the source in the gate impedance model for the CD case. This modification to the model resulted in achieving a match between the simulated and measured input impedance data shown in Fig. 63. The CD input impedance tracks the measured data after the inclusion of the additional capacitance as shown by the solid line in Fig. 63. This curve fits the measured values (symbols) very well, especially in the high frequency regime of the curve (above 5 GHz). It was observed that the presence of $C_{gs2}$ in the gate impedance model did not affect the input impedance behavior for the CS.

![Figure 62 BSIM3v3 with an additional capacitance for CD [41]]
However, in the CD configuration, as $C_{gs2}$ value is increased, the curve moves higher and closer to the measured data. Also, in the case of CS configuration, the placement of $C_{gs2}$ between the gate and source terminals is same as from gate to ground, since the source terminal is connected to the ground. But for the CD configuration, the source is not connected to ground. Hence it was necessary to verify whether the capacitance $C_{gs2}$ was between gate and source terminals or gate and ground terminals or whether it was split between gate-to-source and gate-to-ground. When the capacitance $C_{gs2}$ was split between gate-to-source and gate-to-ground, the simulated data moved away from the measured data. It was confirmed from simulation that all the capacitance was between gate and source terminals only.

![Figure 63 CD Measured and simulated input impedance with BSIM3v3 [41]](image)

The gate impedance model was verified over different devices by varying the number of gate fingers (NF), for both CS & CD. Measured and simulated input impedances for CS are shown in Fig. 64. Measured data is shown by symbols in the figure and the corresponding simulated results are shown by solid line for NF=8 and by dotted line for NF=4. As seen from the figure, for both the cases the simulated curves fit the data very well for the entire frequency range, thus verifying the new gate model.
The simulated and measured input impedance for CD with the modified gate input impedance are shown in Fig. 65. From the results it can be seen that for NF=4 the fit is better at lower frequencies (below 5GHz) compared to NF=8.

5.1.3. BSIM4 model

The performance of the gate model of Fig. 60 with BSIM4 for CS and CD cases is presented here. The test structures were simulated using BSIM4 model to verify the gate impedance behavior for both CS and CD. This section describes required changes to the
gate impedance model that makes it compatible with BSIM4 for both CS and CD configurations.

I. **CS Configuration**

The simple gate model consisting of only parameters \( R_p, R_s, C_p \) and \( C_{gs} \) with BSIM4 tracked the measured data very well for CS devices. This is due to the high frequency gate resistance model of BSIM4 that provides the gate-source capacitance (that was used externally to fit the measured data for CS with BSIM3v3 component \( C_{gs2} \) in Fig. 62). Hence there was no modification necessary in the gate impedance model for CS. The fit between the simulated (solid line) and measured (symbols) data is shown in Fig. 66. It can be seen from the figure that the model and measured data match very well for the entire frequency range.

![Figure 66 CS Simulated and measured input impedance for BSIM4 [41]](image)

II. **CD Configuration**

From the results observed for CS with BSIM4 model, it was expected that the model could be used for CD without modifications. But it was found that for CD, the simulations results did not match the measured data, due to the need for higher series resistance in the gate impedance model. Unlike BSIM4, intrinsic BSIM3 model does not
have provision for adding a series resistance to the gate impedance model and hence the correct value was used in the enhanced model. However, BSIM4 accounts for part of the gate resistance through the intrinsic gate resistance model, so another series resistance was needed to match the measured data. The simulation results in Fig. 67 show that the gate impedance model (dashed line) did not track the measured results (symbols).

![Figure 67 CD Simulated and measured input impedance for BSIM4 [41]](image)

However in this case, due to the provision of $C_{gs}$ in BSIM4, the frequency dependency of the curve is maintained but the magnitudes of the measured and simulated data differ. Addition of series resistance to the gate impedance model rectifies this difference and the simulated input impedance (solid line) after the modification fits the measured data very well. It can also be noticed that, even for BSIM4 model, the simulated results fit the measured data accurately in the high frequency regime. Hence in the case of BSIM4, the simple gate impedance model works for both CS and CD. However, CD requires more series resistance in the model when compared to CS.

