A NEW NON-QUASI STATIC MOSFET MODEL

by

Ibrahim Chamas

BS, Lebanese University, 2001

Submitted to the Graduate Faculty of

School of Engineering in partial fullfilment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

SCHOOL OF ENGINEERING

This thesis was presented

by

Ibrahim Chamas

It was defended on

December 6th, 2004

and approved by

Hong Koo Kim, Professor, Electrical Engineering Department

Dietrich W. Langer, Professor, Electrical Engineering Department

Thesis Advisor: Mahmoud El Nokali, Professor, Electrical Engineering Department

A NEW NON-QUASI STATIC MOSFET MODEL

Ibrahim Chamas, MS

University of Pittsburgh, 2004

Recent progress in wireless communication is sustained through integrated circuit technologies that offer a low cost and low power devices that operate in the Radio Frequency (RF) range with relatively low noise figure. The submicrometer CMOS technology presents a serious alternative to the more expensive, high power GaAs and Si bipolar technologies that have been used for the design of high frequency ICs. Design testing and verification through circuit simulation is a critical step in the design cycle of RF integrated circuits (RFICs). Accurate device models are therefore required to reduce design cycles and to achieve success when the circuit is finally committed to silicon.

This thesis addresses the Radio Frequency (RF) small-signal and large-signal models for the MOS transistor. The quasi-static (QS) and non-quasi-static (NQS) models are discussed and the assumptions used in their development are examined. The various charge components are briefly introduced and the source/drain charge partitioning is presented. The limitation of the QS approach at high frequency is investigated using the Bsim3v3.1 model. The development of a first order NQS small-signal model is briefly presented and its suitability for RF applications is indicated. The effect of the distributed gate, channel, and substrate resistances on the high

frequency characteristics of the MOS transistor is examined. We propose a Radio Frequency small-signal equivalent circuit (EC) together with an efficient parameter extraction algorithm that is necessary for the device optimization and the development of accurate large-signal models. The validity of the proposed model and the accuracy of the extraction method are verified by comparing Pspice simulation results of the EC to experimental data and the Bsim3v3.1 model up to 10GHz.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 THESIS ORGANIZATION 3

2.0 QUASI VERSUS NON-QUASI-STATIC MODELING OF MOSFETS 5

2.1 QUASI-STATIC MODELING (LARGE SIGNAL ANALYSIS) 6

2.2 NON-QUASI STATIC MODELING (LARGE SIGNAL ANALYSIS) 12

2.3 A QUASI –STATIC MEDIUM FREQUENCY SMALL SIGNAL MODELING 18

2.4 A COMPLETE QUASI-STATIC SMALL SIGNAL MODEL 24

2.5 Y-PARAMETER MODELING 34

2.6 A SMALL SIGNAL NON-QAUSI STATIC MODEL 38

2.7 CONCLUSION 52

3.0 EFFECT OF EXTRINSIC ELEMENTS ON THE HIGH FREQUENCY PERFORMANCE OF MOSFETS 54

3.1 MODELING AND IMPACT OF THE GATE RESISTANCE ON THE DEVICE OPERATION AT RF 55

3.2 CUT-OFF AND MAXIMUM OSCILLATION FREQUENCY 60

3.3 EFFECT OF SEMI-CONDUCTING SILICON SUBSTRATE ON THE RF PERFORMANCE OF MOSFETS 68

3.4 SCATTERING PARAMETERS (S-PARAMETERS) 70

3.5 CONCLUSION 71

4.0 RADIO FREQUENCY MODELING AND PARAMETER EXTRACTION OF SUB-MICROMETER CHANNEL LENGTH MOSFETS ACCOUNTING FOR FIRST ORDER NQS EFFECTS 73

4.1 PROPOSED MODEL ANDPARAMETER EXTRACTION PROCEDURE 76

4.2 Y-PARAMETER ANALYSIS AND PARAMETER EXTRACTION 84

4.2.1 EXTRACTION OF SMALL SIGNAL PARAMETERS 87

4.3 MODEL VERIFICATION 95

4.4 CONCLUSION 101

5.0 CONCLUSIONS AND FUTURE WORK 102

5.1 POTENTIAL IMPROVEMENT 103

5.2 FUTURE WORK 104

APPENDIX 105

Y-PARAMETER ANALYSIS OF THE PROPOSED RF SMALL-SIGNAL MODEL 105

BIBLIOGRAPHY 111

LIST OF TABLES

Table 1 High frequency figures-of-merit for different technologies 64

Table 2 Average and optimized values of the extracted model parameters 94

LIST OF FIGURES

Figure 1 Intrinsic part of the MOSFET is shown in the dashed box 7

Figure 2 Simulated turn-on drain current using the Bsim3v3.1 QS model 10

Figure 3 Simulated turn-on drain current using the Bsim3v3.1 QS model for the three charge partitioning schemes 11

Figure 4 Simulated turn-on source current using the Bsim3v3.1 QS model for the three charge partitioning schemes 12

Figure 5 dividing the transistor into smaller sections along the its length is the key to NQS modeling 13

Figure 6 Characteristic of the Drain current using NQS analysis resulting from applying a step input at the gate with zero rise-time..... 15

Figure 7 A medium frequency (QS) small-signal model [9] 19

Figure 8 A medium frequency small signal model including the drain to bulk path..... 21

Figure 9 A complete Quasi-static small signal model 28

Figure 10 Circuit schematic used to simulate the frequency response of the drain current for the Bsim3v3.1 complete QS model 33

Figure 11 The magnitude response of the ac drain current generated by the Bsim3v3.1 complete QS model 34

Figure 12 A setup for deriving the Y-parameter model..... 35

Figure 13 Y-parameter model..... 37

Figure 14 The oxide capacitance and channel resistance form a distributed RC network 39

Figure 15 A first order NQS gate-to-drain equivalent circuit..... 45

Figure 16 A first order Non-quasi-static model 47

Figure 17 Magnitude and phase of the drain current as predicted by the NQS Bsim3v3.1 model52

Figure 18 The gate resistance forms a distributed RC network with the oxide capacitance. The transistor can be divided into subsections to formulate the lumped equivalent gate resistance. 56

Figure 19 The frequency response of the drain current considering for a polysilicon gate resistance of 100Ω. The distributed structure of the gate resistance can be approximated by a lumped resistance one-third the DC value 59

Figure 20 MOS transistor showing the small-signal elements needed to calculate the cutoff frequency w_T. 61

Figure 21 small signal equivalent circuit including the electrode gate resistance. This model is used to calculate the maximum oscillation frequency. 62

Figure 22 The bulk depletion capacitance and the substrate resistance form a distributed RC network at RF 74

Figure 23 Equivalent circuit representation of the test structure used to de-embed the measured high frequency data of the MOS transistor. 78

Figure 24 An RF small-signal MOSFET model based on the first order non-quasi-static effect. The effective gate-to-source and gate-to-drain channel resistance is modeled by R_{chgs} and R_{chgd} respectively. Substrate signal coupling through the intrinsic body node and junction capacitances is examined by introducing a new 3-resistor T-network. 79

Figure 25 Extraction of the gate transconductance was obtained from the Y-intercept of $\mathrm{Re}[Y_{21}]$ versus w^2 88

Figure 26 The drain-to-source conductance was obtained from the Y-intercept of $\mathrm{Re}[Y_{22}]$ versus w^2 89

Figure 27 The extracted gate-to-source, gate-to-drain, and drain-to-gate capacitances are almost constant with frequency, asserting the physical validity of the model and accuracy of the parameter extraction method 90

Figure 28 The effective gate-to-source channel resistance R_{chgs} is determined from the slope of $\left(\mathrm{Re}[Y_{12}]+\mathrm{Re}[Y_{11}]\right)C_{gs}^{-2}$ as a function of w^2 91

Figure 29 The effective gate-to-drain channel resistance R_{chgd} is determined from the slope of $-\mathrm{Re}[Y_{12}]C_{gd}^{-2}$ as a function of w^2 92

Figure 30 Frequency dependence of R_{chgs} and R_{chgd}. The results show that these two resistors are almost constant with frequency, verifying the accuracy of the model. 93

Figure 31 Y_{11} parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated Y_{11} parameter of the proposed model matches the measured data very well and is more accurate than the Bsim3v3.1 model. 96

Figure 32 $\mathrm{Im}[Y_{21}]$ parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated $\mathrm{Im}[Y_{21}]$ parameter of the proposed model shows a slight deviation from the experimental data above 5GHz but becomes more accurate than the published model at 10GHz. 97

Figure 33 $\mathrm{Re}[Y_{21}]$ parameter of measured data, proposed model, Bsim3v3.1 model, and published model. Being based on NQS analysis, the simulated $\mathrm{Re}[Y_{21}]$ of the proposed model is much more accurate than that predicted by the published above 5GHz. 98

Figure 34 Y_{21} parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated Y_{21} parameter of the proposed model matches the measured data very well and is more accurate than the Bsim3v3.1 model. 99

Figure 35 Y_{22} parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated Y_{22} parameter of the proposed model matches the measured data very well and is more accurate than the Bsim3v3.1 model. 100

Figure 36 An RF small-signal MOSFET model based on the first order non-quasi-static effect. The effective gate-to-source and gate-to-drain channel resistance is modeled by R_{chgs} and R_{chgd} respectively. Substrate signal coupling through the intrinsic body node and junction capacitances is examined by introducing a new 3-resistor T-network. 105

Figure 37 The proposed EC in the Y_{22} measurement setup. The effect of the source-to-bulk junction capacitance C_{jsb} and resistance R_{sbi} on the output admittance is neglected for simplicity..... 107

ACKNOWLEDGEMENTS

Without the help, advice, and encouragement of many individuals this work would have never been completed. First, I would like to thank my advisor Professor Mahmoud El Nokali for his support and advice. I believe that I have been truly blessed by working with him over the past two years. I have benefited not only from this knowledge in device modeling, but also from his keen personality. I wish him well in all his future endeavors. Appreciation is also extended to the members of my advisory committee, Professors Hong Koo Kim and Dietrich W. Langer for reviewing this work. Special thanks go to my best friend Ghassan Azour and his parents, Nazih Azour and Nazira Azour for their support, help and encouragement. I would also like to thank Maya Shehab for her love and devotion.

I would also like to acknowledge my friends, Edward Saad, Dany Azour, Wessam Hassanein, Majd Sakr, Amit Gupta, Elizabeth Ellis, Chela Nagieh, and many others for making my life as a graduate student more enjoyable. I would like to thank the graduate program administrator, Sandy Weisberg for her help.

Although I mention my family last, they certainly made the greatest contribution to this work. Their love, care and support over so many years have really helped to shape my future. I thank my father Ramez Chamas for his invaluable advice and my late mother Zakie Chamas for her love and patience. I certainly appreciate the tremendous love from my two sisters

Lara and Rania, and my brother Hassan. I cannot thank them enough for their support and assistance throughout my life. I am also thankful to my brother-in-law Salim Bark for his continuous encouragement. My love goes to Lara's daughters Cherine, Rawan, and Dian, and to Rania's daughter Joel. Finally, I would like to acknowledge the University of Pittsburgh for its support and for the invaluable knowledge that I have gained throughout my stay.

1.0 INTRODUCTION

The possibility of communicating through the "ether" was demonstrated in 1901 when Guglielmo Marconi successfully transmitted radio signals across the Atlantic Ocean. The consequences and prospects of this demonstration were simply overwhelming. However, for many decades, two-way phone conversations would still go over wires and wireless transmission remained limited to one-way radio and television broadcasting. The invention of the transistor, the development of Shannon's information theory, and the conception of the cellular system paved the way for affordable wireless communications. Nowadays, mobile communication systems are moving rapidly from supporting voice only towards integrating digital data and multimedia transmissions as well. Thus, the projected applications for wireless technology are expanding beyond simple cellular phone handsets to include wireless internet connectivity in automobiles, cellular handsets, and personal data assistants (PDAs).

The push for wireless capabilities in the consumer market, in particular, is therefore accompanied by the demand for low-cost, wireless transceivers. Over the past three decades, the number of transistors in silicon (*Si*) based integrated circuits (ICs) has doubled about every 18 months. This well-known trend is referred to as "Moore's law," after Gordon E. Moore of Intel Corporation. Moore recognized the trend in 1965 that continued into the 21st century. Moore's primary intent for predicting future levels of integration was to push the improvement of the microprocessor. Thus, the research and development investments to keep track with Moore's law

have typically focused on digital applications. The corresponding economy-of-scale for Si digital ICs has, therefore, dramatically reduced the cost of microprocessors. On the other hand, Si has not been the ideal semiconductor for high frequency analog applications. Radio frequency ICs (RFICs) and monolithic microwave ICs (MMICs) have historically used compound semiconductors synthesized from elements in columns III and V of the periodic table (III-V semiconductors). III-V semiconductors have characteristically high electron mobility and are readily grown on semi-insulating substrates; features that are ideal for high frequency applications. However, high-speed analog and wireless ICs have recently sought to take advantage of the same Si economy-of-scale in an effort to reduce cost. The potential for high integration and lower cost has spurred research and advances in silicon-based technologies that include both bipolar and submicron complementary metal-oxide silicon (CMOS) devices (BiCMOS technologies).

The quadratic improvement in the microwave properties of CMOS devices with downscaling the channel length combined with the possibility of a system-on-chip integration has motivated extensive research on implementing a CMOS radio transceiver. Subtle physical mechanisms that govern the properties of deep sub-micron Mosfets have to be adequately described and incorporated in CAD tools to empower circuit designers with the ability to extract the best performance out of these devices. Design testing and verification through circuit simulation is a critical step in the design cycle of RF integrated circuits (RFICs). Accurate device models are therefore required to reduce design cycles and to achieve success when the circuit is finally committed to silicon. In essence, critical to the success of “RF CMOS” is the development of accurate and scalable RF Mosfet models.

1.1 THESIS ORGANIZATION

The objective of this research is to develop a physical small-signal equivalent circuit of a MOS transistor that can simulate the device characteristics in the Gega-Hertz range. Extraction of the model parameters is critical and should be considered at the same time the model is developed. To this end, an accurate and efficient parameter extraction procedure is presented. Chapter 2 discusses the quasi-static (QS) and non-quasi-static (NQS) models and the underlying assumptions used in their development. The QS approach assumes that the channel charge can adjust itself instantaneously and is only valid at low frequencies. A complete QS model is presented and the notion of trans-capacitance is introduced. The Bsim3v3.1 model is used to indicate the limitation of the QS model at high frequency. The NQS formulation is briefly introduced and it is shown that the channel resistance plays an important role at high frequencies. This resistance forms a distributed bias dependant RC network with the gate oxide and results in a signal delay between the transistor terminals. Non-quasi-static effects have been demonstrated to exit for both long and short channels and should be included in an RF Mosfet model. A first order NQS model is derived and discussed. Although this model extends the region of validity of the QS models, more elements should be added to enhance the accuracy at RF. These elements are related to the extrinsic parasitics and are discussed in chapter 3. At RF, the extrinsic components of the device play a prominent role in degrading the transistor performance and therefore have to be added to the intrinsic small signal NQS model. Six extrinsic capacitances are to be added: the gate-to-source capacitance C_{gsex}, the gate-drain-capacitance C_{gdex}, the gate-to-bulk capacitance C_{gbex}, the drain-to-bulk capacitance C_{dbex}, the source-to-bulk capacitance C_{sbex},

and the drain-to-source capacitance C_{dsex}. As for the parasitic resistances they can be divided into four parts: the resistance of the gate material, the substrate resistance, the resistance of the source and drain regions and their contacts. At high frequency the impedances of the capacitive components are comparable or even smaller than that of the resistive components that despite their distributive nature are represented by lumped elements in most models. The resistance of the drain and source regions are less important than the other two resistances and are most of the time omitted for simplicity. This chapter is mainly concerned with the effect of the gate and substrate resistances on the device operation at RF. The rest of the chapter is devoted to the discussion of some important parameters required for RF MOSFET modeling and to briefly introduce the scattering parameters.

Chapter 4 presents a radio-frequency (RF) small signal MOSFET model together with a simple parameter extraction algorithm. The intrinsic part of the proposed equivalent circuit (EC) is based on a first order non-quasi-static (NQS) formulation. The intrinsic and extrinsic components are extracted by performing Y-parameter analysis on the proposed model. Simple analytical expressions for extracting the effective gate-to-source and gate-to-drain channel resistances are developed for the first time. The distributed nature of the substrate resistance at high frequency is modeled using a novel lumped three-resistor T-network. Substrate signal coupling through the intrinsic body node and its effect on the output admittance are carefully examined. Simplified new expressions for the real and imaginary part of the output admittance are developed that prove essential in extracting the substrate related parameters. The physical validity of the model and the accuracy of the extraction method are verified by comparing Pspice simulation results of the EC to experimental data up to 10GHz.

2.0 QUASI VERSUS NON-QUASI-STATIC MODELING OF MOSFETS

This chapter discusses the quasi-static (QS) and non-quasi-static (NQS) models and the underlying assumptions used in their development. In order to model the frequency response of a semiconductor device the charge storage effect should be considered. For this reason, the various charge components in the MOS transistor are briefly discussed and the source/drain charge partitioning is presented [1]. To gain more insight into the limitations of the quasi-static approach, the large-signal behavior of the transistor is examined using the Bsim3v3.1 model [2]. Qualitatively, the QS formulation assumes that electrons travel along the channel with infinite speeds and hence ignores the finite time it takes them to cross the channel. The longer the channel is, the more time it takes the electron to travel from the source to the drain and the more the QS model will be in error. We introduce the non-quasi-static (NQS) formulation that divides the channel into sections, each being short enough to operate quasi-statically.

A small-signal quasi-static model for the intrinsic part of the transistor is briefly discussed [3-6]. The model uses five capacitances to simulate the behavior of the transistor in the mid-frequency range. In order to extend the region of validity of the model all capacitive effects have to be considered leading to an improved quasi-static model [7-9]. The failure of the quasi-static models to predict the high frequency behavior of the transistor is examined and the small-signal non-quasi static model is discussed. These small-signal circuits model the intrinsic part of

the device. The impact of the extrinsic elements on the radio frequency operation of the transistor will be discussed in chapter 3.

2.1 QUASI-STATIC MODELING (LARGE SIGNAL ANALYSIS)

Most circuit applications involve using the MOSFET in dynamic operation where either a small or a large time-varying signal is applied to one of its terminals causing a dynamic change in the transistor charges. The change in the transistor charge is supplied from the outside circuitry through the "*charging currents"* [10] that exist only in dynamic operation. The intrinsic part of the device is included between the drain and the source diffusion regions and contains the inversion layer, the depletion region, the oxide, and the gate material as depicted in figure 1 in the dashed box. The intrinsic part of the transistor is responsible for the actual operation while the extrinsic components act as parasitics that will limit the frequency performance of the transistor.

Three bias-dependent charge components are present in a MOS transistor namely the inversion, the bulk and the gate charges. The inversion charge is divided into a drain and source charges that are used to evaluate the source and drain charging currents and the corresponding node capacitances. The three charges are found by solving Poisson's equation. A closed form analytical solution is possible for a one-dimensional case by imposing the gradual-channel approximation and by using the charge sheet approximation [11]. As the channel length is scaled down, the electric field along the channel becomes more significant. In this case, a two-dimensional analysis is required. The channel is considered short if the effective channel length

(L_{eff}) is not much larger than the sum of the drain and source depletion widths. Moreover, if the channel width is narrow, i.e. the width of the transistor is not much larger than the depletion region depth underneath the gate, the electric field along the width of the transistor is significant. Accounting for both short and narrow channel effects requires a three-dimensional numerical analysis [12-14] that gives accurate results, but fails to provide a simple computationally efficient model. A viable alternative consists of proposing empirical and semi-empirical approximations that yield simple equations that are reasonably accurate and computationally efficient [15], [16].

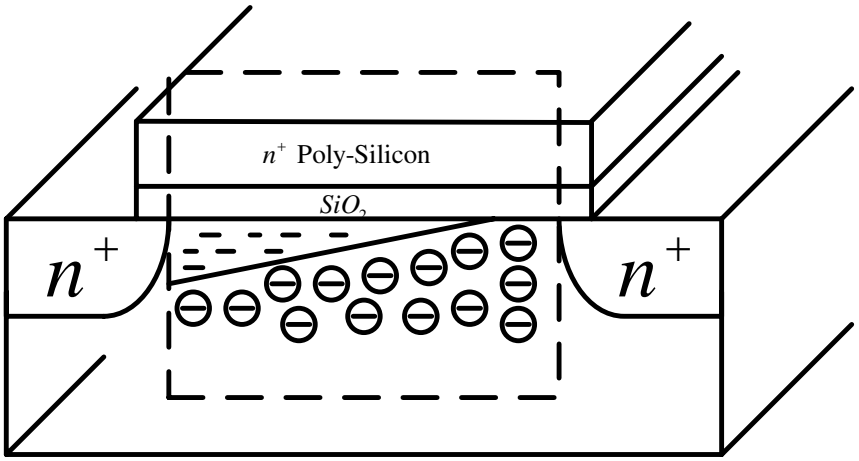

Figure 1 Intrinsic part of the MOSFET is shown in the dashed box

Until recently, most of the MOSFET models in Spice are based on the quasi-static assumption and are inadequate at high frequencies [7], [17-19]. In the quasi-static formulation (QS), the finite time it takes the channel charge to reach equilibrium when a time-varying voltage

is applied is ignored. The channel charge is assumed to follow the signal with no delay and with no degradation in its absolute value. Obviously, this assumption leads to erroneous results if the frequency of the applied signal is of the order of the unity gain transition frequency (f_T) [20]. As will be seen later, f_T is inversely proportional to the square of the effective channel length for long-channel devices. Hence, the shorter the channel is, the larger f_T will be and the transistor can be successfully described by the quasi-static analysis. For digital applications, the quasi-static model fails if the rise- or fall-time of the applied signal is less than or comparable to the channel transit time.

The Bsim3v3.1 will be used to demonstrate the failure of the QS model in predicting the high frequency characteristics of the MOS transistor. The source/drain charge partitioning ratio can be set to the desired value by using the *XPART* parameter in Cadence. Existing charge partitioning ratios are 0/100, 50/50, and 40/60. They correspond to *XPART* = 1, 0.5, and 0 respectively. The inversion layer charge is supplied from the voltages applied to the source and drain terminals. The 0/100 charge partitioning scheme assigns all the inversion layer charge to the source region. This partitioning scheme is commonly used by circuit designers to suppress unrealistic large drain current spikes during transient simulation [21]. Unfortunately, this non-physical solution shifts the problem to the source terminal and should only be used when the source terminal is grounded. The 50/50 partitioning scheme assumes that the source and drain regions contribute equally to the total inversion charge. The 40/60 charge partitioning is the most physical and widely used scheme. It allocates the channel charge to the source and drain regions by assuming that the channel charge is linearly dependent on the distance along the channel [10]. Using Pspice simulation, the drain and source currents that the QS Bsim3v3.1 model predicts for a step input applied at the gate of an NMOS transistor will be investigated. To illustrate the

difference between the three charge partitioning schemes, the simulation is conducted for *XPART=0, 0.5, and 1.* The effect of the rise time on the simulated currents will also be shown. Intuitively, the QS model will result in unrealistic terminal currents if the rise-time (fall-time) of the gate signal is small. The parameters used in the Bsim3v3.1 model are extracted from the $0.35\mu m$ TSMC CMOS process. The simulated MOS transistor has a channel length and width of 0.35 um and 20 um respectively. The XPART parameter is set to 0 (40/60 partitioning) in this experiment. An input step voltage with a rise-time of 10ps, 100ps and 1ns and a maximum value of 1v is applied at the gate. Using different rise times for a single pulse signal in Pspice is possible through the use of the *parameter* element. Consequently, both a parametetric and a transient analysis are needed. The drain terminal is tied to the supply voltage (3.3 volts) and the source and bulk terminals are tied to ground. This biasing setup ensures that the transistor stays in saturation even when the input signal reaches its maximum value of 1 v. During a transient, the drain (or source) current flows through the intrinsic and extrinsic capacitors of the device. In order to examine the drain and source currents of the intrinsic device, the extrinsic gate-to-source and gate-to-drain overlap capacitors denoted as *CGSO* and *CGDO* are set to zero.

Figure 2 shows an initial large negative drain current when the rise time of the gate signal is 10ps. The drain current unrealistically decreases to -0.2 *mA* and then gradually increases and reaches its steady-state value of $1.1mA$ around 10ps. This negative drain current is not observed in practice if the transistor is in the saturation region. In reality, the channel is depleted of electrons until the input reaches the threshold voltage which for this technology is 0.51 volts. At that time, electrons enter the channel through the source electrode and start their journey towards the drain. It is only when the electrons reach the drain electrode that the drain current can be observed. The QS formulation ignores this fact and assumes that the channel is instantaneously

charged to its equilibrium value and that the electrons move along the channel with infinite speed. Hence the QS approximation fails to determine the fine details of the current and breaks down if the input changes too fast. As the input signal is made slower (increase its rise-time) the negative drain current decreases and the QS model is more applicable. For a rise-time of 10ns, the drain current stays at zero for a *delay* time around 0.6ps and then increases gradually to its steady state. This is a more realistic result as explained above. Figures 3 and 4 show the simulated drain and source current for the three charge partitioning schemes and for a rise-time of 10ps. Although, the negative drain current spike is suppressed for XPART = 1, the source current shows the maximum negative spike for this partition since all the channel charge is allocated to the source terminal.