It is established using the measured data that the input impedance of a device when used in CS and CD configurations is different. This initiated the need to study the extraction of MOS device models, as most of the models are extracted using grounded-
source or CS configurations. The gate impedance behavior of CS and CD devices show that for BSIM3v3, an additional capacitive component between gate and source is required to match the measured data. For BSIM4, the provision of gate resistance eliminates the use of additional capacitance for both CS and CD. However, BSIM4 gate impedance model for CD requires higher series resistance to model the gate impedance.

5.2. Effect of input impedance on circuit performance

The input impedance studies so far included the impact of device dimensions and bias conditions on input impedance. It was shown in the previous chapters that the present day models have no mechanisms to accurately predict the high frequency behavior. It was also seen that present day models do not perform over an entire frequency range. These drawbacks can induce errors when the models are used in circuit applications operating at high or varied frequencies. The performance of the gate impedance models were evaluated comparing the simulation results of different models with measured data. The models extracted using CS are validated for CD configuration with the help of BSIM3v3 and BSIM4.2.1 models. All of these studies were carried out at the device level and the performance of these models in circuit environment needs to be studied. Impact of the models on circuit applications is studied in this section, with the help of the gate impedance model and a current mode logic buffer as a test circuit.

5.2.1. Current mode logic: Brief introduction

Current Mode Logic (CML) has been the design methodology of choice for high-speed and high-performance broadband communication systems such as synchronous optical network (SONET) [42-44]. The major advantages of CML over static CMOS logic are the speed of operation due to low voltage swings and low power consumption at
GHz frequencies. These advantages enable the use of CML circuits for applications at GHz. There is a considerable amount of work reported in the literature on development of CML circuit topologies. A design methodology for a chain of tapered CML buffers and an ultra high-speed regenerative latch was presented in [45]. Design for optimum gain and bandwidth for CML gates and a relation between circuit properties and deterministic jitter for broadband applications have been obtained in [46]. Adaptable MOS Current Mode Logic (MCML) gates were developed in [47] where adaptation was achieved by using a real-time bias control loop to control the drive current and voltage swing of the logic gates.

Various techniques have been reported to optimize CML design based on several performance parameters [48-50]. However, there is not much work reported either to design or to optimize the performance of CML based on input data run-time variations. The CML circuits employed in any application are usually designed to operate at one particular frequency. But as the data run-length or the number of consecutive ones and zeros in the input data-stream change, the frequency of operation varies. For example, an 8-bit input data consisting of a stream of 01010101 bits transmitting at a rate of 5 Gbits/sec, changes to a 2.5 Gbits/sec data stream if the incoming input bit stream is 00110011. This situation is also possible in any high-speed digital application.

This section presents a methodology to estimate the performance of CML circuits for different input data run-lengths. The effect of gate impedance is assessed using the PD model that can precisely predict the frequency dependence of input impedance of a MOSFET at GHz frequencies. A typical CML buffer operating at 16 GHz is used as a test circuit. Simulations are performed on the CML buffer and a ring oscillator to show
the performance degradation due to inaccurate gate impedance models, along with the run-length variation.

5.2.2. CML buffer

The CML buffer/inverter considered for this study is shown in Fig. 68, which is a basic building block for many CML circuits. The basic principle of operation of the buffer is to switch the constant current $I_s$ between the two output branches. When the transistor $M_1$ is ON, the output is given by:

$$V_{on} = V_{DD} \text{ and } V_{op} = V_{DD} - \Delta V, \text{ where } \Delta V = I_s \times R$$

Similarly when $M_2$ is ON,

$$V_{on} = V_{DD} - \Delta V \text{ and } V_{op} = V_{dd}, \text{ thus providing the inverting action.}$$

5.2.3. Data run-length error

The input data to the buffer is a stream of bits, at a particular run-length or frequency. Fig. 69 shows an example of frequency variation due to run-length. The 8-bit continuous stream of 1’s and 0’s, for example, translates to 4 GHz frequency, at which a circuit will be designed to operate. But as the input data stream varies, the frequency of the input data changes as shown in Fig. 69.
This will cause anomalies in the circuit operation as the circuit is designed to operate at a given frequency (4 GHz) but the effective frequency of input data changes to 8 GHz or 16 GHz. The high frequency BSIM4.2.1 model along with the gate impedance will be utilized to predict the performance of the buffer shown in Fig. 68 for various run-lengths. Different run-lengths (RL) used for simulations are shown in Table 2. The simulation results are discussed in the next section.