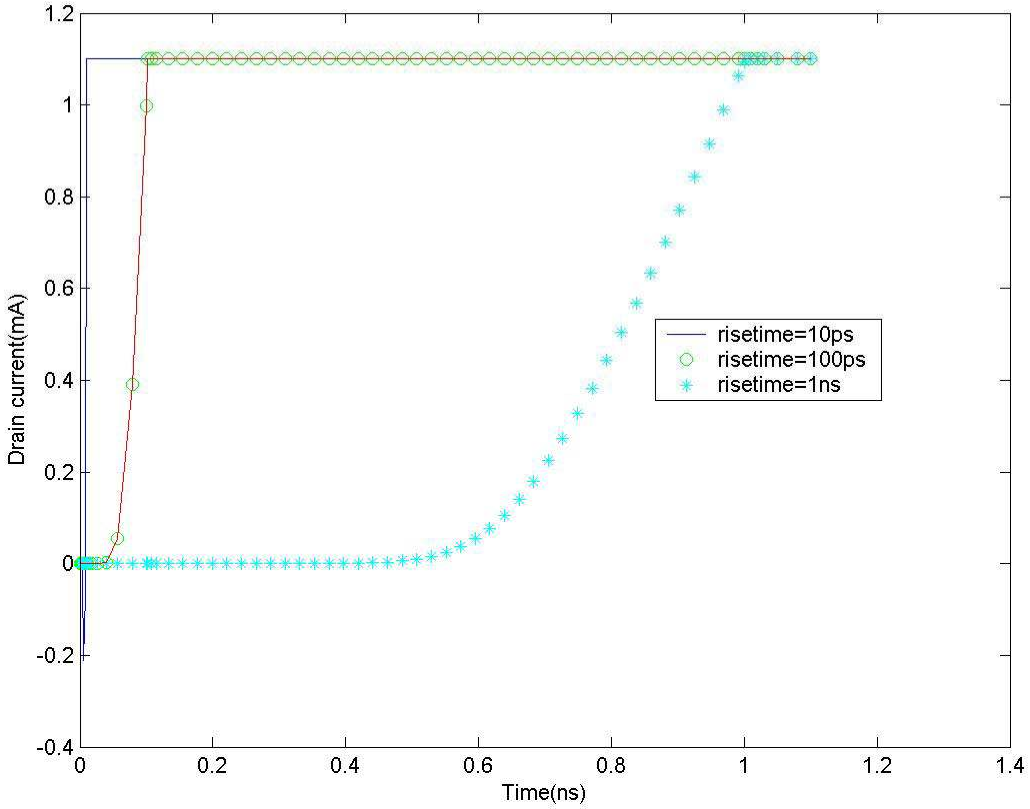

Figure 2 Simulated turn-on drain current using the Bsim3v3.1 QS model

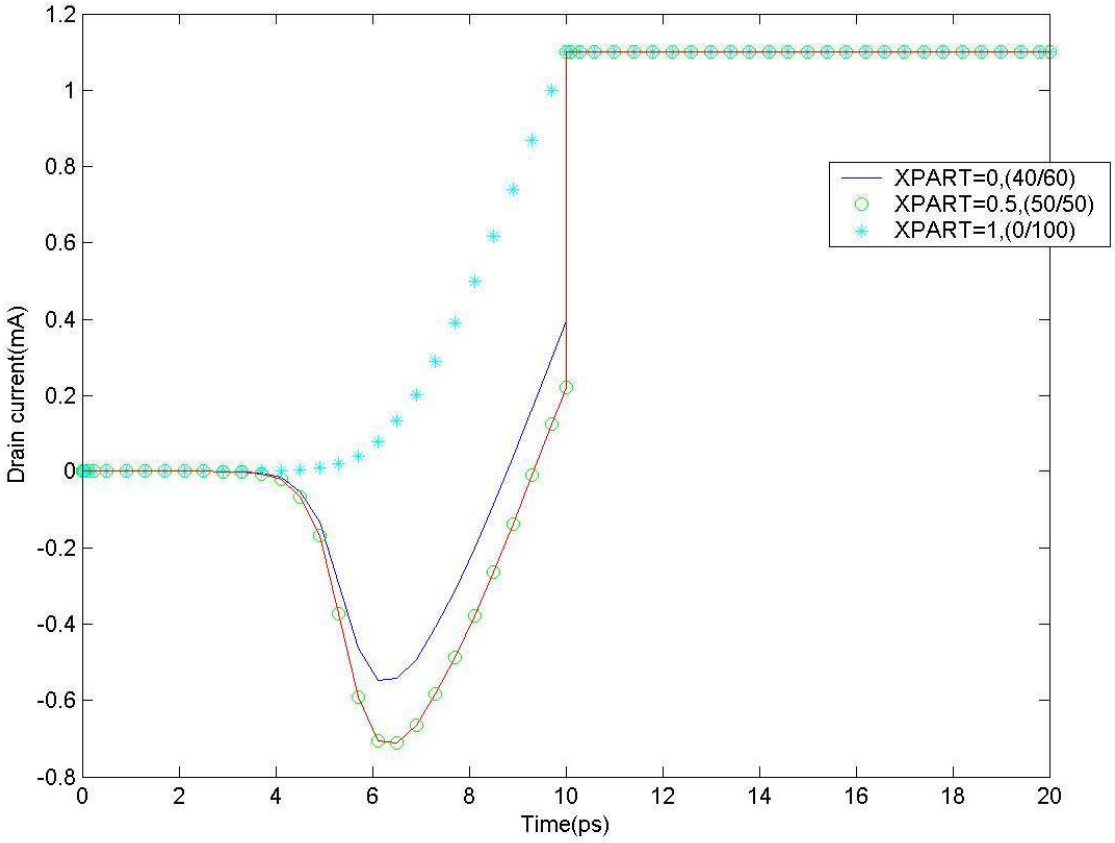

Figure 3 Simulated turn-on drain current using the Bsim3v3.1 QS model for the three charge partitioning schemes

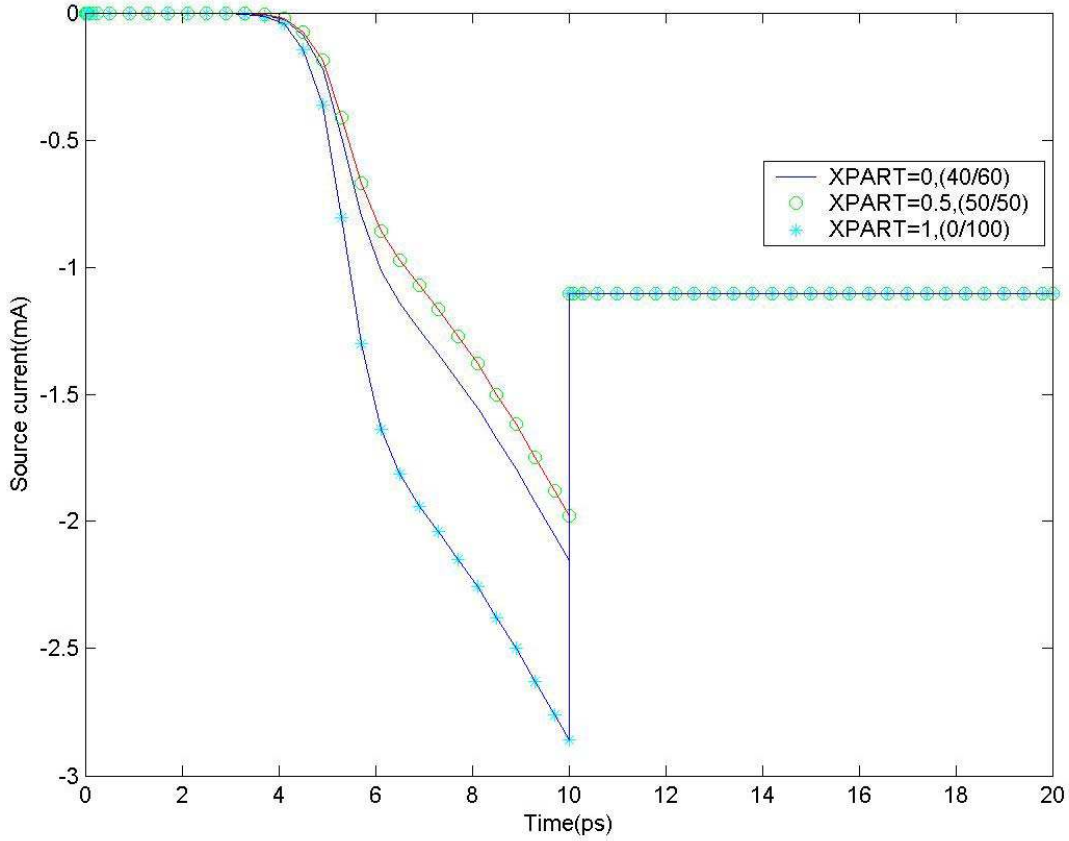

Figure 4 Simulated turn-on source current using the Bsim3v3.1 QS model for the three charge partitioning schemes

2.2 NON-QUASI STATIC MODELING (LARGE SIGNAL ANALYSIS)

In the previous section we found that the quasi-static model breaks down if the input voltage changes too fast. To alleviate this problem the transistor can be divided into smaller sections, each section being small enough to be modeled quasi-statically. The higher the

frequency of operation, the more sections will be needed and the shorter each section will be. Figure 5 illustrates this point.

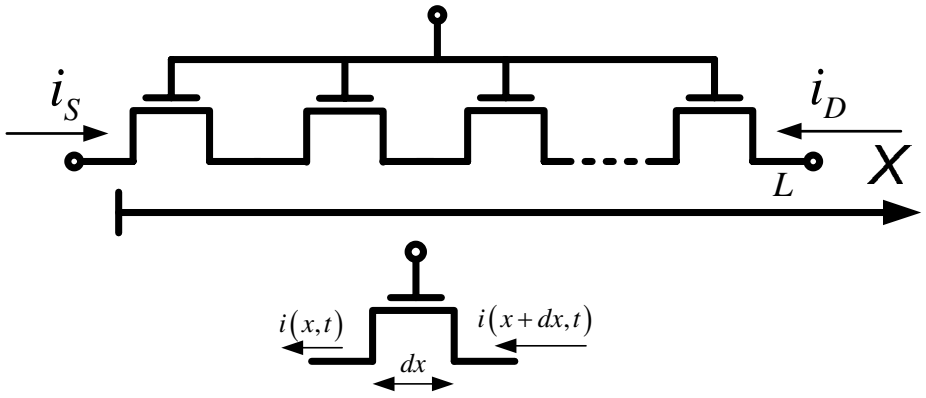

Figure 5 dividing the transistor into smaller sections along the its length is the key to NQS modeling

The source and drain currents of each sub-transistor are usually different during transient to account for the inversion layer charge build up. Hence the current through each section is a function of time and position. First, we need to write the equations that govern the behavior of this collection of transistors:

a) The continuity equation:

$$\frac{\partial i(x,t)}{\partial x} = W \frac{\partial q_I'(x,t)}{\partial t}, \tag{2-1}$$

where $q_I'(x,t)$ is the instantaneous inversion layer charge per unit area.

In steady state, the inversion layer charge is independent of time and hence the current is independent of position.

b) Inversion layer charge:

$$q_I^{'}(x,t) = -C_{ox}^{'}\left[v_{GB}(t) - V_{FB} - \psi_s(x,t) - \gamma\sqrt{\psi_s(x,t)} \right], \tag{2-2}$$

where, $C_{ox}^{'}$ is the oxide capacitance per unit area, $v_{GB}(t)$ is the instantaneous gate-to-bulk voltage, V_{FB} is the flat-band voltage, γ is the body coefficient, and $\psi_s(x,t)$ is the surface potential in the silicon material.

c) Current equation [9]:

$$i(x,t) = -\mu W q_I^{'}(x,t)\frac{\partial \psi_s(x,t)}{\partial t} + \mu W \phi_t \frac{\partial q_I^{'}(x,t)}{\partial x}, \tag{2-3}$$

where μ is the mobility and ϕ_t is the thermal voltage.

Notice that the first term on the right hand side of the equation corresponds to the drift current while the second term represents the diffusion current. Equations (2-1), (2-2), and (2-3) contain three unknowns: $i(x,t)$, $\psi_s(x,t)$, and $q_I^{'}(x,t)$. Solving these equations requires a set of initial and boundary conditions which primarily depend on the applied voltages. The mathematics involved in solving the above system can get tedious. In [1], every point along the channel was assumed to be strongly inverted which simplifies the analysis considerably. In strong inversion, the drift current dominates and the diffusion current can be ignored. Moreover, the surface potential $\psi_s(x,t)$ can be approximated by $v_{CB}(x,t) + \phi_0$, where: $\phi_0 = 2\phi_F + 6\phi_t$ for uniformly doped silicon [9], and $v_{CB}(x,t)$ is the channel-to-bulk voltage at position x and at time t. In [22], an input step with zero rise time and amplitude (V) is applied to the gate of the transistor. The drain is biased at Vdd to insure operation in the saturation region at all times and the source and bulk terminals are grounded. In this setup $v_{GB}(t)$=V. Numerical techniques have

been adopted to solve equations (2-1, 2-2, and 2-3) [22]. The instantaneous drain and source current are calculated from $i(x,t)$ as follows:

$$i_D(t) = i(L,t) \tag{2-4}$$

$$i_S(t) = -i(0,t). \tag{2-5}$$

The characteristic of the drain current is sketched in Figure 6 where τ_d is the delay time between the application of the input voltage and the flow of the drain current and τ_0 is the time at which the current reaches 98% of its steady state value.

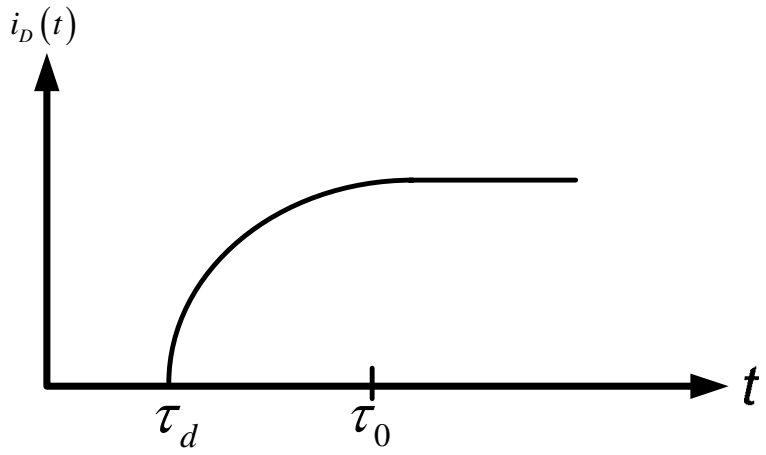

Figure 6 Characteristic of the Drain current using NQS analysis resulting from applying a step input at the gate with zero rise-time

At this point it is instructive to estimate the average time (τ) it takes the electron to cross the entire channel at DC or steady-state conditions. In the strong-inversion saturation region the electron transit time can be calculated as follows [10]:

$$\tau = \frac{|Q_I|}{I_{DS}} = \frac{\frac{2}{3} WLC_{ox}^{'} \left(V_{GS} - V_T\right)}{\frac{1}{2} \mu C_{ox}^{'} \frac{W}{L} \left(V_{GS} - V_T\right)^2} = \frac{4}{3} \tau_0 , \tag{2-6}$$

where:

$$\tau_0 = \frac{L^2}{\mu \left(V_{GS} - V_T\right)} . \tag{2-7}$$

Note that equation (2-6) ignores the electron velocity saturation associated with short channel devices and also neglects narrow channel effects that would otherwise make the equation more complicated. However, equation (2-6) is a good estimate of the channel carrier transit time τ which is proportional to L^2. The transit time decreases quadratically with scaling the channel. This result makes scaling a very attractive feature of CMOS in addition to its low cost and its high density of integration. Obviously, the above derivation breaks down if carrier velocity saturation is present along the channel [22]. In this case, the expressions for the inversion layer charge (Q_I) and the dc current (I_{DS}) will be different. In the extreme case where velocity saturation occurs along the whole channel the value of (τ) can be calculated as:

$$\tau = \frac{L}{v_{dsat}} . \tag{2-8}$$

In this case, the transit time (τ) is linearly proportional to the channel length. In the case when velocity saturation span part of the channel, (τ) will be proportional to L^{α} where α is between 1 and 2. On the other hand, the value of the delay time τ_d was found to be [1]:

$$\tau_d \approx 0.38 \tau_0 , \tag{2-9}$$

As seen from Figure 6, the drain current starts increasing at τ_d, the time it takes the electron to reach the drain terminal and increases gradually to its steady state value. At τ_0, the drain current would have reached around 98% of its steady state value [1]. This picture of the

drain current is a more realistic result than that obtained from the QS formulation which assumes that the channel will be filled with charge instantaneously at t=0 and that the drain current flows immediately and can be negative as seen in Figure 2. A more practical situation occurs when the input signal has a finite rise time [17, 23] rather than being the input step considered before. It has been shown that if the signal rise time is much smaller than τ_0, then the current delay time is given by equation (2-9). However, if the rise time is on the order of τ_0, the delay time if given by:

$$\tau_d = \sqrt{t_R \tau_0} \ . \tag{2-10}$$

Hence the delay time depends on the rise time. The faster the input changes (short rise time) the faster the drain current reacts (delay is reduced). This result can be explained as follows: as t_R is made smaller, the transistor will turn on sooner ($V_{GS} > V_T$) and the channel charge entering the source terminal will reach the drain in a shorter time. It has been shown that if the rise time of the input signal is larger than $20\tau_0$, the numerical solution of (2-1), (2-2), and (2-3) gives roughly the same results as the quasi-static model. The above analysis has been performed considering a long-channel device. Short channel effects such as velocity saturation will change the picture. The extreme case is given by equation (2-8) when a step input is applied and is found to be much larger than the delay time predicted by the long-channel theory [17]. On the other hand, if the rise time of the input signal is much larger than the above limit and the carrier velocity did not saturate then the delay time is found to be less than that predicted by the long-channel theory.

We have seen so far that the QS model breaks down if the input changes too fast compared to the channel transit time. We will find out later on that the small signal quasi-static

model fails if the frequency of the input signal is on the order of the unity gain transition frequency (f_T). Interestingly, f_T is on the order of the inverse of the channel transit time. Moreover, the channel length plays an important role in determining the validity of the QS model as seen from equations (2-7) and (2-8).

2.3 A QUASI –STATIC MEDIUM FREQUENCY SMALL SIGNAL MODELING

The small-signal model of the intrinsic part of the transistor at medium frequencies is shown in Figure 7. The small signal parameters g_m, g_{mb} and g_{ds} represent the gate transconductance, bulk transconductance, and output conductance respectively. These three elements model the transport portion of the drain-to-source current and can be evaluated as follows:

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}}\Big|_{V_{DS},V_{BS}} \tag{2-11}$$

$$g_{mb} = \frac{\partial I_{DS}}{\partial V_{BS}}\Big|_{V_{DS},V_{GS}} \tag{2-12}$$

$$g_{ds} = \frac{\partial I_{DS}}{\partial V_{DS}}\Big|_{V_{GS},V_{BS}} . \tag{2-13}$$

These small signal parameters are bias dependent and have different expressions for short and long channel devices [10]. These parameters depend on the slope or first-order derivative of the current equation thus presenting a major challenge in modeling MOS transistors for analog and RF applications. It has been shown [7] that although the DC drain current can be accurate, the

error in predicting the drain-to-source conductance g_{ds} may exceed 50%. Consider a small change in the gate-to-source voltage. Regardless of the pace of that change, the model in Figure 7 predicts that a small signal drain current will flow instantaneously through the voltage-controlled current source $g_m v_{gs}$ and therefore the Quasi-static assumption is clearly implied by this model.

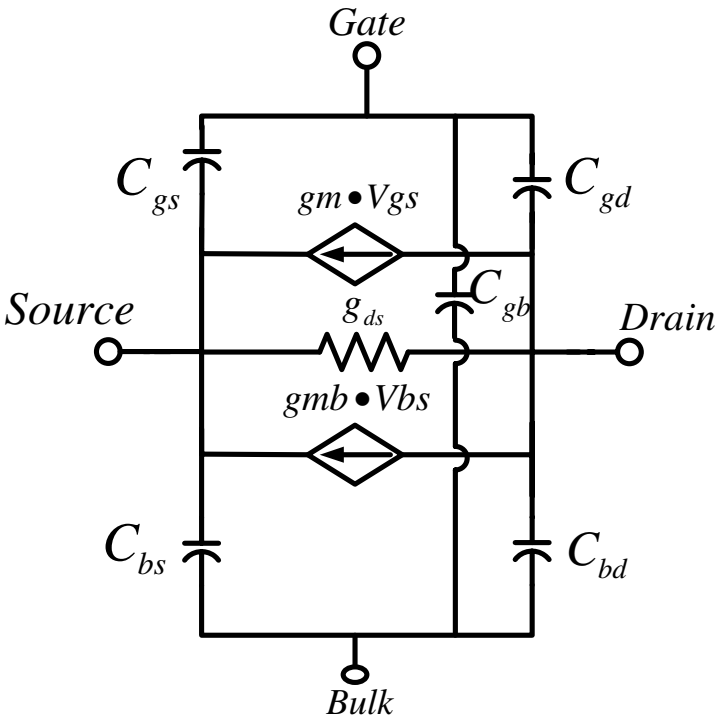

Figure 7 A medium frequency (QS) small-signal model [9]

The presence of impact ionization, which is more pronounced in short channel devices, adds more components to the model. High electric fields along the channel of sub-micrometer devices can cause carrier velocity saturation. These high fields usually exist when the transistor

is in the saturation region. Although the velocity of the electrons saturate, their random kinetic energy continues to increase. Some of these carriers will have enough energy to cause impact ionization where an electron collides with the silicon lattice and generates an electron-hole pair. The drain absorbs the generated electrons and the holes drift towards the substrate terminal resulting in a drain-to-bulk current [24]. The transistor is then said to be in weak avalanche. In general, the drain current contains two components and is given by:

$$I_D = I_{DS} + I_{DB}\,, \tag{2-14}$$

where I_{DS} and I_{DB} are the drain-to-source and drain-to-bulk currents respectively. It should be mentioned that the drain-to-bulk current is usually several orders of magnitude smaller that I_{DS} when the device is operated nominally. Figure 8 shows a small-signal model including the drain to bulk path that has been accounted for through two dependent current sources $g_{bg}v_{gb}$, $g_{bs}v_{sb}$ and the drain-to-bulk conductance g_{bd} [10]. The contribution of the current sources to the total current is negligible and they can be omitted from the model with no loss in accuracy. The drain-to-bulk conductance on the other hand can impact the output conductance, g_o, of the transistor [26] defined as:

$$g_o = \frac{\partial I_D}{\partial V_{DS}}\Big|_{V_{GS},V_{BS}}\,. \tag{2-15}$$

Using equation (2-14):

$$g_o = \frac{\partial I_{DS}}{\partial V_{DS}}\Big|_{V_{GS},V_{BS}} + \frac{\partial I_{DB}}{\partial V_{DS}}\Big|_{V_{GS},V_{BS}} = g_{sd} + g_{bd}\,. \tag{2-16}$$

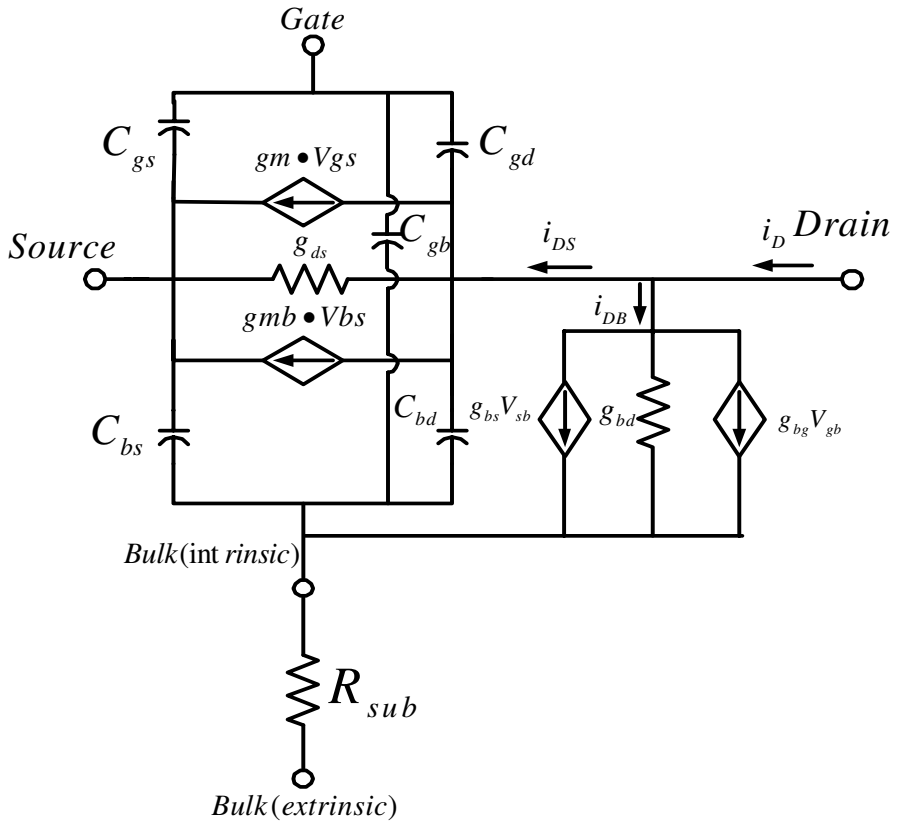

Figure 8 A medium frequency small signal model including the drain to bulk path

The intrinsic gain of the MOS transistor is inversely proportional to the output conductance and is given by $g_m g_o^{-1}$. Hence large output conductance is undesirable for getting high gain. So far we considered the output conductance of the intrinsic transistor which is given by (2-16). CMOS chips are fabricated on resistive substrates. The effect of the substrate resistance on the high frequency performance of the transistor will be discussed later. For the time being suppose that the substrate resistance can be represented by a single resistor R_{sub} connected between the intrinsic and extrinsic body nodes. With the help of the bulk transconductance g_{mb}, the output conductance of the transistor can be shown to be [26]:

$$g_o \approx g_{ds} + g_{mb} R_{sub} g_{db} + g_{bd} . \tag{2-17}$$

The middle term in the above equation can be explained qualitatively as follows. An increase in the drain-to-source voltage will cause an increase in I_{DB}. The latter flows through R_{sub} and increase the intrinsic bulk-to-source voltage v_{bs} by and amount equal to $dI_{DB} R_{sub}$. Due to this effect the total drain current will increase by $g_{mb} dI_{DB} R_{sub}$. Hence the increase in the output conductance due to R_{sub} will be:

$$\frac{dI_{DS}}{dV_{DS}} = \frac{g_{mb} dI_{DB} R_{sub}}{dV_{DS}} = g_{mb} R_{sub} g_{bd} . \tag{2-18}$$

Impact ionization can increase the output admittance by an order of magnitude [25] and thus it is crucial to include g_{bd} in the small-signal model. The capacitances *Cgs* and *Cgd* model the charging or capacitive effect of the source and drain regions on the gate terminal. Consider an experiment where the drain, gate, and bulk terminals are shorted (ac-wise) and the source potential is increased by an amount Δv_s. Since the transistor is assumed to be in strong inversion, the surface potential, ψ_s, will increase proportionally to the increase in v_s. This in turn will cause a decrease in the potential across the oxide $\left(V_G - \psi_s\right)$ and hence, a decrease in the gate charge ($\Delta Q_g < 0$). Considering small-signal operation, this decrease is proportional to the increase in the source potential and the constant of proportionality is represented by Cgs. Notice that Cgs models the capacitive effect of the source on the gate which might be different from the capacitive effect of the gate on the source as will be discussed when the complete quasi-static model is introduced. Similar arguments apply for the other four capacitances in Figure 7. The five capacitances can be evaluated at the bias point by simply taking the partial derivative of the charge with respect to voltage:

$$C_{gs} = -\frac{\partial Q_G}{\partial V_S}\Big|_{V_G, V_D, V_B} \tag{2-19}$$

$$C_{gd} = -\frac{\partial Q_G}{\partial V_D}\Big|_{V_G, V_S, V_B} \tag{2-20}$$

$$C_{gb} = -\frac{\partial Q_G}{\partial V_B}\Big|_{V_G, V_D, V_S} \tag{2-21}$$

$$C_{bs} = -\frac{\partial Q_B}{\partial V_S}\Big|_{V_G, V_D, V_B} \tag{2-22}$$

$$C_{bd} = -\frac{\partial Q_B}{\partial V_D}\Big|_{V_G, V_S, V_B} . \tag{2-23}$$

Two assumptions are inherent in the structure of the model presented in figure 7. First, this model assumes quasi-static operation as mentioned earlier. Second, the capacitive effects of the four terminals on each other were not completely accounted for. It has been shown that for critical applications this model will give satisfactory results if the operating frequency (w) is less than $0.1w_0$, where w_0 is defined as [10]:

$$w_0 = \frac{\mu(V_{GS} - V_T)}{L^2} . \tag{2-24}$$

It is instructive to note that the model fails at high frequency not because of the values of the small signal parameters but because of the nature of the model itself. In order to explain qualitatively the upper region of validity noted in (2-24) we need to revisit the assumptions behind the QS approximation. This formulation assumes that the inversion layer charge and hence the current respond instantaneously to the applied voltage. In other words, there is no delay between the cause (a change in a terminal voltage) and the effect (a change in the current). We know by now that this assumption breaks down at high frequencies since there is a non- zero delay. A delay in the time domain corresponds to a phase shift in the frequency domain. As a

matter of fact, the objective of a non-quasi-static model is to capture this phase shift (delay) in the drain current as well be seen later. It has been shown [8] that the phase shift in the drain current in response to a variation in the gate-to-source voltage for a non-quasi-static model starts around a frequency $w = 0.1w_0$. Interestingly, w_0 can be expressed as:

$$w_0 = \frac{1}{\tau_o}, \tag{2-25}$$

where τ_o is given by equation (2-7). This is another clear indication of the direct correlation between channel transit time and the frequency at which the quasi-static model breaks down.