![Figure 69 Frequency variation with run-length](image)

**5.2.4. CML buffer simulation results**

The circuit under consideration, CML buffer, was simulated for 110 nm technology using different MOS models to compare their performance capabilities. The models used for the simulations were BSIM3v3, BSIM3v3 with the PD gate model of this work, BSIM4.2.1 and BSIM4.2.1 with the PD gate model. The simulation results comprise of the voltages at the two output branches of the buffer — i.e., $V_{on}$ and $V_{op}$ and are shown in Fig. 70 for various models and run-length/frequency. In Fig. 70 the square symbols represent the bare BSIM3v3 model, the dashed line represents bare BSIM4.2.1 model, the triangle symbols represent the BSIM3v3 with the gate impedance and the
solid line represents the BSIM4.2.1 with the gate impedance. From Fig. 70 it can be seen that for 16 GHz operation all the models predict the same output. But as the run-length varies the outputs of different models differ from each other.

<table>
<thead>
<tr>
<th>Run-length (RL)</th>
<th>Number of 1's</th>
<th>Number of 0's</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>32</td>
<td>32</td>
<td>500 M</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>16</td>
<td>1 G</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>8</td>
<td>2 G</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>4 G</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>8 G</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>16 G</td>
</tr>
</tbody>
</table>

Table 2 Run-length used in simulations

The BSIM3v3 and BSIM4.2.1 models with the gate impedance deviate from the bare BSIM3v3 and BSIM4.2.1 model. This is very apparent for 500 MHz case. The run-length variations considered so far are symmetrical in nature. To simulate the real scenario a set of 16-bit asymmetrical data was applied to the buffer. The differential output of the buffer for a case of 10 consecutive logic 1’s followed by 6 logic 0’s is shown in Fig. 71a, and for 15 consecutive logic 0’s followed by a logic 1 is shown in Fig. 71b. The results show that the output deteriorates by 40% for the latter case. Hence it can be seen that by using the accurate models for the transistors, the CML circuit performance can be predicted precisely for the worst-case scenario.
Figure 70 Buffer outputs using various models at a) 16 GHz; b) 4 GHz; and c) 500 MHz

Figure 71 Buffer outputs for asymmetrical input data
5.3. Ring oscillator application

Ring oscillators widely find applications in phase locked loops (PLLs) for clock and data recovery, on-chip clock distribution, frequency synthesis circuits and applications in many digital and communication ICs [51-53]. A simple CMOS ring oscillator can be built by connecting an odd number of inverters in a ring, as shown in Fig. 72. Due to better noise performance by common mode rejection of power and substrate coupled noises, the inverters in Fig. 72 are implemented using the CML buffers discussed in Section 5.2. The ring oscillator is designed to operate at 3.125 GHz and simulated in 110 nm technology using BSIM4, with and without the PD gate model. The jitter in the ring oscillator is used as a performance metric, and the jitter is compared with and without the PD gate model.

![Figure 72 A simple ring oscillator built using odd number of inverter stages](image)

5.3.1. Jitter in ring oscillators

In an ideal ring oscillator, the spacing between the transitions of the oscillator is constant, but in practical implementations this spacing varies. This variation in the output transitions shown in Fig. 72 is known as jitter and is mainly affected by device noise and switching noise coupled through power supply and substrate [54] [55]. Consider the output of the ring oscillator V(out) with zero-crossing from minus to plus at $t_n$. The $n^{th}$ time period is then given as, $T_n = t_{n+1} - t_n$. Jitter can be defined in terms of the variation
\( \Delta T_n \), given by \( \Delta T_n = T_n - \bar{T} \), the deviation of \( T_n \) from \( \bar{T} \), the mean period. There are various ways to quantify jitter in ring oscillators and they are defined as follows [55].

**Figure 73 Variations in the zero-crossing of the output from an ideal reference due to jitter**

The long term jitter or absolute jitter is defined as:

\[
\Delta T_{abs}(N) = \sum_{n=1}^{N} \Delta T_n
\]  

(23)

Absolute jitter is the measure of long term variations in the jitter and is very useful in quantifying jitter for PLL applications. However a better measure of jitter in ring oscillator applications is cycle jitter defined as the *rms* value of \( \Delta T_n \):

\[
\Delta T_c = \lim_{N \to \infty} \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\Delta T_n)^2}
\]

(24)

Cycle jitter is a measure of fluctuations in the time periods. Another measure of jitter that captures the dynamics in the variations of the output is cycle-to-cycle jitter, and is defined as:
\[ \Delta T_{cc} = \lim_{N \to \infty} \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\Delta T_{n+1} - \Delta T_{n})^2} \]  

which is *rms* value of the difference between two consecutive periods. Cycle jitter and cycle-to-cycle jitter (that are best suited for ring oscillators) are used as figures of merit in this study. The results of the ring oscillator are presented in the next section.