2.4 A COMPLETE QUASI-STATIC SMALL SIGNAL MODEL

It was mentioned in section 2.3 that in order to extend the validity of the model in figure 7, the capacitive effect of every terminal on the other has to be considered. This leads to a *complete* quasi-static model. In this section, we will briefly discuss the development of the model. The terminal currents in an MOS transistor can be decomposed into two parts: charging currents and transport or conductive currents. Three elements are needed to model the small signal transport phenomenon along the channel and were shown in Figure 7 as the two dependent current sources $g_m v_{gs}$, $g_{mb} v_{bs}$ and the drain-to-source conductance g_{ds}. The transport currents measure the change (an increase or a decrease) in the drain-to-source current in the long run i.e. at steady state. The charging currents (capacitive currents) on the other hand are transient currents that exist only when the terminal voltages are varying with time and are responsible for changing the charge corresponding to each terminal. The four terminal currents (charging and transport) are

defined as entering the device. The four charging currents *entering* the transistor can be written as follows:

$$i_{dch}(t) = \frac{\partial q_D\left(v_D(t), v_G(t), v_B(t), v_S(t)\right)}{\partial t} \tag{2-26}$$

$$i_{sch}(t) = \frac{\partial q_S\left(v_D(t), v_G(t), v_B(t), v_S(t)\right)}{\partial t} \tag{2-27}$$

$$i_g(t) = \frac{\partial q_G\left(v_D(t), v_G(t), v_B(t), v_S(t)\right)}{\partial t} \tag{2-28}$$

$$i_b(t) = \frac{\partial q_B\left(v_D(t), v_G(t), v_B(t), v_S(t)\right)}{\partial t}, \tag{2-29}$$

where q_D, q_S, q_G, and q_B are the drain, source, gate, and bulk charges respectively. The currents in the above equations add up to zero by applying KCL. Hence knowing three of the above currents is enough to model the charging mechanism. In a source referenced model it is convenient to consider the drain, gate and bulk charging currents. Since the MOS transistor has four terminals, equations (2-26) to (2-29) can be written in terms of sixteen *dependent* capacitors [10]. It turns out that only nine independent capacitors are needed to completely describe the charging mechanism. The three independent charging currents are expressed as follows:

$$i_{dch}(t) = C_{dd}\frac{dv_{ds}}{dt} - C_{dg}\frac{dv_{gs}}{dt} - C_{db}\frac{dv_{bs}}{dt} \tag{2-30}$$

$$i_g(t) = -C_{gd}\frac{dv_{ds}}{dt} + C_{gg}\frac{dv_{gs}}{dt} - C_{gb}\frac{dv_{bs}}{dt} \tag{2-31}$$

$$i_b(t) = -C_{bd}\frac{dv_{ds}}{dt} - C_{bg}\frac{dv_{gs}}{dt} + C_{bb}\frac{dv_{bs}}{dt}. \tag{2-32}$$

It is instructive not to think of the above capacitors as "parallel plate" capacitors. In general, these capacitors are defined as follows:

$$C_{xy} = -\frac{\partial q_x}{v_y}\Big|_{operating\ point} \tag{2-33}$$

$$C_{xx} = +\frac{\partial q_x}{v_x}\Big|_{operating\ point}, \tag{2-34}$$

where C_{xy} represents the capacitive effect of node y on node x, and C_{xx} models the capacitive effect of node x on itself. The algebraic signs in the above equations have been defined as such for convenience. This choice will lead to positive capacitance values. Note that if the terminal voltages are independent of time (constant), the charging currents drop to zero as obvious from equation (2-30), (2-31), and (2-32). Many small-signal circuits can be constructed to represent these equations. However, it would be more attractive if we can arrange the current equations so that the resulting circuit resembles that in figure 7 albeit with some added elements. We will go over the development of a more useful representation of the drain charging current since it is not included in [10]. It can be shown from the indefinite admittance matrix that the capacitance C_{dd} can be expressed as:

$$C_{dd} = C_{gd} + C_{bd} + C_{sd}. \tag{2-35}$$

Substituting equation (2-35) in (2-30), one gets:

$$\begin{aligned} i_{dch}(t) &= \left(C_{gd} + C_{bd} + C_{sd}\right)\frac{dv_{ds}}{dt} - C_{dg}\frac{dv_{gs}}{dt} - C_{db}\frac{dv_{bs}}{dt} \\ &= C_{gd}\frac{d\left(v_{dg} + v_{gs}\right)}{dt} + C_{bd}\frac{d\left(v_{db} + v_{bs}\right)}{dt} + C_{sd}\frac{dv_{ds}}{dt} - C_{dg}\frac{dv_{gs}}{dt} - C_{db}\frac{dv_{bs}}{dt} \\ &= C_{sd}\frac{dv_{ds}}{dt} + C_{gd}\frac{dv_{dg}}{dt} + C_{bd}\frac{dv_{db}}{dt} - C_{mb}\frac{dv_{bs}}{dt} - C_m\frac{dv_{gs}}{dt}, \end{aligned} \tag{2-36}$$

where:

$$C_m = C_{dg} - C_{gd} \tag{2-37}$$

$$C_{mb} = C_{db} - C_{bd}\,. \tag{2-38}$$

Equations (2-31) and (2-32) can be rearranged in the same manner. The resulting expressions for the drain, gate, and bulk charging currents are as follows:

$$i_{dch}(t) = C_{sd}\frac{dv_{ds}}{dt} + C_{gd}\frac{dv_{dg}}{dt} + C_{bd}\frac{dv_{db}}{dt} - C_{mb}\frac{dv_{bs}}{dt} - C_{m}\frac{dv_{gs}}{dt} \tag{2-39}$$

$$i_{g}(t) = C_{gd}\frac{dv_{gd}}{dt} + C_{gb}\frac{dv_{gb}}{dt} + C_{gs}\frac{dv_{gs}}{dt} \tag{2-40}$$

$$i_{b}(t) = C_{bd}\frac{dv_{bd}}{dt} + C_{gb}\frac{dv_{bg}}{dt} - C_{mx}\frac{dv_{gb}}{dt} + C_{bs}\frac{dv_{bs}}{dt}\,, \tag{2-41}$$

where:

$$C_{mx} = C_{bg} - C_{gb}\,. \tag{2-42}$$

The charging currents given by equations (2-39), (2-40), and (2-41) can be easily represented by an equivalent circuit. The total terminal current is the sum of a charging and a transport component. The complete quasi-static model is depicted in Figure 9. This model can be viewed as an extension of the model presented in figure 7 with the addition of four elements that have resulted from considering the complete charging mechanism. These four elements are:

1. The source to drain capacitance: C_{sd} .
2. The trans-capacitance current source: $C_{m}\frac{dv_{gs}}{dt}$.
3. The bulk trans-capacitance current source: $C_{mb}\frac{dv_{bs}}{dt}$.
4. The gate-to-bulk trans-capacitance current source: $C_{mx}\frac{dv_{gb}}{dt}$.

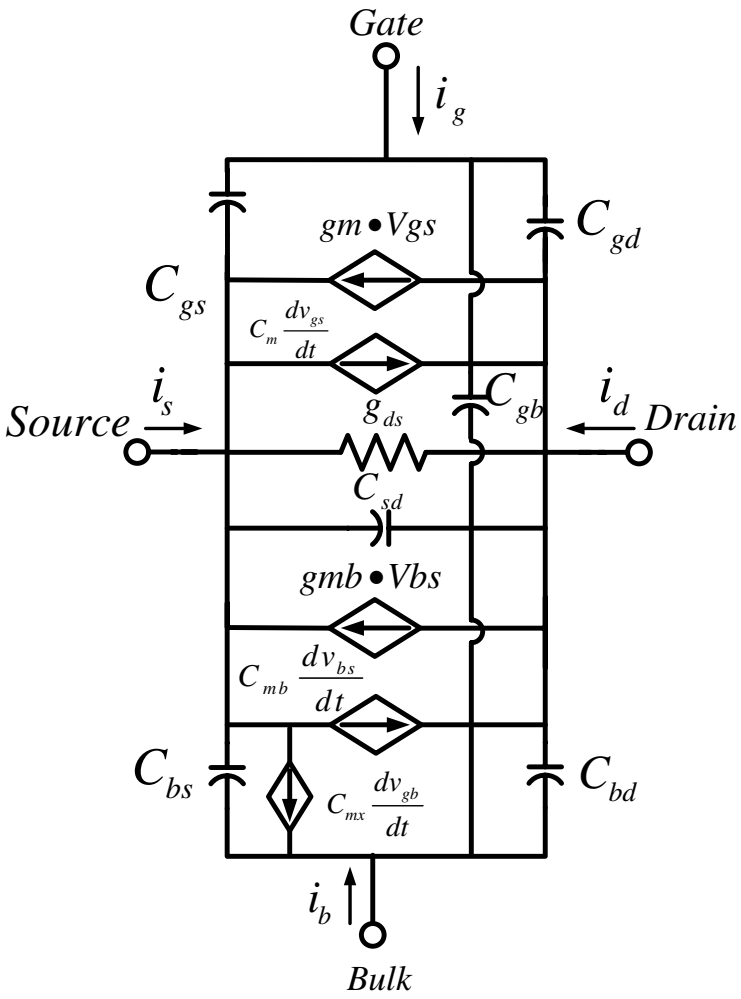

Figure 9 A complete Quasi-static small signal model

It is interesting to examine the response of this model to medium frequency signals. At sufficiently low frequencies, the currents due to the dependent sources ($C_m \frac{dv_{gs}}{dt}$ and $C_{mb} \frac{dv_{bs}}{dt}$) will be small compared to the drain transport currents produced by the $g_m v_{gs}$ and $g_{mb} v_{bs}$ current sources. Consequently, these two elements can be neglected. Moreover, at low frequencies the source-to-drain capacitance C_{sd} can be neglected since its charging current will be much smaller than the transport current conducted by g_{sd}. Finally, in the strong inversion region, the trans-

capacitance C_{mx} is very small and its contribution to the overall charging currents is negligible [10]. Hence the complete quasi-static model in Figure 9 simplifies to that in Figure 7 at medium frequencies. The effect of each terminal on the other is not generally symmetric. Intuitively, the symmetry exits only when the transistor is in strong inversion with the DC drain-to-source voltage set to zero. At this bias condition the trans-capacitors reduce to zero. However, in the saturation region, the device is not symmetric anymore and the trans-capacitors play an important role in the model. To show the asymmetric property, consider an experiment where the drain, source, and bulk terminals are *ac* shorted and the effect of the gate on the drain current is to be examined. From Figure 9, it is easy to show that the drain current is given by:

$$\begin{aligned} i_d\Big|_{v_d=v_s=v_b=0} &= g_m v_g - C_{gd}\frac{dv_g}{dt} - C_m\frac{dv_g}{dt} \\ &= g_m v_g - C_{gd}\frac{dv_g}{dt} - \left(C_{dg} - C_{gd}\right)\frac{dv_g}{dt} \\ &= g_m v_g - C_{dg}\frac{dv_g}{dt}. \end{aligned} \tag{2-43}$$

Notice that the gate terminal affects the drain current, i_d, through a transport component equal to $g_m v_g$ and a charging component $C_{dg}\dfrac{dv_g}{dt}$. The capacitive effect of the *gate* on the *drain* which is represented by C_{dg} and is completely different from the capacitive effect of the drain on the gate symbolically represented by C_{gd}. As a matter of fact, the latter effect is approximately zero in the saturation region since at this bias point the channel end at the drain terminal is pinched-off and the drain has no control whatsoever on the device charges or current (ignoring channel length modulation and two-dimensional charge sharing effects). It is instructive to note from equation (2-43) that the capacitive effect of the gate on the drain tends to decrease the drain current from

its transport value if the gate voltage is increasing with time. This unexpected behavior can be explained as follows. Increasing the gate voltage will demand an increase in the inversion layer charge density. To accommodate this increase, the number of electrons per unit time entering the source terminal increases above its steady state rate while the rate of electrons leaving the drain terminal decreases from its steady state. Hence the *total* drain current (entering the device) decreases due to this action. It is instructive to compare this charging action to that obtained in the analysis done in section 2.2.1 where the large signal drain current was examined under the quasi-static approximation. The unrealistic negative drain current show in Figure 2 can be compared to the negative charging effect described above.

Now consider applying a small signal drain voltage while ac shorting the gate, source, and bulk terminals. The gate current in this configuration will simply be:

$$i_g = -C_{gd}\frac{dv_d}{dt}. \tag{2-44}$$

It is clear from equations (2-43) and (2-44) that the effect of the drain and the gate terminals on each other is very different.

For long channel devices, the capacitance C_{sd} has a negative value in the linear region and goes to zero in the saturation region [10]. This behavior can be explained qualitatively as follows: Increasing the drain voltage when the transistor operates in the linear region will decrease the inversion layer charge. Consequently, the fraction of this charge allocated to the source terminal will decrease. Since the charge in our case is composed of electrons, the change in the source charge is positive ($\Delta Q_s > 0$). Using the capacitance definition in equation (2-33), a negative source to drain capacitance will result. In the saturation region, the drain has a very

negligible control on the device charges resulting in a zero drain-to-source capacitance. This analysis is in agreement with measurements [1].

Augmenting the model of figure 7 with the four capacitances $C_{sd}, C_m, C_{mb},$ and C_{mx} extends the limit of validity of the QS model to high frequencies. However, the improvement depends on the region of operation for the transistor and on the terminals being considered. For instance, the drain-to-source capacitance C_{sd} is maximum when the drain to source bias V_{DS} is zero while it drops to zero in the saturation region as mentioned above. Therefore, it is at $V_{DS}=0$ where the error between the two models of Figure 7 and Figure 9 is maximum with respect to the drain-to-source action. On the other hand, the trans-capacitance C_m is maximum in the saturation region and goes to zero when the transistor operates in the linear region. Hence, it is in the saturation region that the two models differ as far as the gate-to-drain action is concerned. Since the transistor operates in saturation for analog and RF applications, the complete Quasi-static model in Figure 9 extends the region of validity for the model in Figure 7 through the use of the trans-capacitance. It has been shown that this model although based on the quasi-static assumption will predict the device response to time varying signals up to about $w_0/3$ where w_0 is defined in (2-24) [4]. It should be mentioned that although this model extends the region of validity, it can still give physically unacceptable results *worse* than that of the model in Figure 7 if this region is exceeded. We will now elaborate more on this point. We found out that the drain current resulting from applying a gate voltage can be expressed as in equation (2-43) and repeated here in the frequency domain:

$$i_d\Big|_{v_d=v_s=v_b=0} = g_m v_g - jwC_{dg}v_g \,. \tag{2-45}$$

The drain current expression in the above equation contains a right-half-plane zero! which can be evaluated by setting the expression of the drain current to zero, yielding:

$$w_z = \frac{g_m}{C_{dg}}. \tag{2-46}$$

Hence beyond this frequency, the drain current, as predicted by the complete quasi-static model increases. This is physically unacceptable and is clearly false. To show this effect graphically we again use the Bsim3v3.1 complete QS model. The circuit shown in Figure 10 is a single transistor having a width of $20\,\mu m$ and channel length of $1\mu m$. The gate-to-source voltage is set to 1V (to ensure strong inversion) and the drain voltage is 3.3 V keeping the device in saturation. For the given bias condition and device dimensions, the DC analysis using Pspice shows that $g_m = 1.2mS$ and $C_{dg} = 5.49\,fF$. Using equation (2-46), the right hand plane zero was found to be around 34.7 GHz. To prove that this zero indeed exists in the drain current of the Bsim3v3.1 complete QS model, an *AC* signal of 1mv amplitude is superimposed on the DC gate voltage and the frequency was swept from 0 to 50 GHz. The circuit schematic and the magnitude of the ac drain current are shown in Figures 10 and 11 respectively. Notice that the drain current increases with frequency and that the 3-dB frequency is around 34.27 GHz confirming the theoretical calculations. Also note that the small signal drain current at the lower end of the spectrum is equal to its transport value:

$$i_d\big|_{DC} = g_m v_g = 1.2\mu A. \tag{2-47}$$

Since the gate transconductance g_m is inversely proportional to the gate length and C_{dg} is directly proportional to *L*, equation (2-46) indicates that the location of w_z is inversely proportional to L^2. Consequently, the problem of the zero becomes more severe for long channel devices.

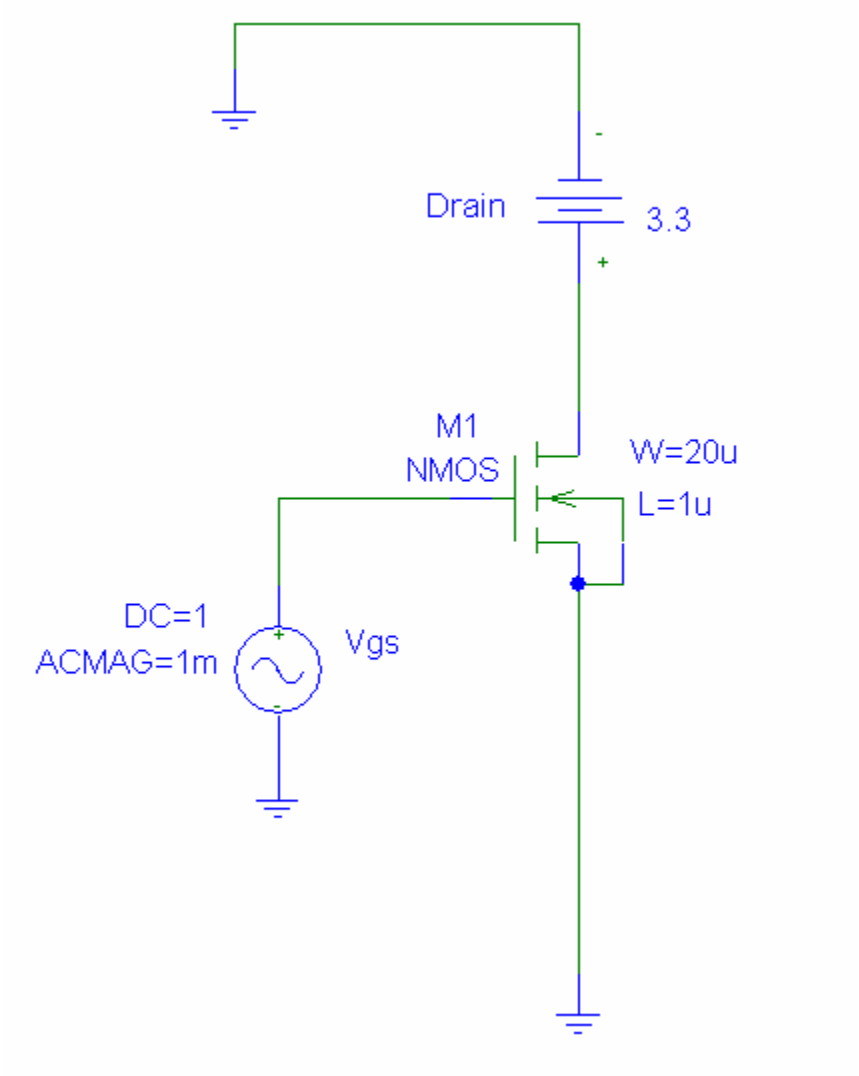

Figure 10 Circuit schematic used to simulate the frequency response of the drain current for the Bsim3v3.1 complete QS model

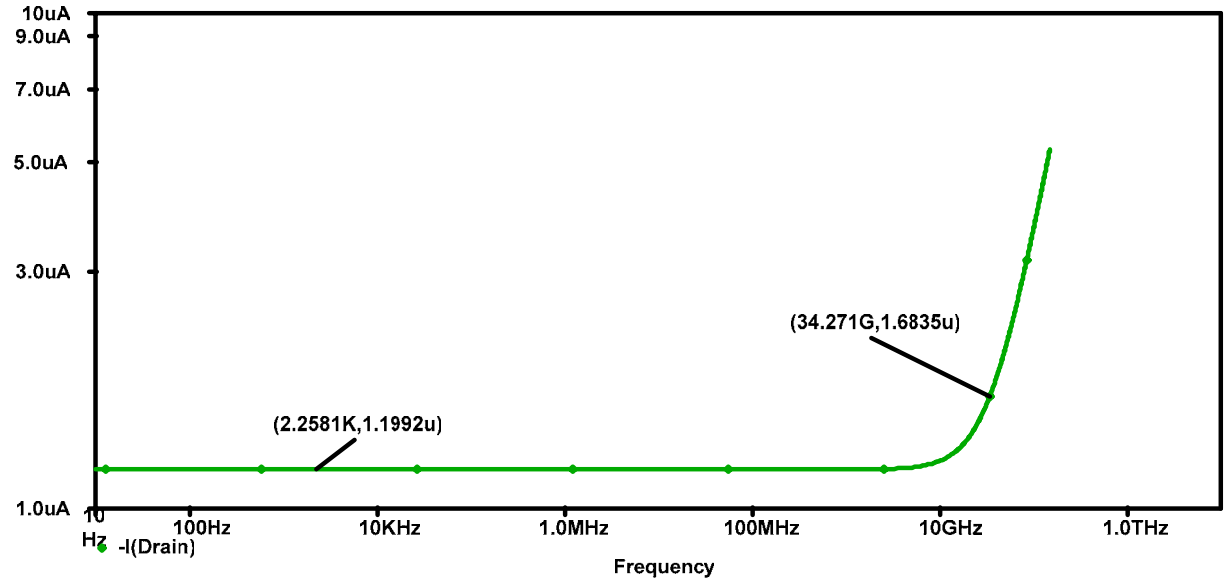

Figure 11 The magnitude response of the ac drain current generated by the Bsim3v3.1 complete QS model

2.5 Y-PARAMETER MODELING

Y-parameters are frequently used in the design of high frequency circuits because they are easy to derive from the S-parameters and are also useful in the computer analysis of circuits. For example, the overall y-parameters of two-port networks connected in parallel is simply the summation of the individual y-parameters of each network. Whether the transistor is a HEMT, BJT, MESFET, or a MOS transistor the y-parameter model is independent of the physics of the transistor and takes a general form that can be applied to any three or four terminal structure. Actually the transistor can be looked at as a black box with the given terminals. Since the MOS transistor is a four terminal device we will present the y-parameter of a four terminal structure. The model will prove helpful in developing the small-signal non-quasi-static model. The derivation of the model will be briefly presented. Consider an MOS transistor that is driven by dc

and small-signal voltages at each terminal. For convenience, the small-signal voltages are assumed to be sinusoidal and to have the same angular frequency. At steady state, the resulting currents will also be sinusoidal and of the same frequency. To simplify the analysis, the phasor representation of each small-signal voltage or current is used. The small signal equivalent circuit of the transistor driven by the phasor voltages is shown in Figure 12.