### 5.3.2. Ring Oscillator simulation results

The simulation schematics of the inverters used in the ring oscillator with and without the PD gate model are shown in Fig. 74. The ring oscillator outputs with and without PD gate model are compared with an ideal output for the whole simulation time, as shown in Fig. 75.

![Figure 74 Simulation schematic of CML buffer used in the ring oscillator a) without and b) with gate PD model](image)

It is observed that the output of the ring oscillator with the PD gate model tracks the ideal curve better than the ring oscillator without PD gate model. This observation is clear in Fig. 76, which shows the low-high transition times of the three outputs — i.e., with PD, without PD and ideal. From the figure it can be seen that the error between the ideal and the PD gate model is minimal and the two curves track each other compared
with the output of the ring oscillator without PD model. The variation in the time periods of the three outputs is shown in Fig. 77.

![Graph showing the output of the ring oscillator with and without PD model compared to an ideal output.](image)

**Figure 75** Ring oscillator output with and without PD model are compared with an ideal output

![Graph showing the comparison of the low-high transition times for the three outputs.](image)

**Figure 76** Oscillator output’s low-high transition times for the three outputs

As seen from the figure, the time period of the ideal output remains constant at 0.32 ns for the entire simulation time, while the output with the PD model is closer to the ideal curve with the average time period of 0.316 ns. The average time period of the output without PD model is 0.269 ns.
The cycle and cycle-to-cycle jitter is calculated using the above equations for both the ring oscillators with and without the PD model. The cycle jitter for the ring oscillator output with the PD model is 16.98 ps and without PD model is 17.93 ps and the cycle-to-cycle jitter is 16.45 ps and 15.04 ps for the ring oscillator with and without the PD model, respectively. The above results show that the ring oscillator with the PD model predicts the output behavior closer to the ideal oscillator than the ring oscillator without the PD model. Since in this analysis, there are no additional noise sources that can contribute to the jitter, the output of the oscillator should be closer to the ideal behavior, as achieved by considering PD gate model. It is also shown that neglecting the PD model in the GHz frequency range can result in over-estimation of the cycle and cycle-to-cycle jitter. The above observations stress the point that the PD gate model has significant impact on the device and circuit applications making it imperative to include this model for GHz applications.
Chapter 6

Conclusions

The incessant demand for simultaneous improvements in performance, cost, size and power consumption in the semiconductor industry heavily favors monolithic System-On-Chip (SoC) implementations. Although suffering from inferior device properties compared to other process technologies, CMOS offers the unique advantage of high levels of integration of RF, digital, and analog circuitry. While faster and smaller transistors are consistently being developed in conformance to Moore’s Law, MOS device modeling has been a major stumbling block in the implementation of these devices. SoC solutions require the coexistence of circuit blocks operating at varied frequencies (e.g. RF, analog, and digital). Typically, device modeling methodologies for these blocks are radically different. A device of specific dimensions, in a given technology is normally represented by different models based on its frequency of operation. Hence, co-simulation across a wide frequency range of interest is time consuming. This is especially true when complicated sub-circuits must be used to simulate device behavior for frequencies in the RF range. State-of-the-art SPICE/BSIM models represent CMOS devices extremely well at the low frequencies for which they were intended. However, these same models suffer many shortcomings in representing high frequency behavior. In particular, they are generally poor in predicting the impedance and noise properties at GHz frequencies. For high frequency circuits which depend on an accurate representation of these parameters, more complicated sub-circuit models are generally required. Unfortunately, development of these high frequency models demands characterization tools, specialized test structures, and extraction
techniques specifically targeted for this purpose. The additional overhead of supplying these models often means that foundries assign them lower priority and deliver them well after a new technology has been introduced. One of the major barriers in migrating to the sub-100 nm high-speed design regime is the lack of scaleable, frequency independent device models. As a first step towards such a model, a semi-empirical MOS gate model was developed and validated for frequency range of 1-20 GHz in this work. Various device structures with different dimensions were fabricated and characterized across three technology nodes using LSI Logic’s 180, 110 and 90 nm processes. It is predicted that for future CMOS technologies polysilicon will be replaced by metal gates for better device performance and process compatibility. This work addresses issues faced by the design and device modeling community currently and provides solution for present and near future problems.