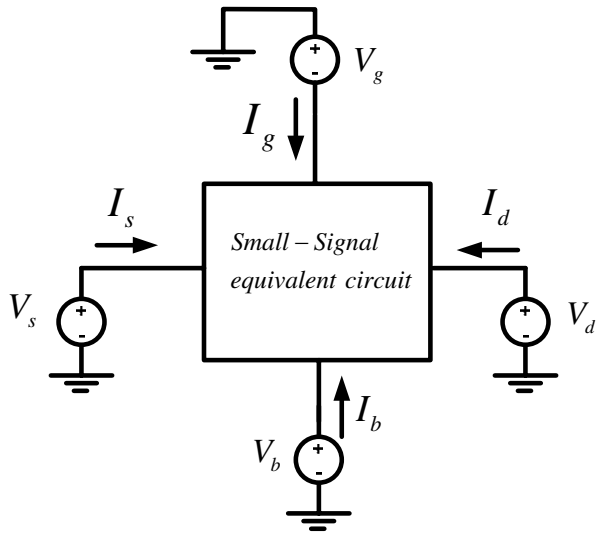

Figure 12 A setup for deriving the Y-parameter model

It is straightforward to show that the following three independent equations are needed to characterize the behavior of the transistor:

$$I_d = -y_{gd}V_{dg} - y_{sd}V_{ds} - y_{bd}V_{db} + y_m V_{gs} + y_{mb}V_{bs} \tag{2-48}$$

$$I_g = -y_{gd}V_{gd} - y_{gb}V_{gb} - y_{gs}V_{gs} \tag{2-49}$$

$$I_b = -y_{bd}V_{bd} - y_{gb}V_{bg} + y_{mx}V_{gb} - y_{bs}V_{bs} \tag{2-50}$$

where:

$$y_m = y_{dg} - y_{gd} \tag{2-51}$$

$$y_{mb} = y_{db} - y_{bd} \tag{2-52}$$

$$y_{mx} = y_{bg} - y_{gb}\,, \tag{2-53}$$

and:

$$y_{xy} = \frac{\partial I_x}{\partial V_y}\Big|_{V_z=0, z\neq y}\,. \tag{2-54}$$

Equations (2-48), (2-49), and (2-50) can be represented by the equivalent circuit as shown in Figure 13 and can be verified by writing Kirchhoff's current law equations for the drain, gate, and bulk terminals. It is instructive to compare the y-parameter model to that in Figure 9. Actually, the latter is a special case of the former. By comparing the two figures, the small signal y-parameters can be expressed as:

$$-y_{gd} = jwC_{gd} \tag{2-55a}$$

$$-y_{gs} = jwC_{gs} \tag{2-55b}$$

$$-y_{bd} = jwC_{bd} \tag{2-55c}$$

$$-y_{bs} = jwC_{bs} \tag{2-55d}$$

$$-y_{gb} = jwC_{gb} \tag{2-55e}$$

$$-y_{sd} = g_{sd} + jwC_{sd} \tag{2-55f}$$

$$y_m = g_m - jwC_m \tag{2-55g}$$

$$y_{mb} = g_{mb} - jwC_{mb} \tag{2-55h}$$

$$y_{mx} = -jwC_{mx}\,. \tag{2-55i}$$

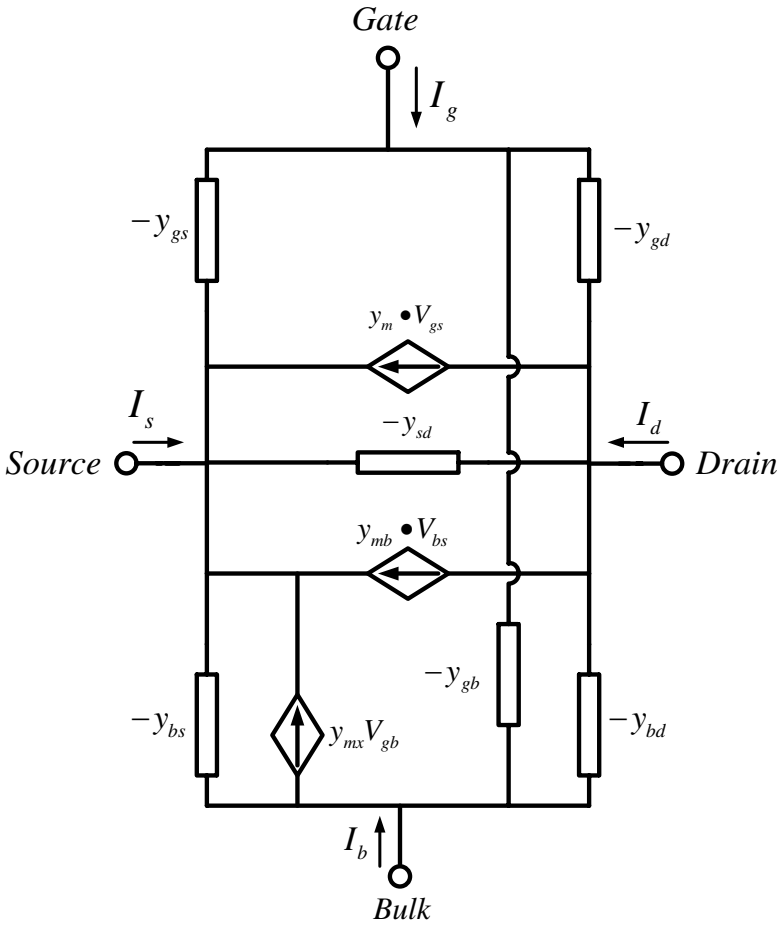

Figure 13 Y-parameter model

Notice that all the y-parameters in equation (2-55) have imaginary parts that vary linearly with frequency. As a matter of fact, the capacitance values can be determined experimentally as the constant of proportionality in the imaginary part of the measured admittance [1]. Moreover, the three y-parameters $y_m, y_{mb},$ and $-y_{sd}$ have a constant, real, positive part independent of frequency. The above observations hold true as long as the angular frequency w is less than $w_0/3$ where w_0 is defined in equation (2-24) [1]. As the frequency increases above that limit, the behavior is found to be different from what is predicted above. We know now that this is true

because the model of Figure 9 is based on the quasi-static assumption which is not adequate to predict the behavior of the MOS transistor at high frequencies. For example, it will be shown later that both the real and imaginary part of y_m decrease in magnitude with frequency and that -y_{gs} will not be purely imaginary (capacitive) but will have a nonzero real part. This observation can be explained by developing a non-quasi-static model which is the topic of the following section.

2.6 A SMALL SIGNAL NON-QAUSI STATIC MODEL

Without embarking on a detailed derivation of the small signal non-quasi-static model let us anticipate the results qualitatively. The NQS large signal analysis of the MOS transistor was described in section 2.2. It was found out that if the rise time of the gate signal is much smaller than the channel transit time, there will be a non zero delay between the cause (an increase in the gate voltage) and the effect (an increase in the drain current). However if the rise time is greater than $20\tau_0$ (a slow signal), the quasi-static model yields accurate results and can be used with minimum error. A similar behavior is observed in the small-signal domain where the corresponding figure of merit is w_0. Consider the experiment where the charging effect of the source on the gate was examined. This effect was represented by a capacitor C_{gs}. This implies that this charging mechanism is instantaneous in that the channel charge responded to the increase in the source voltage with no delay and the gate charge changed correspondingly. In other words, the electrons are assumed to have an infinite velocity. At high frequencies this is not true. If the source voltage is varying fast, then the finite time it takes the inversion layer charge to

respond to this variation has to be considered. This finite time results from the fact that electrons indeed move with a finite speed in the channel. We will now relate the speed of the electron to the channel resistivity. Assuming no velocity saturation, the velocity of the electron in the inversion layer can be expressed as:

$$v_d = \mu_n \vec{E} \approx \frac{1}{qn\rho} \vec{E} \tag{2-56}$$

where $\vec{E}$ is the electric field along the channel, n is the number of electrons per unit volume in the inversion layer, and ρ is the channel resistivity. Assuming that the electrons respond with infinite velocity is equivalent to assume that the channel resistivity is zero as can be seen from equation (2-56). Hence it can be concluded that the quasi-static formulation ignores the effect of the channel resistance on the frequency response. At high frequency, this resistance is distributed along the channel and forms a bias dependent distributed RC network with the oxide capacitance as depicted in Figure 14.

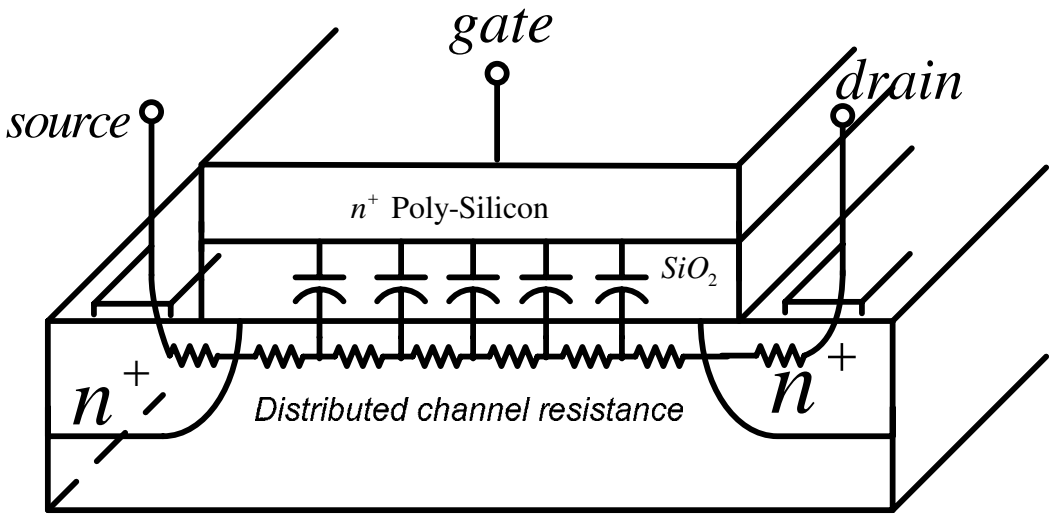

Figure 14 The oxide capacitance and channel resistance form a distributed RC network

The equivalent circuit or path between the source and the gate is therefore not purely capacitive, as predicted by the QS formulation, but should also have a resistive component. As a matter of fact the QS model unrealistically ignores the distributed channel resistance and split the gate capacitance between the source and drain terminals as seen from the small signal quasi-static models in Figures 7 and 9. The inversion layer charge density depends on the surface potential at the silicon silicon-dioxide interface with respect to the bulk. In strong inversion, the surface potential is approximately equal to:

$$\psi_s(x,t) \approx v_{CB}(x,t) + 6\phi_t \,. \tag{2-57}$$

The distributed nature of the channel resistance which is capacitively coupled to the gate through the oxide capacitance will indeed model the finite time it takes for the surface potential to reach its steady state when the terminal voltages change fast. Modeling the channel gate path as an infinite distributed RC network is almost identical to dividing the MOS transistor along its channel length into infinitesimally small sections as was done in section 2.2. From the above analysis it is clear that the inversion layer charge will not respond instantaneously and the delay will be captured by including the effect of the channel resistance in the model. A first order approximation would be to model the gate to source path by a capacitor in series with a resistance. Obviously this resistance should be proportional to the channel resistance with the constant of proportionality being less than unity. Referring to the y-parameter model, this indicates that the admittance $-y_{gs}$ contains a constant, real positive component proportional to the channel resistance in series with C_{gs}. Note that this admittance models the charging effect of the source on the gate terminal and the presence of the resistance indicates that the change in the charge on the capacitor plates is not instantaneous but lags behind its cause (change in source

potential). Non-quasi static effects can significantly influence the behavior of both short- and long- channel devices as verified experimentally [27]. The same analysis applies for the equivalent circuit between any two terminals that are connected through channel resistance like the drain-to-bulk, gate-to-drain paths, etc.

It is also instructive to note that as the frequency increases, the inversion layer charge is affected both in its magnitude and phase. Obviously, the magnitude of the charge will be smaller than that predicted by the QS formulation since it will not have enough time to reach the steady state value and the delay is modeled by a negative phase shift. One way to model the transistor at frequencies where the QS formulation is no longer valid is to divide the transistor into infinitesimally small sections, each section being small enough to be modeled quasi-statically as was done in section 2.2. The only difference here is in the terminal voltages. The transistor is first biased in the strong inversion region. The analysis requires determining the magnitude and phase of the small-signal terminal currents in response to small-signal voltages applied at the transistor terminals. The mathematical details will not be shown here. The reader is referred to literature for more information [8], [28], [29]. We are in particular interested in the resulting expressions of the admittances in the y-parameter model shown in Figure 13 which can generally be expressed in the following form:

$$y_{kl} = \frac{N_{kl}(w)}{D(w)}, \tag{2-58}$$

where, $N_{kl}(w)$ and $D(w)$ are expressed as infinite series in jw, namely:

$$N_{kl}(w) = n_{klo} + (jw)\, n_{kl1} + (jw)^2\, n_{kl2} + \ldots \tag{2-59}$$

$$D(w) = d_o + (jw)\, d_1 + (jw)^2\, d_2 + \ldots \tag{2-60}$$

The coefficients in these infinite series up to second order as well as for $N_{kl}(w)$ where $(k,l=d,g,b)$ are given in Appendix N Ref. [10]. Using these coefficients, the expressions for all the parameters in the model of Figure 13 can be computed using equation (2-58) to any desired accuracy by keeping an appropriate number of terms in the infinite series. In order to compare the NQS model with the complete and incomplete Quasi-static models in Figure 7 and Figure 9 respectively, the y-parameters can expressed in the following convenient form [10]:

$$-y_{gs} = jwC_{gs}\frac{1+jw\tau_2+...}{1+jw\tau_1+...} \tag{2-61a}$$

$$-y_{bs} = jwC_{bs}\frac{1+jw\tau_2+...}{1+jw\tau_1+...} \tag{2-61b}$$

$$-y_{gd} = jwC_{gd}\frac{1+jw\tau_3+...}{1+jw\tau_1+...} \tag{2-61c}$$

$$-y_{bd} = jwC_{bd}\frac{1+jw\tau_3+...}{1+jw\tau_1+...} \tag{2-61d}$$

$$-y_{gb} = jwC_{gb}+\frac{(jw)^2 C_{gb,sat}\tau_4+...}{1+(jw)\tau_1+...} \tag{2-61e}$$

$$-y_{sd} = \frac{g_{sd}}{1+jw\tau_1+...} \tag{2-61f}$$

$$y_m = \frac{g_m}{1+jw\tau_1+...} \tag{2-61g}$$

$$y_{mb} = \frac{g_{mb}}{1+jw\tau_1+...} \tag{2-61h}$$

$$y_{mx} = 0 \tag{2-61i}$$

where $\tau_1,\tau_2,\tau_3,$ and τ_4 are bias dependent time-constants inversely proportional to w_0. In the saturation region, these parameters can be expressed as [10]:

$$\tau_1 = \frac{4}{15}\frac{1}{w_0} \tag{2-62a}$$

$$\tau_2 = \frac{2}{15}\frac{1}{w_0} \tag{2-62b}$$

$$\tau_3 = \frac{5}{30}\frac{1}{w_0} \tag{2-62c}$$

$$\tau_4 = \frac{4}{15}\frac{1}{w_0}. \tag{2-62d}$$

At very low frequencies and in particular when $w << w_0$, the y-parameters in equation (2-61) simplify to: $-y_{gs} \approx jwC_{gs}$, $-y_{bs} \approx jwC_{bs}$, $-y_{gd} \approx jwC_{gd}$, $-y_{gb} \approx jwC_{gb}$, $-y_{sd} = g_{sd}$, $y_m = g_m$, $y_{mx} = 0$. Hence the NQS model simplifies to that in Figure 7 for very low frequencies. In order to compare the NQS model with that in Figure 9 we will now consider a simplified first order NQS model. Consider the gate-to-drain admittance. If the frequency of operation is such that $w\tau_3 << 1$, then equation (2-61c) can be expressed as:

$$-y_{gd} \approx jwC_{gd}\frac{1}{(1+jw\tau_1)(1-jw\tau_3)}$$

$$\approx jwC_{gd}\frac{1}{1+jw(\tau_1-\tau_3)} \tag{2-63}$$

where the high-order terms are neglected for simplicity. The three parameters $-y_{bs}, -y_{bd}$, and $-y_{gs}$ can be arranged in the same manner. The higher order terms are neglected in the y- parameters of equations (2-61e→ 2-61g). The first order y-parameters can then be expressed as:

$$-y_{gs} \approx \frac{jwC_{gs}}{1+jw(\tau_1-\tau_2)} \tag{2-64a}$$

$$-y_{bs} \approx \frac{jwC_{bs}}{1+jw\left(\tau_1-\tau_2\right)} \tag{2-64b}$$

$$-y_{gd} \approx \frac{jwC_{gd}}{1+jw\left(\tau_1-\tau_3\right)} \tag{2-64c}$$

$$-y_{bd} \approx \frac{jwC_{bd}}{1+jw\left(\tau_1-\tau_3\right)} \tag{2-64d}$$

$$-y_{gb} \approx jwC_{gb} \tag{2-64e}$$

$$-y_{sd} \approx \frac{g_{sd}}{1+jw\tau_1} \tag{2-64f}$$

$$y_m \approx \frac{g_m}{1+jw\tau_1} \tag{2-64g}$$

$$y_{mb} = \frac{g_{mb}}{1+jw\tau_1} \tag{2-64h}$$

$$y_{mx} = 0 \tag{2-64i}$$

Given that we expect qualitatively that the equivalent circuit between any two nodes should contain a resistive component, the form of the y-parameters in equations (2-64a → 2-64d) is appealing. The equations can be modeled by this resistance in series with a capacitor. Consider for example the admittance $-y_{gd}$ between the gate and the drain terminals. Figure 15 shows a simple RC circuit that models this admittance. It is easy to show that the gate-to-drain resistance is equal to:

$$R_{gd} = \frac{\tau_1-\tau_3}{C_{gd}}. \tag{2-65}$$

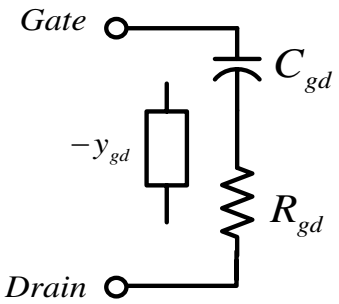

Figure 15 A first order NQS gate-to-drain equivalent circuit

Following the same approach, equations (2-64-a), (2-64-b), and (2-64-d) resulting from a first order NQS analysis can be represented by a similar circuit, where:

$$R_{bd} = \frac{\tau_1 - \tau_3}{C_{bd}} \tag{2-66}$$

$$R_{gs} = \frac{\tau_1 - \tau_2}{C_{gs}} \tag{2-67}$$

$$R_{bs} = \frac{\tau_1 - \tau_2}{C_{bs}} \tag{2-68}$$

It is straight forward to prove that $-y_{sd}$ can be realized by a resistance $1/g_{sd}$ in series with an inductor L_{sd} such that:

$$L_{sd} = \frac{\tau_1}{g_{sd}} \tag{2-69}$$

To summarize, we have managed starting from Figure 13 and using a first order analysis to derive a first order non-quasi-static model shown in Figure 16. The need to include a resistance in series with each capacitor between each two nodes of the transistor was explained at the beginning of this section (2.4.1). In general, all capacitors in the small-signal models are

proportional to the oxide capacitance C_{ox}. Moreover, the time constants τ_i, are inversely proportional to w_o introduced in section 2.3.1 and repeated here for convenience:

$$w_o = \frac{\mu(V_{GS} - V_T)}{L^2} \tag{2-70}$$

Consequently, the four resistances in the NQS model will be proportional to:

$$R_{gs,gd,bd,bs} \propto \frac{1}{w_0 C_{ox}^{'} WL} \propto \frac{L^2}{\mu(V_{GS} - V_T) C_{ox}^{'} WL} \propto \frac{1}{\mu C_{ox}^{'} \frac{W}{L}(V_{GS} - V_T)} \tag{2-71}$$

For long channel devices and in the absence of velocity saturation, it is easy to show that the gate transconductance g_m can be expressed as:

$$g_m = \mu C_{ox}^{'} \frac{W}{L}(V_{GS} - V_T) \tag{2-72}$$

which ultimately means that the four resistances are inversely proportional to the gate transconductance as follows from equation (2-71). As a matter of fact, it can be shown that the gate-to-source resistance (R_{gs}) for a long channel transistor biased in the strong inversion saturation region to be equal to $1/5 g_m$ [10]. But, it was mentioned before that these resistances should be proportional to the channel resistance. To overcome this quandary, we will derive an expression for the channel resistance when the transistor is biased in the strong inversion region. In strong inversion, the drain current is dominated by the drift component and can be written as:

$$I_{DS} = -\mu W Q_I^{'} \frac{dV_{CB}(x)}{dx} \tag{2-73}$$

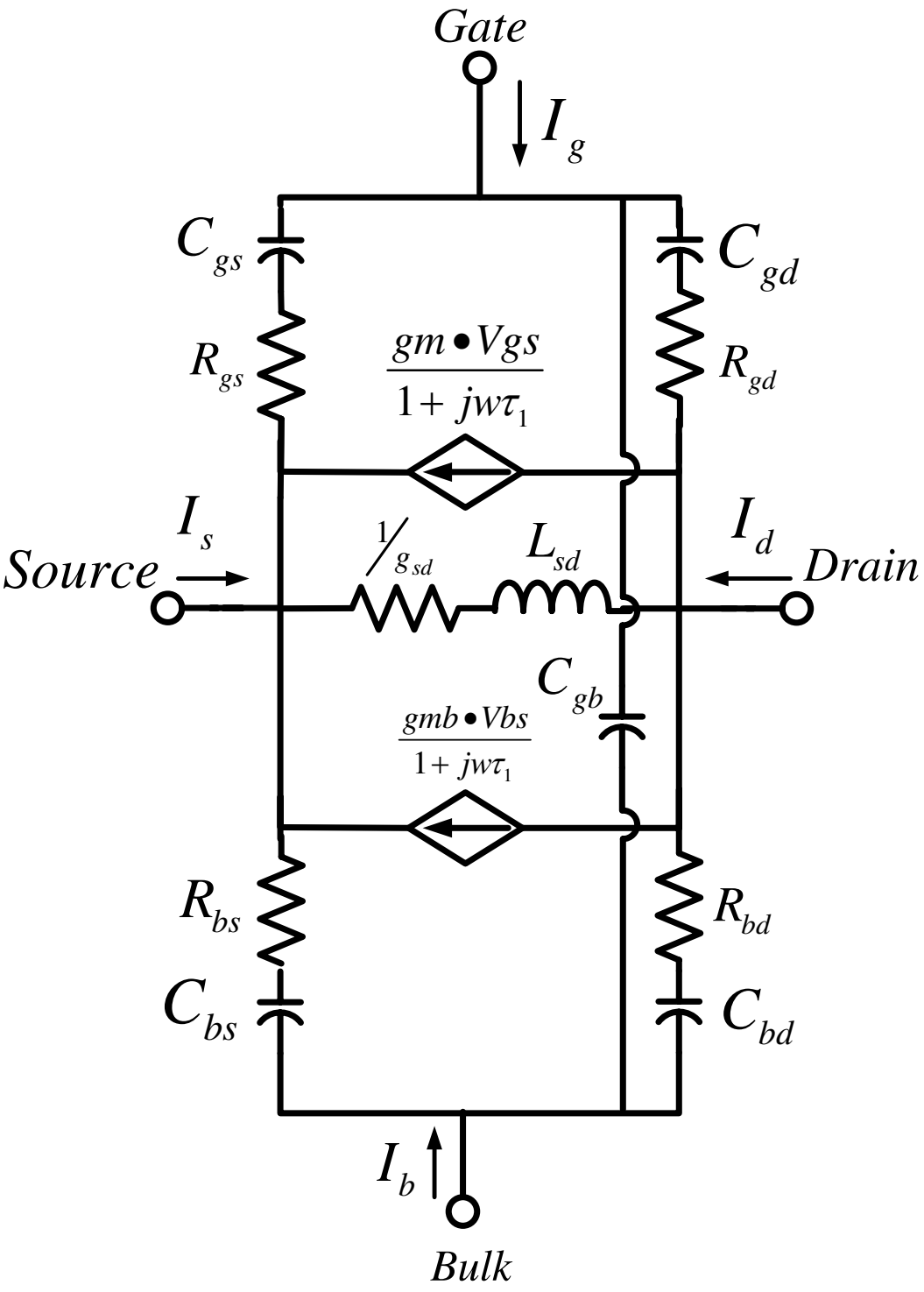

Figure 16 A first order Non-quasi-static model

A small element of the channel of length dx, has a resistance dR which can be expressed as:

$$dR = \frac{dV_{CB}(x)}{I_{DS}} = \frac{-dx}{\mu W Q_I^{'}(x)} \tag{2-74}$$

In general the inversion layer charge per unit area ($Q_I^{'}(x)$) varies along the channel length and decreases in absolute value going from the source to the drain. Expressing $Q_I^{'}(x)$ is not an easy task. We will therefore consider the case when the transistor is in the strong non-saturation region in particular with V_{DS}=0V. At this bias point, the inversion layer charge per unit area is uniform along the channel, and can be written as:

$$Q_I^{'} = C_{ox}^{'}\left(V_{GS} - V_T\right) \tag{2-75}$$

The channel resistance will also be uniform along the channel. Integrating equation (2-74) along the channel length, it is straightforward to show that:

$$R_{ch} = \frac{1}{\mu \frac{W}{L} C_{ox}^{'}\left(V_{GS} - V_T\right)} \tag{2-76}$$

which is nothing but the inverse of the gate transconductance g_m given in equation (2-72). At high frequency, this resistance is distributed along the channel and forms a bias dependent RC network with the gate capacitance. It is instructive to note that the channel resistance is bias dependent as seen from equation (2-76). It can therefore be concluded that the input impedance of an MOS transistor is not purely capacitive as predicted by the small signal QS models but that it contains a resistive component that is bias dependent. In addition to the channel resistance, a distribute gate resistance along the channel resulting from the finite resistivity of the polysilicon gate material will also affect the input impedance as well as other device parameters as will be seen in Chapter 3.

The presence of the inductor L_{sd} in Figure 16 should not be surprising. A key property for the inductor is that the current flowing through it cannot change instantaneously. Recall that g_{sd} models the amount of source current that would flow if the drain voltage is increased. The change in the current and hence the inversion layer charge is not instantaneous and that is where the inductor comes into play. It simply models the inertia of the channel to respond to a fast drain signal. The two dependent current sources in Figure 16 are also a function of frequency. The gate transconductance, g_m, decreases with frequency at the rate of 20dB/dec for a first order NQS model. This intuitive result reflects the fact that at high frequency, the channel will not be able to adjust to the fast varying gate signal and that the drain current will not just lag in phase but will also decrease in magnitude. The same argument applies for the bulk transconductance g_{mb}. As the frequency of operation is decreased, the model of Figure 16 is expected to reduce to a complete QS model shown in Figure 9. This is because impedance of the capacitors in Figure 16 becomes much larger than the series resistance, thus reducing the RC equivalent circuits between each node to a simple capacitor. Moreover, at frequencies satisfying the inequality $w\tau_1 << 1$, the y-parameters $-y_{sd}$, y_m, and y_{mb} reduce to the following expressions:

$$-y_{sd} \approx g_{sd} - jw\tau_1 g_{sd} \tag{2-78a}$$

$$y_m \approx g_m - jw\tau_1 g_m \tag{2-78b}$$

$$y_{mb} \approx g_{mb} - jw\tau_1 g_{mb} \tag{2-78c}$$

where $\tau_1 g_{sd} = -C_{sd}$, $\tau_1 g_m = C_m$, and $\tau_1 g_{mb} = C_{mb}$. Hence the NQS model reduces to the complete QS model if $w\tau_1 << 1$, or in terms of w_0:

$$w << 3.75 w_0 \tag{2-79}$$

In the strong inversion saturation region, it can be demonstrated [10] that the small signal parameters of the model in Figure 16 (assuming no velocity saturation or channel length modulation (CLM)) have the following value: $C_{gd} = C_{bd} = 0$, $R_{bd} = R_{gd} = \infty$, and $L_{sd} = \infty \left(g_{sd} L_{sd} = \tau_1 \right)$. The first four parameters model the control of the drain terminal on the channel. However, in the saturation region the channel end is pinched-off and the drain has no control on the device characteristics given the above conditions. This means that the *intrinsic* path between the infinite series inductance is a consequence of the somehow unrealistic assumptions that have been adopted to derive the model such as the absence of channel length modulation as obvious from equation (2-69). However, at high frequencies the parasitic capacitances connected at the source and drain terminals from adjacent devices or from the extrinsic capacitances of the device itself will shunt the source to drain terminal and the current between these two terminals will flow in these capacitances since they constitute a low impedance branch to the current. Hence for practical cases, the series inductor is usually omitted from the small signal model.

It is obvious that the NQS model avoids the right-hand plane zero that appears in the expression for the drain current (equation 2-46) and that leads to an unrealistic increase in the drain current as shown in Figure 11. Considering the same analysis as was done when deriving equation (2-45), the drain current predicted by Figure 16 can be expressed as:

$$i_d \Big|_{v_d = v_s = v_b = 0} = \frac{g_m}{1 + j \, w / w_p} v_g \tag{2-78}$$

where:

$$w_p = \frac{1}{\tau_1} = 3.75 w_0 \tag{2-79}$$

We will now demonstrate the presence of this pole using the Bsim3v3.1 Non-quasi-static model which can be activated by setting the parameter nqsMod=1. The implementation of the NQS model in the Bsim3v3.1 was done by modifying the current equations rather than including the channel resistance explicitly in the small signal model as shown in Figure 16[30]. Adding elements will require the creation of additional nodes in the pre-existing Bsim3 model significantly increasing the simulation time to solve the Jacobian Matrix in the simulator. Moreover, adding new elements will change the device topology requiring a non trivial modification to the existing model equations. In this alternative approach, the effect of the channel resistance is included by introducing a new state variable Q_{def} to keep track of the amount of channel charge necessary to reach equilibrium at a given time:

$$Q_{def}(t) = \left| Q_{cheq}(t) - Q_{ch}(t) \right| \tag{2-80}$$

where, $Q_{cheq}(t)$ and $Q_{ch}(t)$ are the equilibrium channel charge and the instantaneous channel charge respectively. Note that the QS formulation assumes that these two terms are identical by virtue of assuming that the channel charge responds instantaneously to the applied voltages. $Q_{def}(t)$ is allowed to decay exponentially in the channel with a bias dependent NQS relaxation time τ that approximates the delay of the RC gate-to-channel distributed network and is evaluated using Elmore's approach when the transistor is in strong inversion. A simple sub-circuit is constructed to evaluate $Q_{def}(t)$ at a give time from which the charging currents is calculated. This method is very efficient in modeling the NQS effect without significantly increasing the simulation time and without modifying the pre-existing model. The setup used is as shown in Figure 10.The drain current is expected to decrease with increasing the signal

frequency and to lag behind the gate signal (negative phase shift). Figure 17 shows the magnitude and phase of the drain current versus frequency.

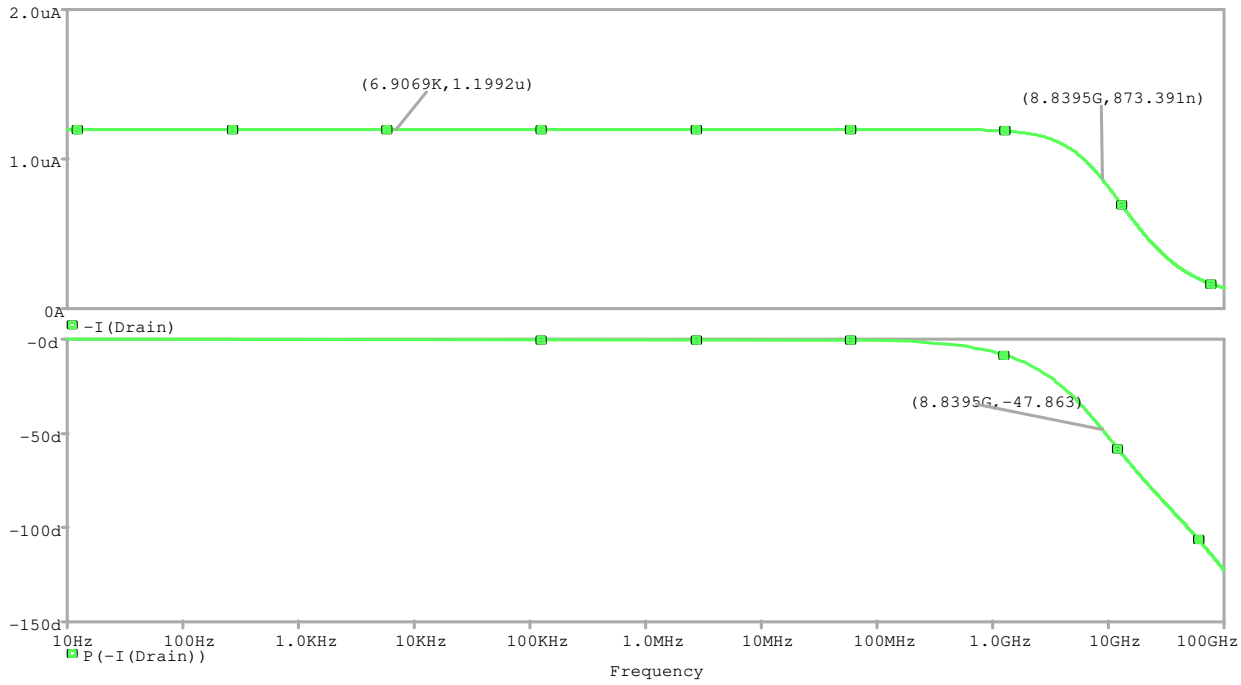

Figure 17 Magnitude and phase of the drain current as predicted by the NQS Bsim3v3.1 model

2.7 CONCLUSION

This chapter has introduced the basic principles of the quasi-static and non-quasi-static modeling. The QS approach assumes that the channel charge can adjust itself instantaneously and is only valid at low frequencies. A complete QS model was presented and the notion of trans-capacitance was introduced. The Bsim3v3.1 model was used to indicate the limitation of the QS model at high frequency. Ultimately, the complete QS model results in an unrealistic increase in the drain current with frequency. The region of validity of the small signal models is

proportional to w_0 (equation 2-24). It was concluded that the region of validity of the model depends quadratically on the channel length. For Digital signals, the QS model fails if the rise-time (fall-time) of the signal is much smaller or comparable to the carrier transit time. Unrealistic negative drain current spikes were observed when a step input was applied at the gate terminal. The NQS formulation was briefly introduced and it was shown that the channel resistance plays an important role at high frequencies. This resistance forms a distributed bias dependant RC network with the gate oxide and results in a signal delay between the transistor terminals. Non-quasi-static effects have been demonstrated to exist for both long and short channels and should be included in an RF Mosfet model. A first order NQS model was derived and discussed. Although this model extends the region of validity of the QS models and is accurate up to w_0, more elements should be added to enhance the accuracy at RF. These elements are related to the extrinsic parasitics and are discussed in chapter 3.

3.0 EFFECT OF EXTRINSIC ELEMENTS ON THE HIGH FREQUENCY PERFORMANCE OF MOSFETS

The behavior of the intrinsic part of the MOS transistor was examined in Chapter 2. It was concluded that at high frequency the QS assumption gives erroneous simulation results and should be abandoned. The NQS formulation was introduced and a first order small-signal NQS model was derived. It was noticed that the NQS model accounts for the finite channel charging time and simulates this physical phenomenon by incorporating the channel resistance effect on the charging mechanism. Essentially, the effect of each terminal on the device performance is not instantaneous, and the delay between the cause and effect is captured by a resistive element in series with each capacitor. At RF, the extrinsic components of the device play a prominent role in degrading the transistor performance and therefore have to be added to the intrinsic small signal NQS model. Six extrinsic capacitances are to be added: the gate-to-source capacitance C_{gsex}, the gate-drain-capacitance C_{gdex}, the gate-to-bulk capacitance C_{gbex}, the drain-to-bulk capacitance C_{dbex}, the source-to-bulk capacitance C_{sbex}, and the drain-to-source capacitance C_{dsex}. As for the parasitic resistances they can be divided into four parts: the resistance of the gate material, the substrate resistance, the resistance of the source and drain regions and their contacts. At high frequency the impedances of the capacitive components are comparable or even smaller than that of the resistive components that despite their distributive nature are represented by lumped elements in most models. The resistance of the drain and source regions

are less important than the other two resistances and are most of the time omitted for simplicity. However, an accurate model ought to include their effects. This chapter is mainly concerned with the effect of the gate and substrate resistances on the device operation at RF. In section 2, we consider the effect of the gate resistance and the methods used in the literature to include this element in the small signal model. Some of the issues will be proved using Pspice and Cadence tools. Section three discusses the effect of the substrate resistance at RF. The rest of the chapter is devoted to the discussion of some important parameters required for RF MOSFET modeling and to briefly introduce the scattering parameters, followed by a conclusion.

3.1 MODELING AND IMPACT OF THE GATE RESISTANCE ON THE DEVICE OPERATION AT RF

In some applications, the need for wide transistors is inevitable. For a given technology, the width of the device has to be increased to attain sufficient gate-transconductance for high gain amplifiers or to generate large current drive for driving large capacitances in a digital circuit. At low frequencies, the resistance of a strip of polysilicon material can be expressed as:

$$R_g = \frac{W}{L} R_{gsh} \tag{3-1}$$

where, W and L are the material dimensions and R_{gsh} is the polysilicon sheet resistance. At high frequency, this resistance is distributed along the transistor's width and forms a distributed RC network with the gate capacitance C_{ox}. To reduce simulation time and avoid computational complexity, it is useful to model the gate-distributed effect (GDE) along its width using lumped elements. In this approach, the transistor can be divided into *n* parallel-connected devices as

shown in Figure 18. Each device has a channel length L, a channel width W/n, a gate resistance R_g/n, a transconductance g_m/n and a gate capacitance C_g/n.

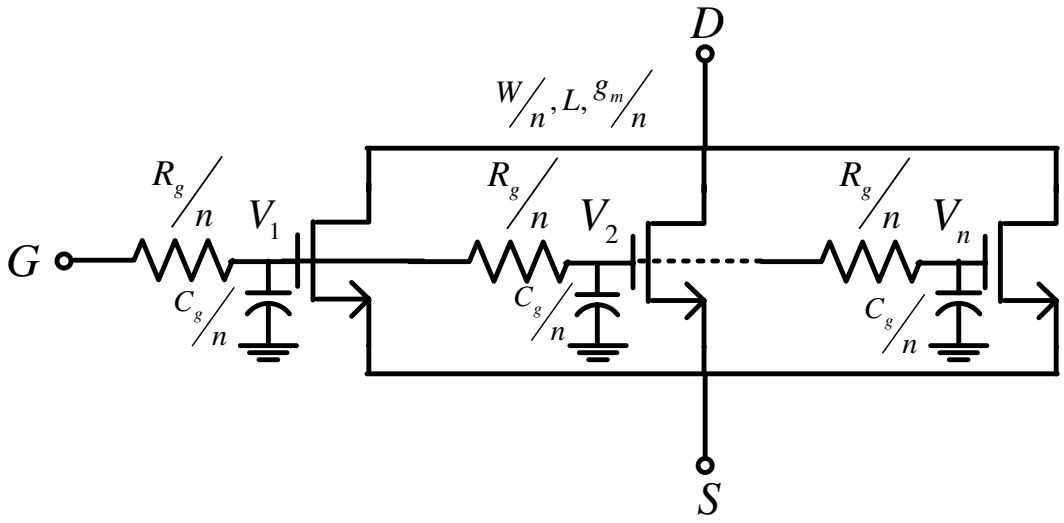

Figure 18 The gate resistance forms a distributed RC network with the oxide capacitance. The transistor can be divided into subsections to formulate the lumped equivalent gate resistance.

The effect of the gate distributed resistance on the cutoff frequency (w_T), the maximum oscillation frequency ($w_{\max}$), input referred thermal noise, and time response was investigated [31]. To avoid repeating the quantitative analysis, we will point out what we should expect intuitively from this approach. Consider for example applying a step voltage at the gate terminal in Figure 18. It is clear that voltage V_1 experiences less phase shift than V_2. More specifically, sub-transistors toward the left end "see" only a fraction of the gate resistance R_g, while those

near the right end experience the effect of most of R_g . Hence, we would expect that the lumped resistor modeling the time response of the GDE to be smaller than R_g . It was found that the unity gain cut-off frequency is independent of the distributed gate resistance. However, for noise and $w_{\max}$ calculations, as well as for time response analysis, a lumped effective resistor in series with the gate can approximate the distributed gate resistance. In particular:

$$R_{g_{eff}} = \frac{W}{\alpha L} R_{gsh} \tag{3-2}$$

where α is 3 if the gate terminal is connected on one side and 12 when connected on both sides. We will prove the above statement by performing transient circuit simulation. The ac drain current for the distributed structure in figure 18 has the following expression [31]:

$$I_D(s) = g_m \frac{\tanh\sqrt{R_g C_g s}}{\sqrt{R_g C_g s}} V_{in}(s) \tag{3-3}$$

where $V_{in}(s)$ in the input signal applied at the gate, g_m is the gate trans-conductance, and s stands for Laplace . In deriving equation (3-3), the transistor was assumed to be strongly inverted and in saturation and the drain was biased at V_{dd} . For typical range of frequencies we can assume that $sR_g C_g << 1$ which is a valid assumption for the current technology. Using the above assumption, the tanh function in equation (3-3) can be approximated by the first two terms of the Taylor series expansion:

$$\begin{aligned} \tanh\sqrt{R_g C_g s} &\approx \sqrt{R_g C_g s} - \frac{\left(\sqrt{R_g C_g s}\right)^3}{3} \\ &\approx \frac{\sqrt{R_g C_g s}}{1 + \frac{\left(\sqrt{R_g C_g s}\right)^2}{3}} = \frac{\sqrt{R_g C_g s}}{1 + \frac{R_g C_g s}{3}}. \end{aligned} \tag{3-4}$$

Thus the ac response of the drain current including the distributed effect of the gate resistance can be approximated by:

$$I_D(s) \approx \frac{g_m}{1+\frac{R_g C_g s}{3}} V_{in}(s). \tag{3-5}$$

Hence, the gate resistance influences the device frequency response if the driving circuit has an output impedance comparable to one-third the gate resistance. To prove the above result, we will compare the simulated ac drain current for three circuits. The first circuit consists of a single transistor biased at VGS = 1 v, VDS = Vdd = 3.3 v and has a channel width and length of $200\,\mu m$ and $1\,\mu m$ respectively. A 100 Ω resistor is attached to the gate. The second circuit resembles the distributed structure in figure 18 where n is chosen to be 10. Hence, the resistance attached to the gate of each sub-transistor is $10\,\Omega$. The third circuit consists of the lumped equivalent circuit of the distributed structure which from equation (3-5) is composed of a single transistor with a gate resistance equal to $33.33\,\Omega$. The three circuits were implemented in Pspice. An ac signal with an amplitude of 1V is applied the input of each circuit. The frequency of the signal was swept from 0 to 10GHz. The simulated drain current is shown in figure 19. Figure 19 clearly shows that a single lumped resistor that is equal to one-third the total gate resistance can model the distributed effect at RF. It is worth noting that the polysilicon gate resistance can be reduced significantly by dividing the width of the gate into multi-fingers. For a device with N-fingers, the effective polysilicon gate resistance is expressed as [32]:

$$R_{g_{eff}} = \frac{1}{N^2}\frac{W}{\alpha L} R_{gsh} \tag{3-6}$$

As a result, the effective gate resistance scales down quadratically with the number of fingers. Nonetheless, using multiple fingers degrade the circuit density and increase the source or drain sidewall capacitance. Moreover, current CMOS technologies use a silicide layer [33] (such

as $T_i S_{i2}$) that is deposited over the gate, drain, and source regions. Although the deposition of the silicide layer requires an additional processing step, the reduction in the terminal resistances is significant. Consequently, the transistor can be made to switch faster. The above two factors combined together lead to a substantial decrease in value of the effective polysilicon gate resistance. This result has an important implication; the real part of the input impedance of a transistor is mainly contributed by the distributed channel resistance, which for a first order NQS model is lumped into two resistors in series with the gate-to-source and gate-to-drain capacitances as shown in Figure 16.

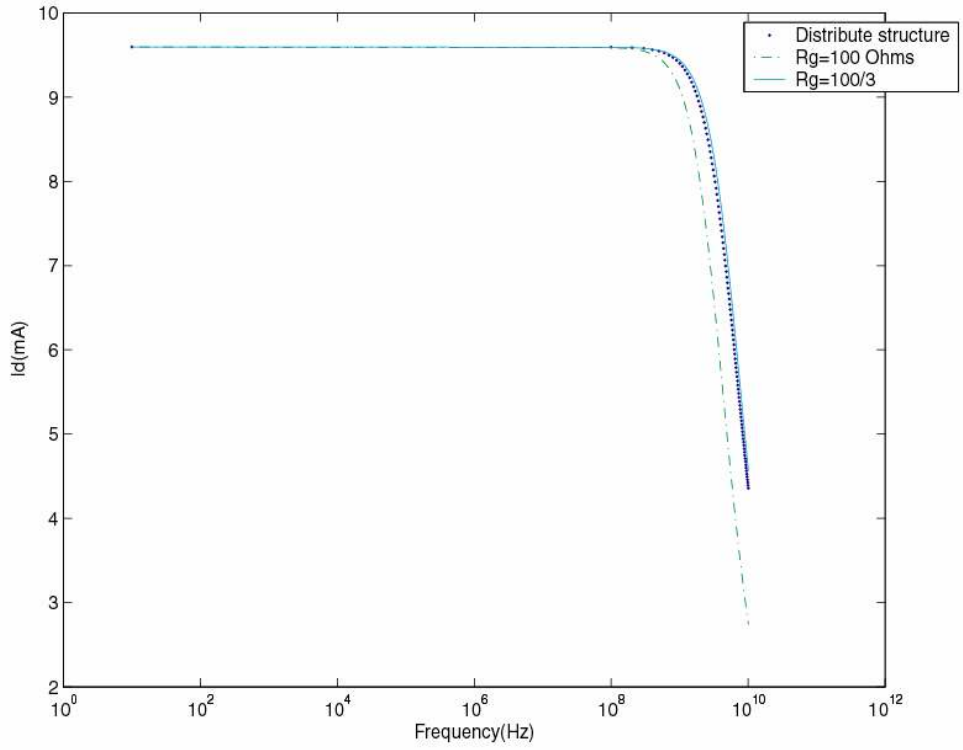

Figure 19 The frequency response of the drain current considering for a polysilicon gate resistance of 100Ω. The distributed structure of the gate resistance can be approximated by a lumped resistance one-third the DC value

3.2 CUT-OFF AND MAXIMUM OSCILLATION FREQUENCY

It was mentioned earlier that the cut-off frequency, w_T, is independent of the gate resistance even if the distributed effect of the gate material is taken into account. This result should not be surprising. This parameter is defined as the frequency at which the short circuit current gain of the transistor drops to unity. Obviously, this gain is infinite at DC since the gate at low frequency acts as an open circuit. As the frequency increases, the input current flows through the gate capacitance, develops a gate-to-source voltage that will subsequently induce a drain current through the gate conductance current source ($g_m v_{gs}$). The circuit configuration used to measure the cut-off frequency is shown in Figure 20. It can be shown that the cut-off frequency using the *QS* model for the transistor in Figure 20 is given by:

$$w_T \approx \frac{g_m}{C_{gs} + C_{gd} + C_{gb}} \approx \frac{\mu(V_{GS} - V_t)}{L^2} \tag{3-7}$$

Where g_m is the gate transconductance given by equation (2-72), μ is the carrier mobility in the channel, $V_{GS} - V_t$ is the overdrive voltage, and L is the channel length. It is worth noting that the region of validity of the small-signal models presented in chapter was proportional to w_o (2-24) which accidentally has the same expression as w_T derived above.

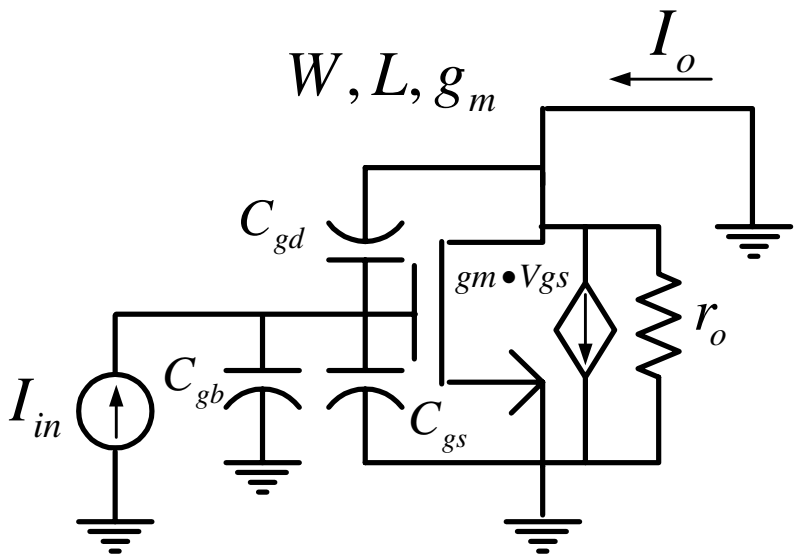

Figure 20 MOS transistor showing the small-signal elements needed to calculate the cutoff frequency w_T .

It is clear from Figure 20 that including the gate resistance will not affect the cut-off frequency since the input current is in series with this resistance. However, at high frequency, the gate resistance forms a distributed RC network with gate capacitance as shown in Figure 18 and yet the unity gain frequency is still given by equation (3-7) [31]. This result implies that the unity gain frequency is not affected by the inclusion of the gate resistance. It is also instructive to note that w_T result from a *magnitude* measurement of the current gain but ignores the phase information. However, it was noted in chapter 2 that a high frequency NQS model predicts a nonzero phase difference between the terminal voltages and currents. It can be therefore concluded that, w_T by itself is not sufficient for high frequency characterization.

Transistors are active devices since they have the ability to amplify the input power from a supply and deliver it to the load, i.e. they exhibit a power gain greater than unity. By excluding

the gate resistance one implies zero input power dissipation since in this case the gate of the transistor is purely capacitive. But since the transistor delivers a nonzero power to its load, an infinite power gain is predicted in this case. Obviously, this is not true and including the gate resistance will affect the power gain simply because resistive elements dissipate power. The power gain of a two port network (in our case it is a transistor) can be maximized if the input and output impedances of the transistor are the complex conjugates of the source and load impedances respectively. It is common to calculate the power gain from the input (gate) to the output (drain) of the transistor while excluding the feedback power from the output due the gate-to-drain capacitance C_{gd}. This is called the unilateral power gain (U) and the condition for its measurement is achieved by using an appropriate network between the gate and drain terminal to cancel the effect of the feedback signal through C_{gd} [34]. It can be shown that (U) can be expressed as a function of the y-parameters as follows:

$$U = \frac{\left|Y_{12} - Y_{21}\right|^2}{4\left(\mathrm{Re}\,Y_{11}\,\mathrm{Re}\,Y_{22} - \mathrm{Re}\,Y_{12}\,\mathrm{Re}\,Y_{21}\right)} \tag{3-8}$$

The small signal model used in [31] to calculate (U) is based on the QS assumption and is given in Figure 21.