Depletion in the polysilicon gate electrode is predominant and cannot be ignored in the GHz frequency range. Poly depletion effect was modeled in the 1-20 GHz range for high performance NMOS devices in 180, 110 and 90 nm technologies. The depletion capacitance and resistive effect in the depleted poly are modeled as a parallel RC combination. The capacitive and resistive spread in the depleted poly are approximated with a lumped capacitor and resistor, respectively. The depletion width is calculated numerically using a program called Schred, assuming uniform profiles for gate and substrate regions. The depletion capacitance is derived from the depletion width based on the device dimensions. The calculations show that the poly becomes inverted around the gate bias of 1V (for 110 nm process and 1.2 V VDD). However the lumped approximation of the resistance is derived empirically from measured data. The behavior of the MOS
input impedance and its dependence on device length, NF, bias conditions and frequency were studied for different technologies with the help of measured data. The PD model explains this behavior of input impedance and the performance of the model is further verified by comparing the simulated and measured input impedances. The gate impedance model along with BSIM4 was simulated for a 110 nm process for various width, length and bias conditions. The results show that the simulated input impedance deviates from measured data for gate bias below the threshold voltage and at $V_{dd}$. However it works very well at all other bias conditions over different widths and lengths. It was also shown that the model works better than existing state of the art high frequency gate models and is compatible with both BSIM3v3 and BSIM4 MOS models in the 1-20 GHz frequency range.

The PD model was further modified for a wide frequency range of 200 M-20 GHz and verified with the help of measured data for all three technologies. The study showed that a difference between the calculated and simulated poly-depletion capacitance parameter exists due to approximations made while calculating the depletion capacitance. It was also observed that this difference is greatest under sub-threshold conditions. However, the calculation of this parameter is very accurate in the saturation region of operation. The PD model is used in device and circuit level applications to evaluate extraction techniques and estimate the performance of high-speed digital circuits. To study the extraction techniques of MOS transistors, test structures were developed for CS and CD configurations and measurements were carried out in 1-20 GHz range for various device dimensions and bias conditions. The PD model behavior was studied for BSIM3v3 and BSIM4 models with the PD model. It was seen that the input impedance
predicted fit very well with the measured data with the standard models (extracted using grounded-source configurations) for both BSIM3v3 and BSIM4 for the CS devices. However, this was not true for CD devices, as the models used for CD were extracted using grounded-source configurations and did not predict the input impedance accurately. To compensate for this discrepancy and to fit the input impedance a capacitance was added for BSIM3v3 model. The simulation results show a better fit in CD case with the modifications for BSIM3v3 and by increasing series resistance for BSIM4 models. By including these modifications a universal model has been developed which can be used for both CS & CD configurations. However, the model needs refinement for low frequencies (below 5 GHz).

The degradation in the performance of CML circuits due to run-length variations was studied. The accurate PD gate model was used to predict the performance of a CML buffer designed in 110 nm technology with varying run length. Simulation results using the PD gate model show that the run-time variations cause performance degradation (which is erroneously masked by existing models) of the buffer, worst case being 40% for an input bit stream of logic 1 followed by a string of logic 0s. This study establishes that by using state of the art modeling approaches, real time errors caused by the input data itself can be predicted. The model is also used in a ring oscillator application to calculate jitter. The ring oscillator without any additional noise sources is simulated with and without the PD model to compare errors in jitter calculation. The results show that the ring oscillator with PD model performs closer to the ideal output and is more accurate. It was also shown that neglecting the PD effect can result in over-estimation of cycle and cycle-to-cycle jitter.
In summary, this work presents an accurate gate model for GHz applications and future directions include developing a fully scalable gate model by theoretically calculating the resistive component of the model. The model can be enhanced by studying the effect of process and layout parameters on the gate and by incorporating parasitics such as gate inductance and oxide transconductance. The frequency independent gate model can be further extended to develop dynamic MOS models to estimate the gate impedance parameters and design techniques to overcome the errors caused by run-length variations.
REFERENCES