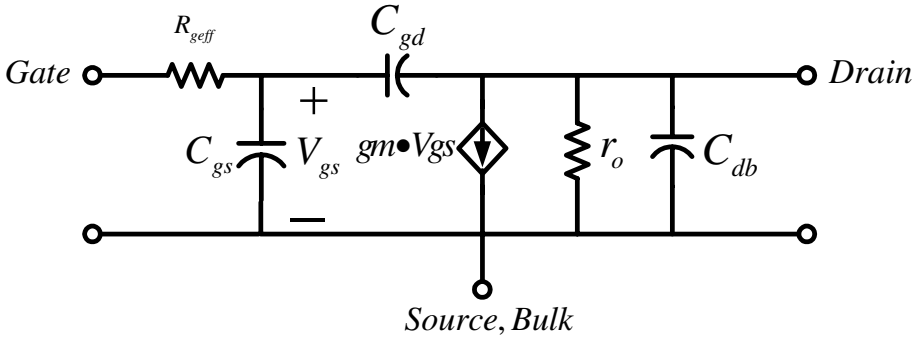

Figure 21 small signal equivalent circuit including the electrode gate resistance. This model is used to calculate the maximum oscillation frequency.

Notice that the intrinsic part of this model is similar to that given in Figure 5 with the bulk and source terminals being short-circuited. It is instructive to note that the transconductance current source does not respond instantaneously to the applied gate-to-source voltage but is delayed due to the RC filtering composed of the gate resistance and the gate capacitance. However the channel resistance and substrate losses were not included in this model. This will impact the accuracy of the result since these components will add to the total power dissipation. The frequency at which the unilateral power gain drops to unity is called the maximum oscillation frequency $\left(w_{\max}\right)$ since this is the maximum frequency at which the device can be made to oscillate in a feedback configuration. The operation beyond this frequency should be avoided since the power gain is less than unity. It can be shown that $\left(w_{\max}\right)$ for the circuit given in Figure 21 is approximately equal to:

$$w_{\max} \approx \sqrt{\frac{g_m^2 r_o}{4R_{geff}\left(C_{gs}+C_{gd}\right)\left[C_{gs}+\left(1+g_m r_o\right)C_{gd}\right]}} \tag{3-9}$$

where R_{geff} is the effective gate resistance and is given by equation (3-2). It is interesting to express $\left(w_{\max}\right)$ as a function of the cut-off frequency $\left(w_T\right)$. First, notice that the model in Figure 24 excludes the gate-to-bulk capacitance which is not a bad assumption since C_{gb} is in most cases much smaller than C_{gs} and C_{gd}. Dividing the numerator and denominator in equation (3-9) by $r_o\left(C_{gs}+C_{gd}\right)^2$, the maximum oscillation frequency can be expressed in terms of w_T, the output conductance g_{sd}, the effective gate resistance R_{geff}, and the gate-to-drain capacitance C_{gd} as follows:

$$w_{\max} \approx \frac{w_T}{\sqrt{4R_{geff}\left(g_{sd}+w_T C_{gd}\right)}} \tag{3-10}$$

where w_T is given by equation (3-7) and neglecting C_{gb}. Equations (3-7) and (3-9) show that the maximum oscillation frequency is inversely proportional to the square-root of the effective gate resistance (keeping the other parameters constant):

$$w_{\max} \propto \frac{1}{\sqrt{R_{geff}}}. \tag{3-11}$$

Since the maximum oscillation frequency depends on the Y-parameters that in turn are dependent on the model topology, $w_{\max}$ of a complete RF model will be different from that predicted by equation (3-9). However, the Y-parameters including substrate losses, channel resistance, source/drain diffusion resistance and a frequency dependant current source result in a very complicated expression for the unilateral power gain *(U)* making it impossible to arrive to a simple expression for $w_{\max}$. Nonetheless, equation (3-9) represents a good estimate and is widely used. Table 1 shows the maximum oscillation frequency, unity gain transition frequency, and minimum noise figure for a variety of technologies [35].

Table 1 High frequency figures-of-merit for different technologies

Channel length, L [nm]	**250**	**180**	**140**	**120**	**100**
$f_T\left(GHz\right)$	**33**	**49**	**70**	**84**	**112**
$f_{\max}\left(GHz\right)$	**41**	**47**	**51**	**52**	**60**
$NF_{\min}\left[dB\right]@2GHz$	**0.5**	**0.35**	**0.23**	**0.2**	**0.15**

Although the unity gain transition frequency is increasing almost quadratically with the downscaling of the channel length, there is only a slight improvement in $f_{\max}$. As the channel length of the transistor shrinks, short channel effects such as velocity saturation, mobility degradation, and drain induced barrier lowering (DIBL) become more severe impacting not only the DC characteristics of the device but also degrading the high frequency performance. Equation (3-10) shows the maximum oscillation frequency depends on four parameters: the unity-gain transition frequency w_T, the effective gate resistance R_{geff}, the drain-to-source conductance g_{sd}, and feedback gate-to-drain capacitance C_{gd}. It is instructive to examine the effect of scaling on these four parameters and indicate the set of parameters that are limiting the improvement of the maximum oscillation frequency. Although equation (3-6) indicates that the polysilicon gate resistance increases if the channel length decreases, using multiple fingers in a salicide (self aligned silicide) process reduces the total gate resistance significantly. In addition, the data in Table 1 as well as equation (3-7) clearly indicate that w_T is improving drastically with scaling. Consequently, R_{geff} and w_T are not the limiting parameters. As the channel length of the transistor shrinks, short channel effects such as velocity saturation, mobility degradation, impact ionization and drain induced barrier lowering (DIBL) become more severe. It was pointed out in chapter 2 that the last two effects can increase the output conductance by an order of magnitude. Moreover, the gate-to-drain feedback capacitance in short channel devices contains an inner fringing capacitance not present in long channels [10]. Where as the intrinsic gate-to-drain capacitance is almost zero for long channel devices, the inevitable DIBL effect in short channels increases the control of the drain on the channel charge and lead to a non-zero gate-to-drain intrinsic capacitance. These two effects tend to increase C_{gd} with scaling. In summary, short

channel effects increases both the output admittance and the gate-to-drain capacitance and are the two parameters that are limiting further improvement in the maximum oscillation frequency with channel scaling [36]. To account for SCEs in the output conductance, the drain-to-source conductance in equation (3-10) should be replaced by g_o defined in (2-17) and repeated below:

$$g_o \approx g_{sd} + g_{mb} R_{sub} g_{db} + g_{bd} . \tag{3-12}$$

However, based on the classical equivalent circuit in Figure 21, the author [31] ignores the second and third terms in equation (3-12). It is also instructive to note that the gate resistance will also impact the input impedance of the transistor and hence an accurate characterization of this resistance is crucial to achieve maximum power transfer.

Most of the proposed RF models use a lumped gate resistance as in Figure 21 whose value is extracted from the measured input impedance data [37,38]. Such an approach has the disadvantages that it does not show the dependence of the gate resistance on device geometry and operating point. Since the impedance of the oxide capacitance decreases with frequency, the effect of the channel resistance will become noticeable and will contribute to the real part of the input impedance at RF. Hence the *effective* gate resistance (probably a better name would be the input resistance) consists of two parts: The polysilicon gate resistance and a *fraction* of the channel resistance. In [39], a *physical* effective gate resistance (EGR) including both the first-order non-quasi-static effect (or simply taking the finite channel resistance into consideration) and the gate distributed resistance along its width was examined. The effective gate resistance was expressed as:

$$R_g = R_{gelt} + R_{gch} \tag{3-13}$$

where R_{gch} is the effective channel resistance seen from the gate and R_{gelt} is the gate electrode resistance expressed as follows:

$$R_{geltd} = R_{eltd}\left(\alpha\frac{W}{L}+\beta\right) \tag{3-14}$$

where R_{eltd} is the polysilicon sheet resistance and β is a factor modeling the external gate resistance which the extension of the polysilicon material over the active region . This factor can be neglected in most practical cases since its effect is negligible. The factor α models the distributed gate electrode resistance along its width. A typical value of α is $1/3$ as was previously discussed. The channel resistance is composed of two parts: a static component (R_{st}) which accounts for the dc channel resistance and an excess diffusion component (R_{ed}) which accounts for the change in the channel charge when a small signal is applied at the gate. The latter component is important in the weak inversion region. The static resistance is obtained from DC measurements by integrating an incremental channel resistance along the channel length under QS assumption. The excess diffusion resistance R_{ed} is derived from the diffusion current can be expressed as [39]:

$$R_{ed} = \frac{qL}{\eta W \mu C_{ox} KT} \tag{3-15}$$

where η is a technology dependent parameter. The effective channel resistance seen from the gate can be expressed as [39]:

$$\frac{1}{R_{gch}} = \gamma\left(\frac{1}{R_{st}}+\frac{1}{R_{ed}}\right) \tag{3-16}$$

where γ is a fitting parameter taking into consideration the distributed effect of the channel resistance at high frequency. Since the channel is connected at both ends, the lumped equivalent

representation of the distributed channel resistance is $1/12$ the total value if and only if the channel resistance is uniform. This condition is satisfied only for a zero drain-to-source DC voltage. In most RF applications the transistor is biased in the saturation region to get high gain and sufficient dynamic range. In the saturation region, the inversion layer decreases from the source to the drain resulting in a non-uniform distributed resistance. As a result, γ is left as a fitting parameter and was found to be 14 to best fit the experimental data [39]. The high frequency characterization of the gate resistance was studied in [40]. It is found experimentally that the gate resistance R_g decreases when either the channel length or the finger width increases, reaches a minimum value and then increases.

3.3 EFFECT OF SEMI-CONDUCTING SILICON SUBSTRATE ON THE RF PERFORMANCE OF MOSFETS

The fact that MOS transistors are fabricated on a silicon substrate has a detrimental effect on the performance of Mosfet's at radio frequencies. There are two major differences between Mosfet's and conventional high frequency transistors such as GaAs HEMTs , MESFETs etc. The first difference is related to the substrate resistivity and the second arises from the different physical structure of the two transistors. For ease of device fabrication, silicon substrates are doped to have a carrier concentration on the order of $10^{15} cm^{-3} \rightarrow 10^{18} cm^{-3}$ which results in a substrate resistivity on the order of $0.01 \rightarrow 10\ ohm \square cm$. On the other hand, the resistivity of a GaAs substrate is on the order of $10^{8} ohm \square cm$. The relatively low resistivity of silicon substrates results in larger parasitics that degrade the performance of integrated CMOS RF circuits and

complicates the modeling of MOS transistors. Signals applied the gate, drain, and source terminals of the device are coupled to the substrate through the gate-to-bulk, drain-to-bulk, and source-to-bulk capacitances. At low frequencies, these capacitors act as an open circuit decoupling the signals at the device terminals from the substrate. As the frequency increases, the impedance of these capacitors decreases and the signals at each terminal couple through the low resistive substrate. Signal coupling through the silicon substrate mainly affects the output admittance (y_{22}) as will be seen in chapter 4.

High frequency FET's have three terminals; a gate, a drain, and a source terminal and can be treated as two port networks. However, the MOS transistor has a fourth terminal called the "body" or the "bulk" node that affects the device characteristics and plays a similar role as the gate terminal but is less efficient. It usually referred to as the "back-gate". To prevent the source/drain-bulk junction of being forward biased, the bulk of the transistor is connected to the most positive supply for a PMOS transistor and to the most negative supply for an NMOS transistor. Therefore the extrinsic body node is connected to *ac* ground. In a common source configuration and at low frequencies, the bulk transconductance current source ($g_{mb}v_{b_i s}$) is deactivated because the ac bulk-to-source voltage ($v_{b_i s}$) is zero. At very low frequencies, the intrinsic and extrinsic body nodes have the same potential. However, as the operating frequency increases, the potential of the intrinsic body node increases because of the charging currents flowing through the distributed substrate resistance. As a result, the bulk transconductance current source is always activated at radio frequencies and has to be included in the small signal equivalent circuit even if the source is tied to the bulk node. Substrate signal coupling through the intrinsic body node mainly affects the small signal output characteristics of the transistor as will seen in chapter 4. The four terminal structure of the MOS transistor complicates modeling

and parameter extraction significantly to the extent that the intrinsic body node is usually neglected for simplicity although an accurate model ought to consider its effect.

3.4 SCATTERING PARAMETERS (S-PARAMETERS)

Measured data for both the transfer and impedance functions are required to characterize and to fully understand the behavior of an N-port network. The most commonly used parameters at low frequencies include the z-parameters, y-parameters, h-parameters, and the chain or ABDC parameters. These parameters require measuring a short-circuit current or an open-circuit voltage at the network terminals which is an easy task at low frequencies. However, these parameters are hard to measure above 1GHz and are not useful for several reasons. First, it is difficult to achieve an "ac-short" or an "ac-open" at radio frequencies over a broad range. Second, the voltage and currents depend on the position along the cable connecting the device. Consequently, if probe is not exactly positioned at the port terminals a measurement error will occur. Finally, applying a short or an open at the terminals of the transistor at RF may cause the transistor to oscillate and self-destruct. The S-parameter (scattering parameters) measurement technique overcomes the above limitations and is the most reliable and widely used method to characterize the high frequency characteristics of devices, circuits, and systems [41]. Scattering parameters are defined in terms of normalized voltages and currents and represent reflected and transmitted power through the network. Each parameter contains a real and an imaginary part or in other words has a magnitude and a phase. The Device Under Test (DUT) is inserted into a transmission line having a characteristic impedance Z_o which is usually 50Ω. Measuring the S-parameters requires

terminating the port with Z_o which is relatively easy at high frequencies. It is worth noting that the incident and reflected power is position independent along a lossless transmission line. They can be easily implemented in a CAD package and can be plotted using a Smith-chart. The S-parameters of a network is measured using a vector network analyzer (VNA) and can be easily converted to Y,Z, H-parameters to explore the data on a linear scale rather than on a Smith-chart. As mentioned earlier, the MOS transistor should be treated as a four terminal device at RF even if the source terminal is connected to the bulk. A four terminal device can be treated as a three-port network where nine S-parameters (eighteen real numbers) are needed to fully characterize the device at RF. On the other hand, a three terminal device can be analyzed as a two port network where only four complex parameters are needed to be measured. Unfortunately, an on-wafer three-port S-parameter technique is not yet well established and requires more measurement points and produces a large amount of data. Until recently, two-port S-parameter measurement has been used to characterize the MOS transistor at RF and is inadequate to extract the substrate related parameters such as the bulk transconductance (g_{mb}). Nonetheless, two-port modeling of the Mosfet including the body related parameters shows satisfactory results up to 10 GHz.

3.5 CONCLUSION

The effect of the distributed gate, channel, and substrate resistance on the high frequency performance of the MOS transistor has been discussed. To enhance the computational efficiency of the model, the distributed effects of the extrinsic elements at RF can be modeled using lumped elements or network. The distributed gate resistance at high frequencies can be

modeled using a lumped resistance in series with gate. Simple analytical expression for modeling this resistance were discussed and analyzed. The modeling and parameter extraction of the substrate and channel distributed resistance will be discussed in chapter 4. Not only does the finite resistivity of the silicon substrate complicate the modeling and parameter extraction of the MOS transistor, but also prevents the successful integration of a system-on-chip using the CMOS technology.

4.0 RADIO FREQUENCY MODELING AND PARAMETER EXTRACTION OF SUB-MICROMETER CHANNEL LENGTH MOSFETS ACCOUNTING FOR FIRST ORDER NQS EFFECTS

The continuous scaling of the channel length has lead to an MOS transistor that exhibits a 50 GHz transition frequency (f_t) for a 0.18- μm CMOS technology and a relatively low noise figure of 0.35 dB at 2-GHz [42]. Moreover, the low cost, low power, and very large scale integration (VLSI) capabilities offered by the CMOS technology makes it very suitable for integrating a system-on-chip. The above two factors have motivated extensive research on using the CMOS technology for RF IC applications [43], [44]. Until recently, high performance radio/microwave frequency integrated circuits have been implemented using either compound semiconductor transistors such as GaAs HEMTs, HBTs, and MESFETs, or silicon bipolar junction transistors. Microwave Small signal models and parameter extraction methods for compound semiconductors are well established [45], [46]. Signal coupling through the semi-conducting silicon substrate and the four terminal structure of the MOS transistor complicates the radio frequency modeling and parameter extraction of MOSFETs. Furthermore, it will be inaccurate to represent CMOS using high frequency small signal models and parameter extraction techniques that characterize the compound semiconductor transistors. The finite resistivity of the silicon substrate $(0.01 \sim 10)$ohm.cm results in a resistive parasitic component forming a distributed RC network with the depletion capacitance underneath the gate, source and drain regions as shown in Figure 22. At low frequency the impedance of the junction

capacitance is large and the substrate resistance has no effect on the device performance and is usually ignored.

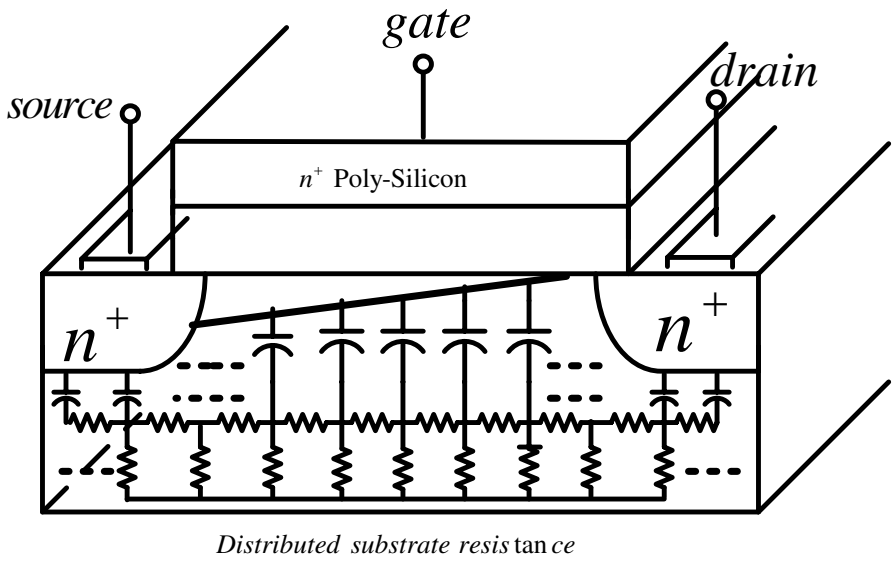

Figure 22 The bulk depletion capacitance and the substrate resistance form a distributed RC network at RF

As the frequency of the signal increases up to the Gega-Hertz range the impedance of the junction capacitance decreases and the signal at the gate, drain and source terminals will capacitively couple through gate-to-bulk and junction capacitances respectively. For example, a signal at the gate can be capacitively coupled to the drain terminal through the gate-to-bulk capacitance, substrate resistance, and drain-to-bulk junction capacitance. In the strong inversion saturation region the substrate losses have a great impact on the output admittance of the transistor [47], [48]. Consequently, a high frequency model should include the substrate losses to better match the output impedance to the experimental data. Due to its finite resistivity, the

intrinsic body node of the MOS transistor is at a different *ac* potential than the extrinsic node. Most often, the body terminal (extrinsic) is connected to a fixed potential with a polarity such as to make sure that the junction diodes formed by the source and drain diffusion regions are reverse biased. Hence the extrinsic body node is at ac ground in almost all practical cases. Because of the various signal coupling to the substrate as the frequency of the signal is increased, the intrinsic body potential will increase from its zero value at low frequency to a value dependent of the impedance of the junction capacitance and substrate resistance. Consequently, the non-zero intrinsic body potential will activate the substrate voltage-controlled current source $g_{mb}v_{bi}$. This current source has a great impact on the output admittance Y_{22}. As a matter of fact, it has been demonstrated [49] that it is impossible to accurately model the output impedance for different gate biases without including the bulk transconductance current source. However, many models exclude the intrinsic body node and hence ignore the effect of bulk transconductance on the output admittance [48], [50]. Although the output admittance predicted by these models match the experimental data at the *given bias point*, nonetheless it would deviate at other bias points. In addition, the substrate resistance $\left(R_{sub}\right)$ extracted from these models cannot be modeled physically and hence is not scalable or predictable from the technology parameters. To achieve both accuracy and computational efficiency, the distributed substrate resistance has been modeled in the literature using different resistive networks. These include the five-resistor network, four-resistor network [37], three-resistor network [51], two-resistor network [38], and one-resistor network [52]. The first two networks are the most accurate and extend the validity of the model in predicting the Y-parameters at higher frequencies. However, they lead to complex circuits and make it harder to extract the model parameters. The one and two resistor networks are simpler to analyze and result in simple parameter extraction

algorithms. They are however less accurate at high operating frequencies. The three-resistor network is a compromise among the substrate equivalent circuits (SEC). It is simpler to analyze, results in a straightforward extraction algorithm, and is accurate for up to 10GHz. A detailed characterization of the substrate resistance at RF has been examined at different bias conditions [51]. It is demonstrated that the substrate resistance shows a very week bias dependence. Consequently, it is sufficient to extract the substrate resistance at one bias point. Exploiting this feature, a very simple and accurate method of extracting a single resistor substrate network of an RF MOSFET in the cut-off region (VGS=0) was formulated [52].

4.1 PROPOSED MODEL ANDPARAMETER EXTRACTION PROCEDURE

In this section a small signal model is proposed together with a straightforward algorithm that extracts the parameters of the model from measured y-parameters. The device under test is a multi-fingered N-MOSFET with $100\mu m$ channel width having 20-unit gate fingers fabricated using the $0.35\mu m$ CMOS technology [48]. The device is biased at a gate-to-source voltage of 1V (to insure strong inversion) and drain-to-source voltage of 2V (to insure that the transistor is in the saturation region). The S-parameters of the DUT are measured in the common source substrate configuration using on-wafer RF probes and an HP 8510C vector network analyzer. The measured S-parameters are then converted to Y-parameters. To extract the small signal parameters of the proposed model, the parasitic components of the test structure should be first removed or deembedded from the measure raw Y-parameters. Figure 23 shows a typical representation of the DUT including the parasitics of the test structure. $Y_{parasitic1,}Y_{parasitic2,}$ and $Y_{parasitic3}$ model the influence of the parallel

parasitics. They can be extracted from the measured data of the open pad structure (S_{open}). Representing the series parasitics, $Z_{series1,}Z_{series2,}$ and $Z_{series3}$ can be extracted from the measured data of both open and short pad structure (S_{short}). The measured S-parameters of the open structure, short structure, and DUT are then converted to Y-parameters represented as Y_{open}, Y_{short}, and Y_{DUT} respectively. A two-step de-embedding is then performed. The purpose of the first step is to remove the influence of the parallel parasitics by subtracting Y_{open} from Y_{DUT} and Y_{short} resulting in a new set of de-embedded Y-parameters:

$$Y_{DUT-Yp} = Y_{DUT} - Y_{open} \tag{4-1}$$

$$Y_{short-Yp} = Y_{short} - Y_{open} \tag{4-2}$$

The Z-parameters of the transistor are obtained by a second de-embedding step according to following equation:

$$Z_{transistor} = Z_{DUT-Yp} - Z_{short-Yp} \tag{4-3}$$

where:
$$Z_{DUT-Yp} = \left(Y_{DUT-Yp}\right)^{-1}, \tag{4-4}$$

and
$$Z_{short-Yp} = \left(Y_{short-Yp}\right)^{-1} \tag{4-5}$$

In summary, the measured Y-parameters corresponding to the transistor can be obtained as follows:

$$Y_{transistor} = \left[\left(Y_{DUT} - Y_{open}\right)^{-1} - \left(Y_{short} - Y_{open}\right)^{-1}\right]^{-1} \tag{4-6}$$

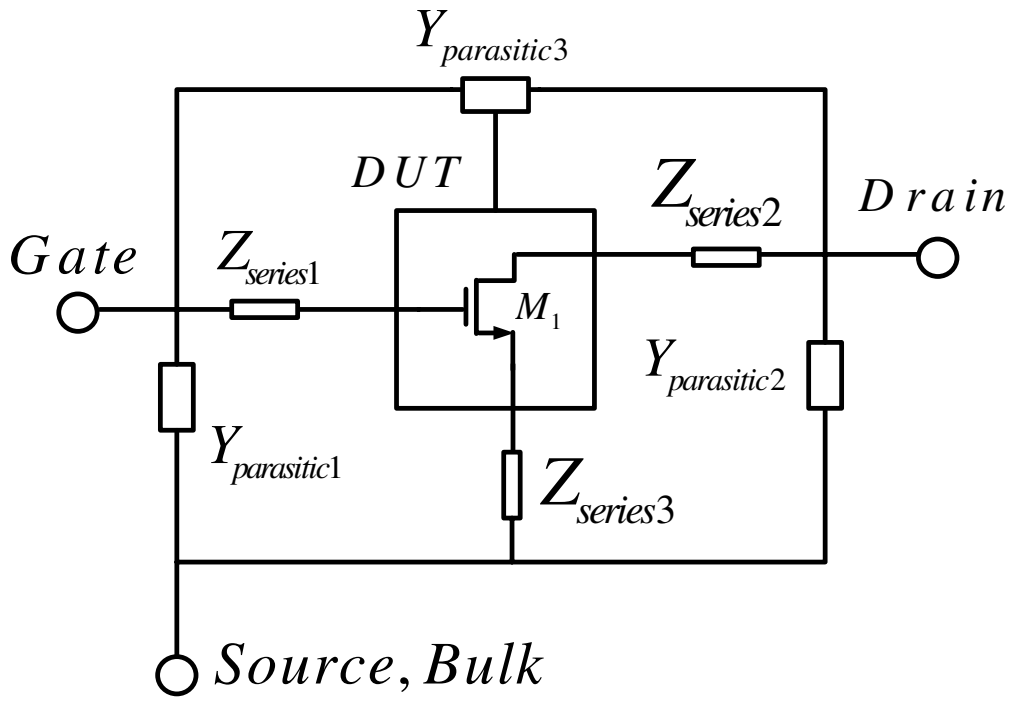

Figure 23 Equivalent circuit representation of the test structure used to de-embed the measured high frequency data of the MOS transistor.

The proposed small signal model is shown in Figure 24. The source and the bulk terminal are tied together resulting in a two port common-source configuration. This setup is applicable to most high frequency applications. The gate resistance, R_{gpoly}, models the distributed gate resistance of the polysilicon material along the transistor width. Its value can be calculated using a standard formula. The polysilicon sheet resistance (R_{shpoly}) of a 0.35um CMOS technology is on the order of $10\,\Omega/\square$. If N_f is the number of fingers used, then the distributed gate electrode resistance can be approximated by a single lumped resistor of value:

$$R_{gpoly} = R_{shply}\frac{1}{3}\frac{1}{N_f^2}\frac{W}{L} \tag{4-7}$$

where the 1/3 factor accounts for the distributed nature of this resistor.

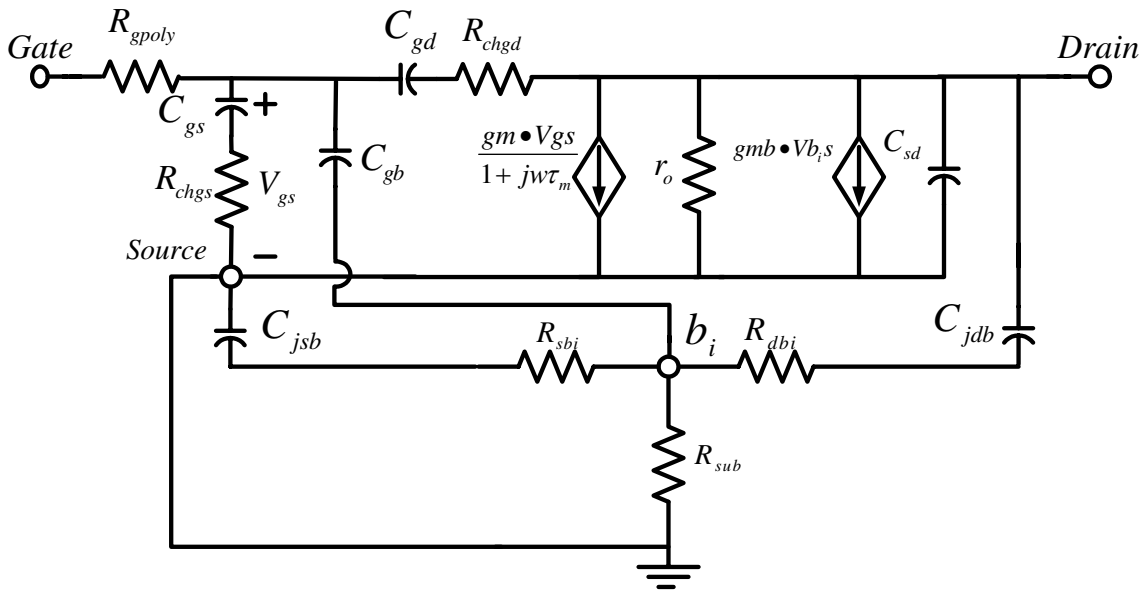

Figure 24 An RF small-signal MOSFET model based on the first order non-quasi-static effect. The effective gate-to-source and gate-to-drain channel resistance is modeled by R_{chgs} and R_{chgd} respectively. Substrate signal coupling through the intrinsic body node and junction capacitances is examined by introducing a new 3-resistor T-network.

Since the width W of the transistor is divided into N_f sections, then the resistance of each section is reduced by N_f. Furthermore, since the sections are connected in parallel the equivalent resistance will be reduced by N_f^2. Hence, the gate electrode resistance decreases quadratically with the number of fingers. This appealing result is the reason why multi-fingered transistors are normally employed in the design of high-speed integrated circuits. The

characterized devices have a channel width of 100um divided into 20 sections (N_f=20) and a channel length of 0.35um. Using equation (4-7):

$$R_{gpoly} = 10\frac{1}{3}\frac{1}{20^2}\frac{100}{0.35} \approx 2.4\Omega . \tag{4-8}$$

This clearly indicates that for a large multi-fingered transistor the gate electrode resistance can be neglected from the extraction procedure with little penalty. On the other hand, the channel resistance is independent on the number of fingers and has a great impact on the real part of the input admittance $\mathrm{Re}[Y_{11}]$. If a single finger transistor is used (N_f=1), the gate electrode resistance, using equation (4-8) is 960Ω degrading the frequency performance, noise characteristics and transient response of the transistor. Moreover, it would be impossible to match the real part of the input impedance to 50Ω which is a prerequisite to achieve optimal performance of most RF IC building blocks. Although, the gate electrode resistance can be lowered considerably by using multiple fingers, this remedy has two limitations. First, it increases the source or drain junction capacitance and second, it degrades the circuit density.

In the equivalent circuit displayed in Figure 24, C_{gs} models the intrinsic and extrinsic gate-to-source capacitance and R_{chgs} models the effective channel resistance between the gate and the source. Similarly, C_{gd} and R_{chgd} model the gate-to-drain capacitance and the effective channel resistance between the gate and the drain respectively. Ignoring short channel effects such as channel length modulation, DIBL, and impact ionization, the value of these small signal elements were found to be zero for the intrinsic gate-to-drain capacitance C_{gdi} and infinity for R_{chgd} [10]. These results should not be surprising. In fact, ignoring channel length modulation simply means that the drain terminal has no effect on the channel charge. Hence the capacitive

effect of the drain on the gate terminal symbolically denoted as C_{gdi} is zero. In short channel devices, the situation is different. The drain affects the channel charge and consequently the gate charge through channel length modulation and the two-dimensional charge sharing effect also known as the DIBL effect. Hence, the capacitive effect of the drain on the gate is not zero ($C_{gdi} \neq 0$). The two resistors R_{chgs} and R_{chgd} simulate the first order non-quasi-static effect and their inclusion in the model is necessary to accurately predict the input admittance Y_{11}. Since the value of R_{gpoly} is very small, it will be ignored from the parameter extraction process for simplicity. The gate-to-drain effective channel resistance R_{chgd} is crucial to match the simulated $\mathrm{Re}[Y_{12}]$ and $\mathrm{Re}[Y_{22}]$ to the experimental data. Evidently, the channel resistance is inversely proportional to the inversion charge density. Since the magnitude of the inversion charge density decreases from the source to the drain, the effective gate-to-drain resistance R_{chgd} is expected to be larger than R_{chgs}. Indeed, the extracted values of these two resistors confirm this fact.

Based on the results obtained from a first order NQS analysis, the gate transconductance is a function of frequency and is given by:

$$g_m(w) = \frac{g_m}{1 + jw/w_p} \tag{4-9}$$

where

$$w_p = \frac{1}{\tau_m} = \frac{g_m}{C_m} \tag{4-10}$$

$$C_m = C_{dg} - C_{gd} \tag{4-11}$$

The transcapacitance C_m has two important roles. First, it insures that the charge conservation condition is satisfied. Second, it is needed to match $\mathrm{Im}[Y_{12}]$ and $\mathrm{Im}[Y_{21}]$ at the same time. Including C_m in the model as a frequency dependent current source, as was done in the complete quasi-static model will lead to a zero in the right hand side of the s-plane for the drain current which result in a discrepancy between the simulated and measured $\mathrm{Re}[Y_{21}]$ at high frequencies (above ~ 6 GHz) [48]. The proposed model reflects the presence of C_m by including it through a frequency dependent gate transconductance as seen from equations (4-9) and (4-10). The gate transconductance $g_m(w)$ decreases with frequency and contains a left-hand-plane pole that is a function of C_m. This implementation is more realistic and is a first order approximation to the NQS effects that occur at high frequencies. Moreover, this implementation is mandatory to match the simulated $\mathrm{Re}[Y_{21}]$ to the experimental data for frequencies above 6GHz and is a major improvement to the model presented in [48] as will be discussed in section 4.3. The output resistance of the transistor presented as $r_o = 1/g_{sd}$ can be extracted at low freq.1uency and is probably one of the most challenging small signal parameters to model physically. It has a great impact on the real part of the output admittance $\mathrm{Re}[Y_{22}]$. The source to drain capacitance C_{sd} models the capacitive effect of the drain on the source. For long channel devices, C_{sd}, was theoretically calculated to be 0 using the same argument as for C_{gdi} [10]. This capacitance is important to model the imaginary part of the output admittance $\mathrm{Im}[Y_{22}]$ because it is connected directly from the output terminal (drain) to ground as obvious from Figure 24. The distributed nature of the substrate resistance at high frequencies is modeled by a lumped three-resistor T-network shown in Figure 24 as R_{dbi}, R_{sbi}, and R_{sub}. The latter resistor is extracted from the

output admittance Y_{22} while the drain-to-bulk and source-to-bulk resistors can be calculated using the following standard formula:

$$R_{sbi}, R_{dbi} = \frac{1}{2N_f}\frac{L_f}{W_f}R_{shsi} \tag{4-12}$$

where R_{shsi} is the sheet resistance of the silicon substrate underneath the channel between the source and drain terminals of a single finger device. Its value is typically $(0.1 \to 1)\,k\Omega/\square$. For $R_{shsi} = 1k\,\Omega/\square$, equation (4-12) yields:

$$R_{sbi} = R_{dbi} = \frac{1}{40}\frac{0.35}{5}1k\Omega = 1.75\Omega\,. \tag{4-13}$$

For devices with a large number of fingers, the value of the two resistors R_{sbi} and R_{dbi} becomes negligible with respect of R_{sub}. In this case, these two resistors can be omitted and the substrate distributed resistance can be modeled as a single lumped resistor R_{sub} [53]. High frequency measurements on multi-fingered devices with $N_f = 50$ have demonstrated that a single substrate resistor is sufficient for accurate RF MOSFET modeling up to 10 GHz [55]. However, the simulated Y-parameters of a model with a single resistor substrate network start to deviate significantly from the experimental data beyond 10GHz even for a large number of fingers. The proposed substrate network (T-network) is then valid for an arbitrary number of fingers and models the Y-parameters accurately for operating frequencies above 10GHz. Albeit the value of R_{sbi} and R_{dbi} is small (1.75Ω) and can be neglected during parameter extraction, they will be included in the analytical expressions developed for the real and imaginary parts of the output admittance. Consequently, the extraction process is generic and can be applied for a device with an arbitrary number of fingers. The drain-to-bulk and source-to-bulk junction capacitances are

represented by C_{jdb} and C_{jsb} respectively. These capacitors couple the signal from the source and drain terminals to the intrinsic body node and vice versa. Each capacitor is a lumped approximation of the distributed channel-bulk intrinsic capacitance and the distributed diffusion region-bulk extrinsic capacitance. The lumped extrinsic capacitor consists of two components: a bottom wall component and two sidewall components. The capacitance C_{gb} represents the sum of the intrinsic and extrinsic gate-to-bulk capacitance. The intrinsic component is bias dependent and has a relatively high value $(\approx C_{ox})$ when the device is off. In the strong inversion saturation region, the value of this capacitance diminishes because of the shielding effect of the channel charge. The extrinsic component extends along the channel length outside the active channel area. It is bias independent and is usually very small. The magnitude of the charging current flowing through C_{gb} is negligible compared to that flowing through the other elements. It has a paltry influence on the Y-parameters [56] and is therefore neglected during parameter extraction for simplicity.

4.2 Y-PARAMETER ANALYSIS AND PARAMETER EXTRACTION

The small signal parameters of the proposed model are extracted from the de-embedded y-parameters. The first step is to derive simple analytical expressions for the y-parameters of the proposed model. The next step is to develop an extraction algorithm to obtain the small signal parameters. The proposed method of extracting the small signal parameters from $Y_{11}, Y_{21},$ and Y_{12} is similar to that in [48] with minor modifications specifically in extracting the two components that

model the effective channel resistance. This approach is straightforward and does not require an optimization process, which may lead to unrealistic values for the model parameters. Including the bulk transconductance (g_{mb}) and the proposed three-resistor substrate network, result in a very complicated expression for the output admittance. After some simplifications that relate to the range of the operating frequency, the output admittance is shown to be a function of five unknown parameters: $g_{sd}, C_{sd}, g_{mb}, C_{jdb}, R_{dbi}$ and R_{sub}. This imposes a big challenge. To avoid tackling a problem with this magnitude, a simple curve fitting technique is proposed to extract the three parameters $C_{sd}, R_{sub},$ and C_{jdb} from the frequency response of the real and imaginary parts of the output admittance Y_{22}. The resistance R_{dbi} is calculated using a standard formula given by equation (4-12). In addition, the drain-to-bulk and source-to-bulk resistances are assumed equal from symmetry consideration. The bulk transconductance (g_{mb}) is difficult to extract from high frequency measurements and is not included in most published models although is important to account for the strong gate bias dependence of the output admittance [49]. We have concluded that for a given device with known dimensions and bias conditions, the Bsim3v3.1 model with parameters extracted for the $0.35\mu m$ CMOS process has a gate transconductance g_m that is very close to that obtained from the measured $\mathrm{Re}[Y_{21}]$ at low frequency [48]. Since the value of g_m that results from a DC analysis in Spice is accurate, and since the Bsim3v3 model predicts reasonably accurate low frequency small signal parameters, one conclude that the bulk transconductance g_{mb} can also be obtained by running the same DC analysis in Spice. The Bsim3v3 model has proved to be both accurate and scalable in determining the device dc characteristics. However, at high frequency it fails to predict the device performance unless two modifications are included. The first modification consists of adding an external gate resistance and the other is to account for the substrate losses [37].

The equivalent circuit of Figure 24 can be analyzed as a two-port network with the input being the gate terminal and the output the drain terminal. The source and the extrinsic substrate terminals are tied to ground. The polysilicon gate resistance is excluded from the parameter extraction for simplicity. A detailed analysis of the Y-parameters is presented in Appendix A. The simplified expressions that are used for extracting the model elements are summarized bellow:

$$Y_{11} \approx w^2 \left[R_{chgs} C_{gs}^2 + R_{chgd} C_{gd}^2 \right] + jw\left(C_{gd} + C_{gs} \right) \tag{4-14}$$

$$Y_{12} \approx -w^2 R_{chgd} C_{gd}^2 - jwC_{gd} \tag{4-15}$$

$$Y_{21} \approx g_m - w^2 R_{chgd} C_{gd}^2 - jwC_{dg} \tag{4-16}$$

$$\begin{aligned} Y_{22} \approx g_{sd} &+ w^2 \left[R_{chgd} C_{gd}^2 + \left(R_{sub} + R_{dbi} \right)\left(1 + g_{mb} R_{sub} \right) C_{jdb}^2 \right] \\ &+ jw\left[C_{gd} + C_{sd} + C_{jdb}\left(1 + g_{mb} R_{sub} \right) \right] \\ &- jw^3 \left[C_{jdb}^3 \left(R_{sub} + R_{dbi} \right)^2 \left(1 + g_{mb} R_{sub} \right) \right] \end{aligned} \tag{4-17}$$

Based on equations (4-14) to (4-17), the impact of each circuit element on the real and imaginary part of the Y-parameters can be explained as follows. The effective gate-to-source and gate-to-drain channel resistances (R_{chgs} and R_{chgd}) are needed to accurately model the real part of the input admittance $\mathrm{Re}[Y_{11}]$. Moreover, without R_{chgd} it is impossible to match $\mathrm{Re}[Y_{12}]$ and $\mathrm{Re}[Y_{22}]$ to the experimental data. The amplification occurring in a transistor implies that the MOSFET is a nonreciprocal network. As a result, the forward and reverse transmissions through

the device represented as Y_{21} and Y_{12} are not equal. Equation (4-15) and (4-16) implies that it is impossible to model $\mathrm{Im}[Y_{12}]$ and $\mathrm{Im}[Y_{21}]$ at the same time without considering the nonreciprocal charging effect of the gate and drain terminals on one another ($C_{gd} \neq C_{dg}$). The gate-to-source capacitance C_{gs} affects $\mathrm{Im}[Y_{11}]$. In essence, the gate-to-source equivalent circuit directly impacts the input admittance. The rest of the model parameters are needed to match the output admittance Y_{22} to the experimental data.

4.2.1 EXTRACTION OF SMALL SIGNAL PARAMETERS

Simple analytical equations are derived from the real and imaginary parts of the y-parameters given in equations (4-14) to (4-17). The gate transconductance is the ratio of the short circuit output current to that of the gate-to-source voltage at DC. This measurement is equivalent to that used in equating Y_{21}. Hence g_m should be extracted from Y_{21} at low frequencies. As a matter of fact, equation (3-16) shows that g_m is equal to the y-intercept of $\mathrm{Re}[Y_{21}]$ versus w^2. $C_{gd}, C_{gs}, C_{dg}, R_{chgs}, R_{chgd}$, and g_{sd} can be calculated from the above equations as follows:

$$g_m = \mathrm{Re}[Y_{21}]\Big|_{w^2=0} \tag{4-18}$$

$$g_{sd} = \mathrm{Re}[Y_{22}]\Big|_{w^2=0} \tag{4-19}$$

$$C_{gd} = \frac{-\mathrm{Im}[Y_{12}]}{w} \tag{4-20}$$

$$C_{gs} = \frac{\left(\mathrm{Im}[Y_{11}] + \mathrm{Im}[Y_{12}]\right)}{w} \tag{4-21}$$

$$C_{dg} = \frac{-\mathrm{Im}[Y_{21}]}{w} \tag{4-22}$$

$$R_{chgs} = slope\left\{\frac{\left(\mathrm{Re}[Y_{12}]+\mathrm{Re}[Y_{11}]\right)}{C_{gs}^2}\right\} vs.\ w^2 \tag{4-23}$$

$$R_{chgd} = slope\left\{\frac{-\mathrm{Re}[Y_{12}]}{C_{gd}^2}\right\} vs.\ w^2 \tag{4-24}$$

Equations (4-18)-(4-24) are used to extract the corresponding model parameters. A gate transconductance g_m of 16.6 mS is obtained from the Y-intercept of $\mathrm{Re}[Y_{21}]$ versus w^2 as depicted in Figure 25 .A drain-to-source conductance g_{sd} of 0.31mS was evaluated from the real part of the output admittance at low frequencies as shown in Figure 26.

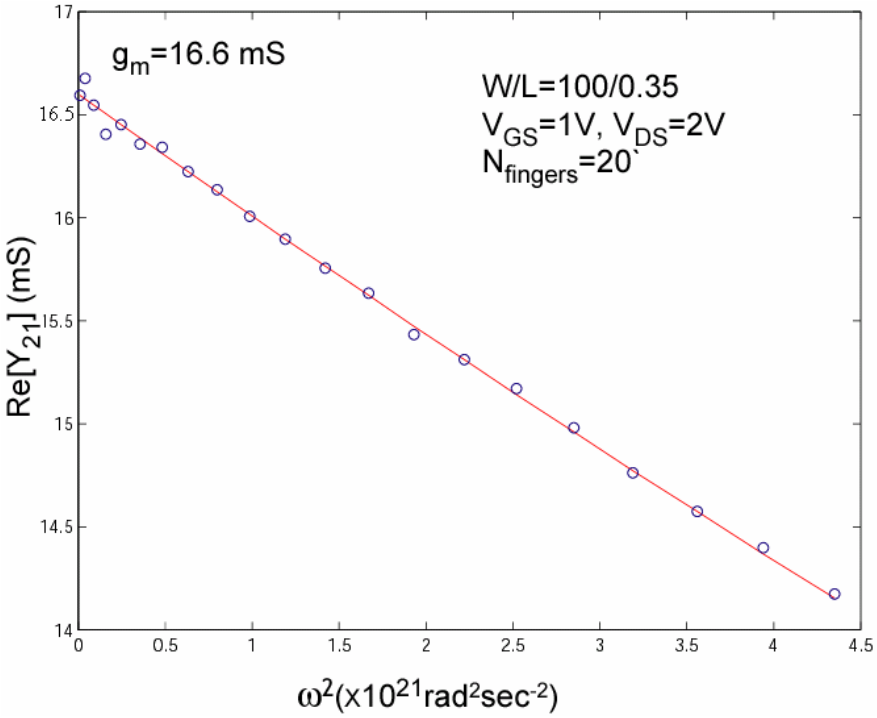

Figure 25 Extraction of the gate transconductance was obtained from the Y-intercept of $\mathrm{Re}[Y_{21}]$ versus w^2

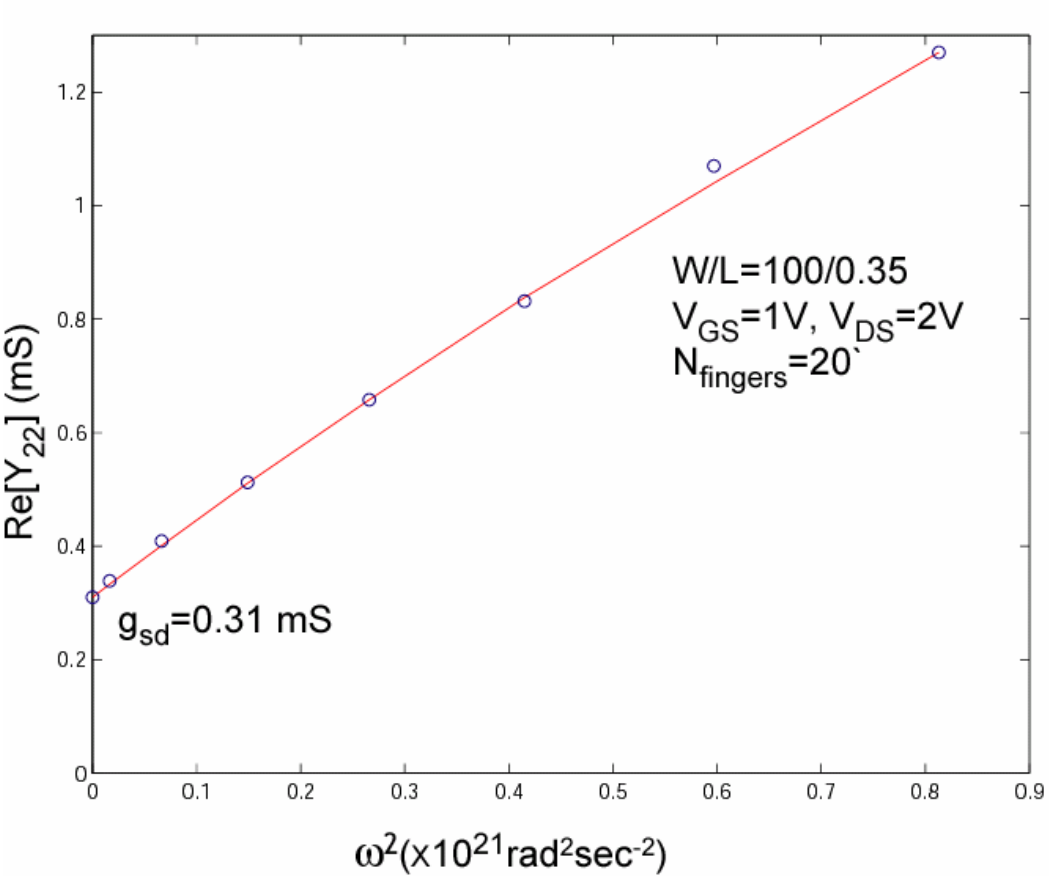

Figure 26 The drain-to-source conductance was obtained from the Y-intercept of $\mathrm{Re}[Y_{22}]$ **versus** w^2

The gate-to-drain, gate-to-source, and drain-to-gate capacitances are extracted using equations (4-20), (4-21), and (4-22) respectively. The extracted value of these capacitors as a function of frequency is displayed in Figure 27. The results show that the extracted parameters are frequency independent verifying the physical validity of the model and the accuracy of the parameter extraction method. The average values of the extracted C_{gs}, C_{gd}, and C_{dg} are $122.4\,fF$, $34\,fF$, and $151\,fF$ respectively. Their relative values can be explained in physical terms as follows. Since the gate terminal has more control on the channel charge than the other terminals

of the device, the capacitive effect of the gate on the drain is expected to be larger than the capacitive effect of the drain/source terminals on the gate ($C_{dg} > C_{gd}, C_{gs}$). Furthermore, in the strong inversion saturation region the drain terminal has a weaker control on the channel charge than the source terminal. Consequently, the capacitive effect of the source on the gate C_{gs} is expected to be greater than that of the drain on the gate ($C_{gs} > C_{gd}$).

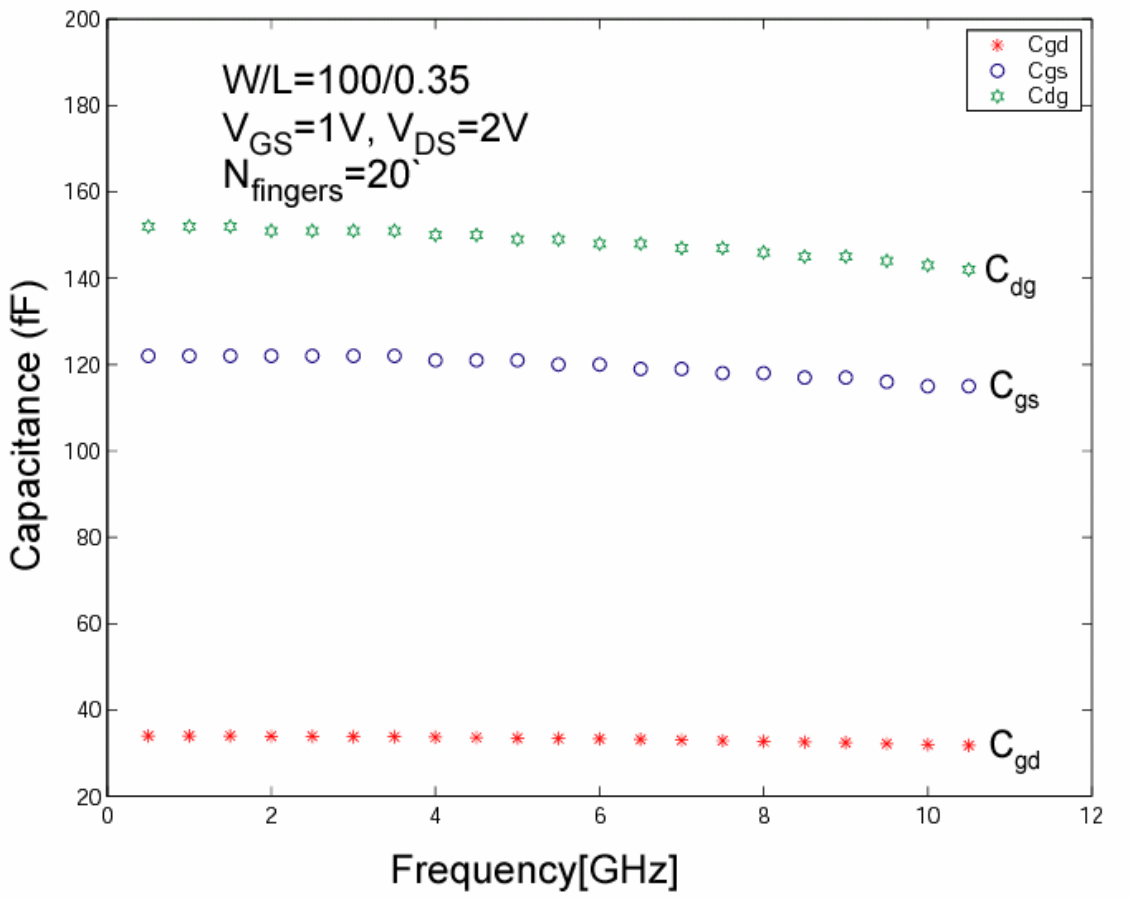

Figure 27 The extracted gate-to-source, gate-to-drain, and drain-to-gate capacitances are almost constant with frequency, asserting the physical validity of the model and accuracy of the parameter extraction method.

The effective gate-to-source and gate-to-drain channel resistances (R_{chgs} and R_{chgd}) are extracted from the slope of $\left(\mathrm{Re}\left[Y_{12}\right]+\mathrm{Re}\left[Y_{11}\right]\right)C_{gs}^{-2}$ and $-\mathrm{Re}\left[Y_{12}\right]C_{gd}^{-2}$ as a function of w^2 as shown in Figure 28 and Figure 29 respectively.

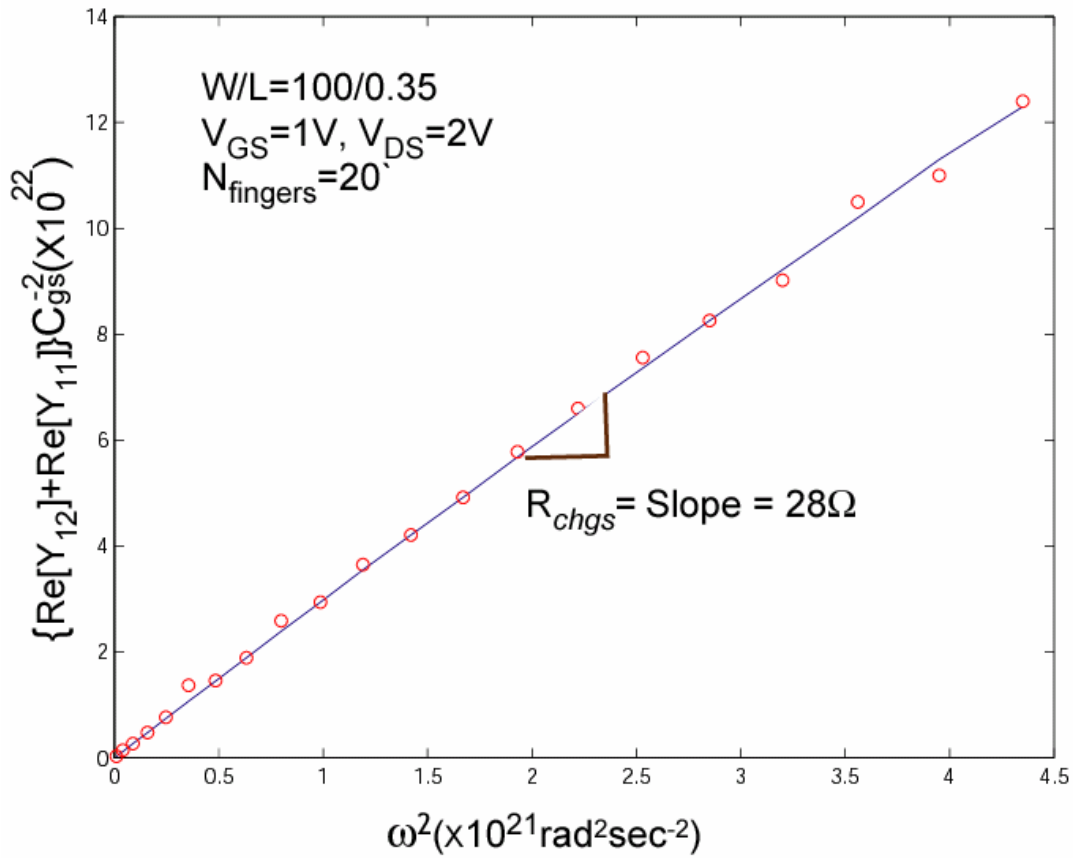

Figure 28 The effective gate-to-source channel resistance R_{chgs} is determined from the slope of $\left(\mathrm{Re}\left[Y_{12}\right]+\mathrm{Re}\left[Y_{11}\right]\right)C_{gs}^{-2}$ as a function of w^2

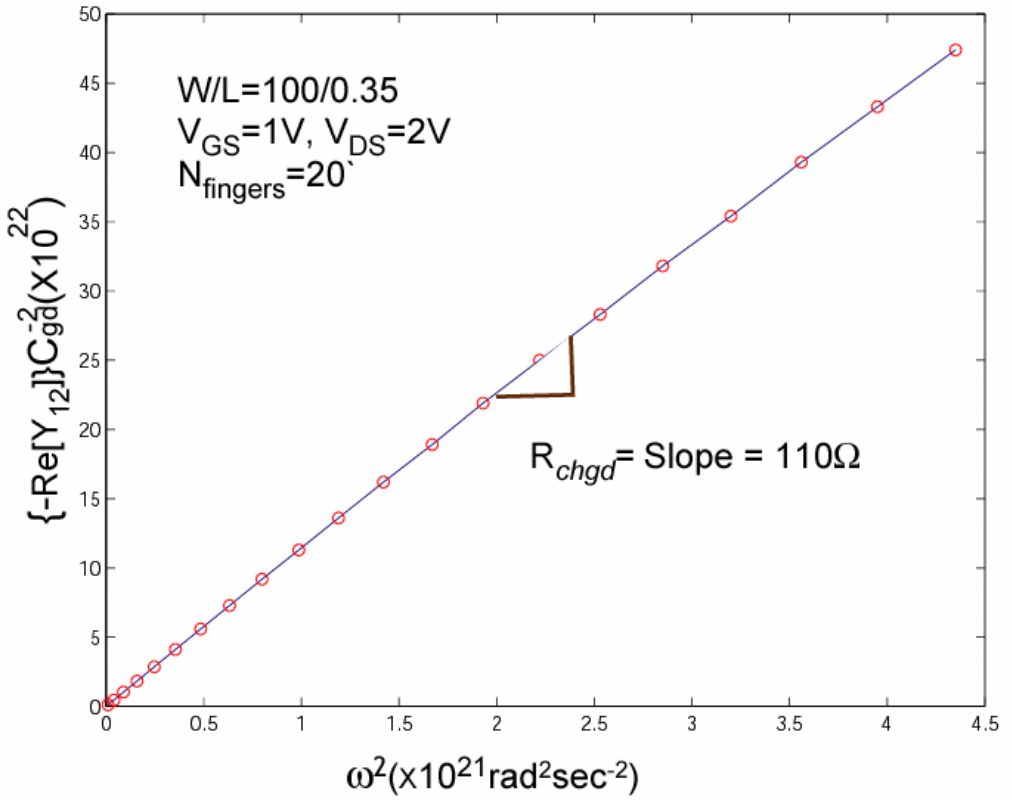

Figure 29 The effective gate-to-drain channel resistance R_{chgd} is determined from the slope of $-\mathrm{Re}[Y_{12}]C_{gd}^{-2}$ as a function of w^2

The frequency dependence of the extracted parameters R_{chgs}, and R_{chgd} is shown in Figure 30. The results show that these two resistors are almost constant with frequency. The source-to-drain capacitance C_{sd}, substrate resistance R_{sub}, and drain-to-bulk junction capacitance C_{jdb} are extracted by curve fitting the experimental data for both the real and imaginary parts of Y_{22} to the following two polynomials:

$$\mathrm{Re}[Y_{22}] = \overbrace{0.31^{mS}}^{g_{sd}} + Aw^2 \tag{4-25}$$

$$\mathrm{Im}[Y_{22}] = Bw + Cw^3 . \tag{4-26}$$

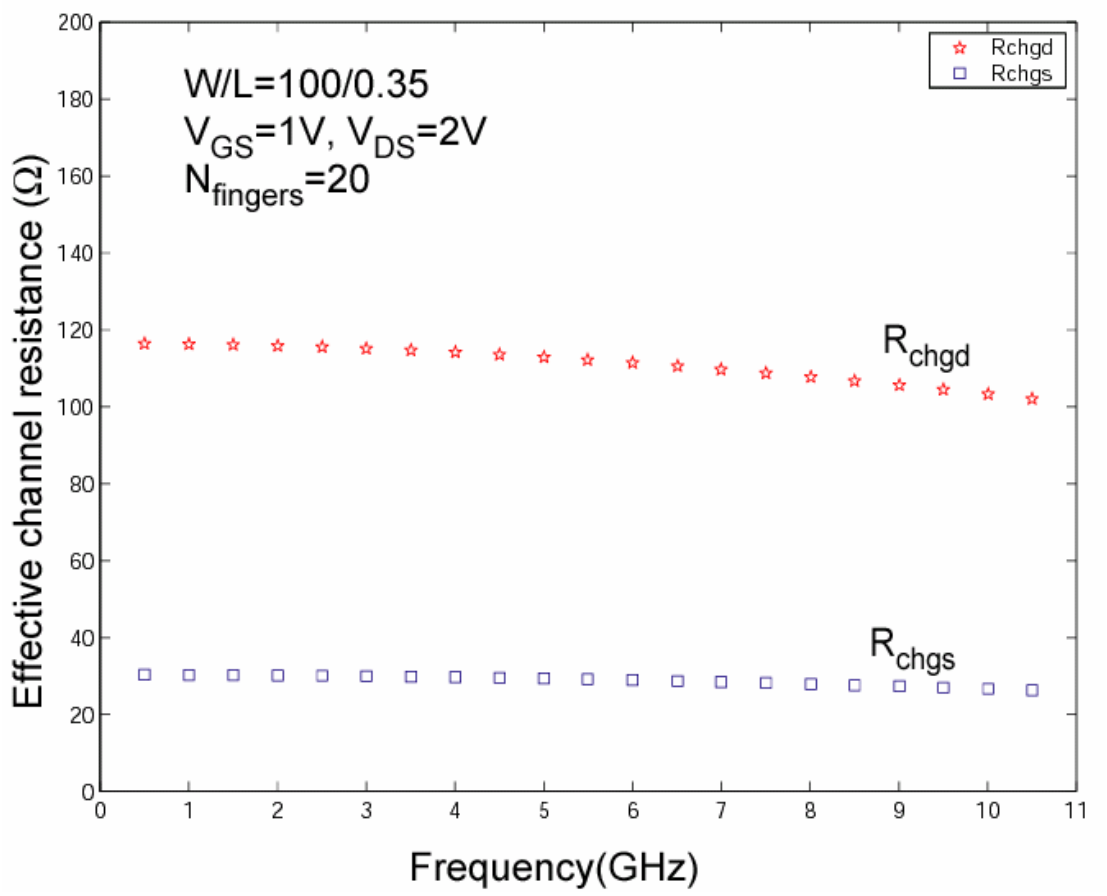

Figure 30 Frequency dependence of R_{chgs} and R_{chgd}. The results show that these two resistors are almost constant with frequency, verifying the accuracy of the model.

Using the DataFit curve fitting software, the value of the three constants A, B, and C are found to be 1.23×10^{-24}, 118.8×10^{-15}, and -9.27×10^{-36} respectively. A drain-to-bulk resistor (R_{dbi}) of 1.75Ω is calculated using equation (3-13) and a bulk transconductance g_{mb} of $4.2mS$ is evaluated by running a DC analysis using Pspice. Using equation (4-17) and the extracted values of R_{chgd} and C_{gd}, three equations in three unknowns are derived as follows:

$$\left(R_{sub}+R_{dbi}\right)\left(1+g_{mb}R_{sub}\right)C_{jdb}^{2}=1\times10^{-24} \tag{4-27}$$

$$C_{gd} + C_{sd} + C_{jdb}\left(1 + g_{mb}R_{sub}\right) = 118.8 \times 10^{-15} \tag{4-28}$$

$$C_{jdb}^{3}\left(R_{sub} + R_{dbi}\right)^{2}\left(1 + g_{mb}R_{sub}\right) = 9.27 \times 10^{-36} \tag{4-29}$$

Using (4-29) and (4-27), the following relation between the R_{sub} and C_{jdb} is derived:

$$R_{sub}C_{jdb} = 9.27 \times 10^{-12} \tag{4-30}$$

A substrate resistance R_{sub} of 134Ω and drain-to-bulk capacitance C_{jdb} of $68.8\,fF$ are obtained from equations (4-27) and (4-30). Finally, the source-to-drain capacitance using equation (4-28) is $-22.9\,fF$. Based on the insight gained from the simulated output admittance of the proposed equivalent circuit the value of C_{sd} and C_{jsb} were chosen to be $-16\,fF$ and $48\,fF$ respectively for best matching the experimental data. The average value of the extracted model parameters are listed in Table 2.

Table 2 Average and optimized values of the extracted model parameters

g_m	$16.6mS$	C_{gs}	$122.4\,fF$
g_{sd}	$0.31mS$	C_{dg}	$151fF$
R_{chgs}	28Ω	R_{subd}	134Ω
R_{chgd}	110Ω	C_{jsb}	$48\,fF$
C_{gd}	$34\,fF$	C_{jdb}	$68.8\,fF$
C_{sd}	$-16\,fF$	g_{mb}	$4.2mS$

4.3 MODEL VERIFICATION

The proposed equivalent circuit with the extracted values is implemented in Pspice. The frequency-dependent voltage-dependent current source ($g_m(w)V_{gs}$) was implemented using a Laplace voltage dependent current source (*GLAPLACE*) found in the *abm.slb* library. The Y-parameters of the model are plotted in Pspice by connecting the desired port to an ac signal voltage (having an amplitude of 1V for simplicity) and by ac short circuiting the other terminal. Once plotted in the waveform analyzer known as Probe, different analog operations and functions can be performed on the y-parameters as desired. This is a very powerful tool to investigate the effect of different circuit elements on the Y-parameters and to plot currents and voltages as a function of frequency. The simulated Y-parameters of the proposed model are compared with the measured data and with the Bsim3v3.1 model as shown in Figures (31)-(35). The results show that the proposed model can accurately predict the device small-signal characteristics up to 10 GHz. Not only is the proposed model more accurate than Bsim3v3.1 model, but also gives a better match in $\mathrm{Re}[Y_{21}]$ than that predicted by the published model [48].

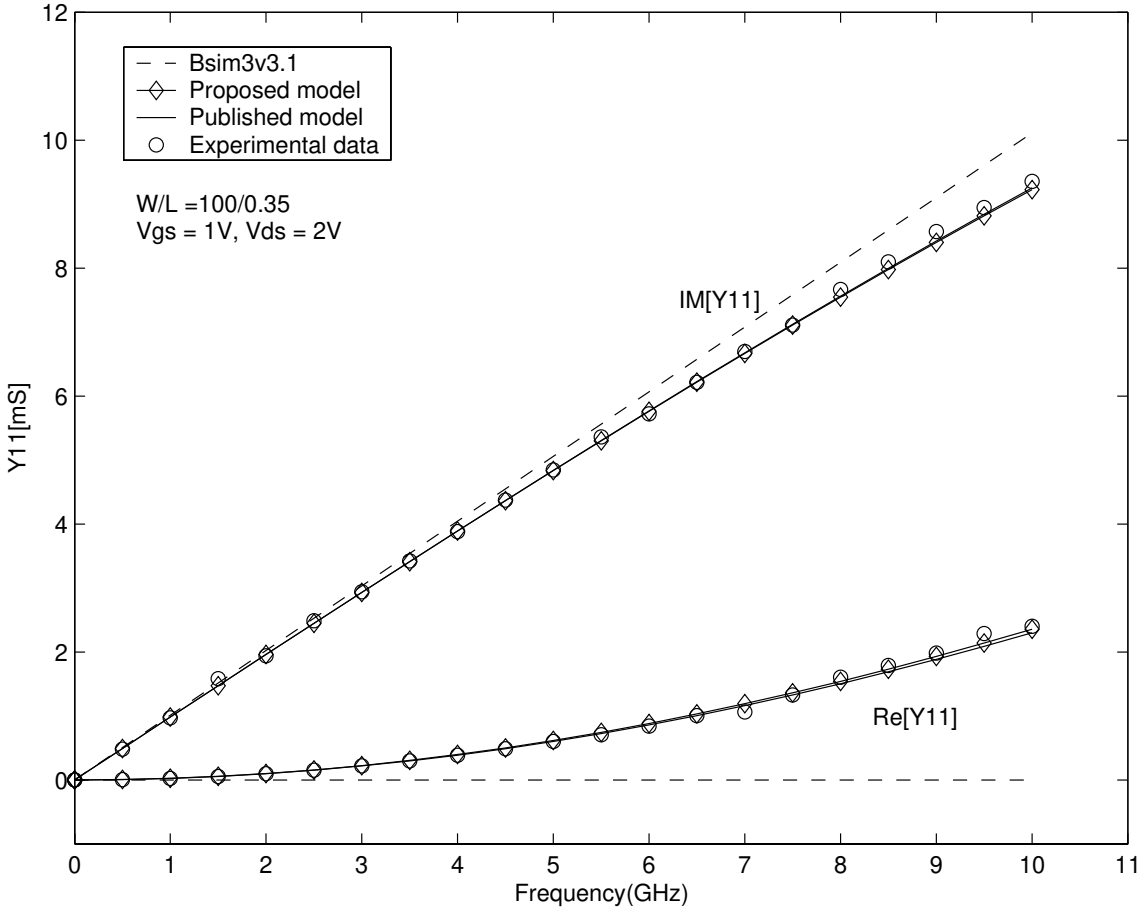

Figure 31 Y_{11} parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated Y_{11} parameter of the proposed model matches the measured data very well and is more accurate than the Bsim3v3.1 model.

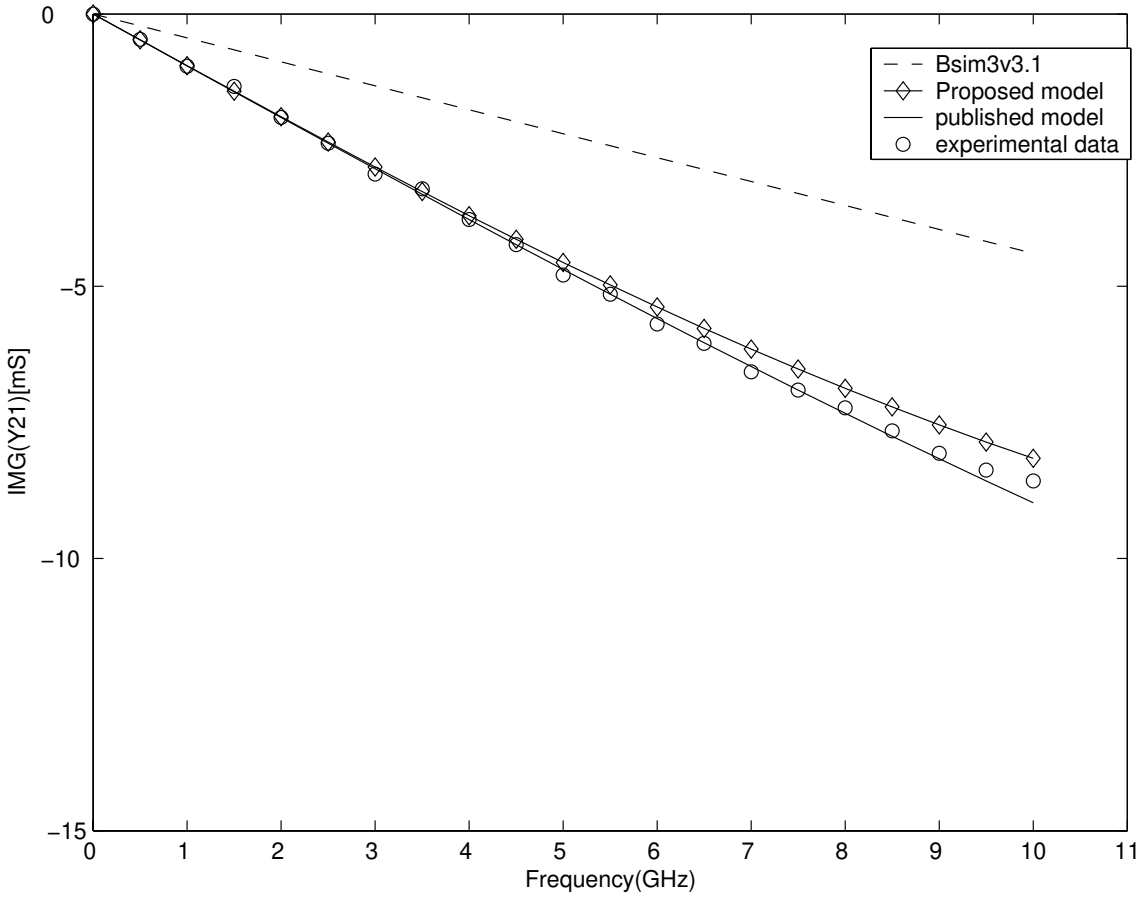

Figure 32 $\mathrm{Im}[Y_{21}]$ **parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated** $\mathrm{Im}[Y_{21}]$ **parameter of the proposed model shows a slight deviation from the experimental data above 5GHz but becomes more accurate than the published model at 10GHz.**

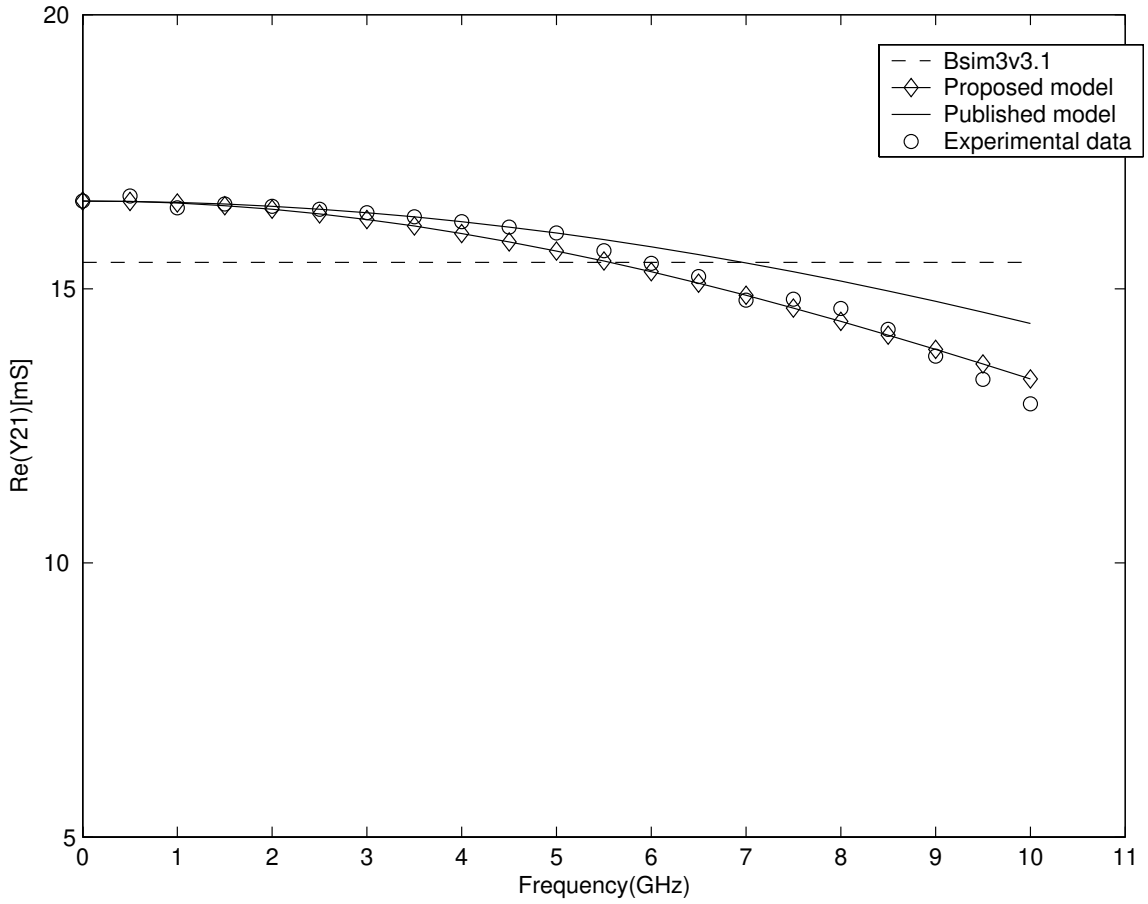

Figure 33 $\mathrm{Re}[Y_{21}]$ **parameter of measured data, proposed model, Bsim3v3.1 model, and published model. Being based on NQS analysis, the simulated** $\mathrm{Re}[Y_{21}]$ **of the proposed model is much more accurate than that predicted by the published above 5GHz.**

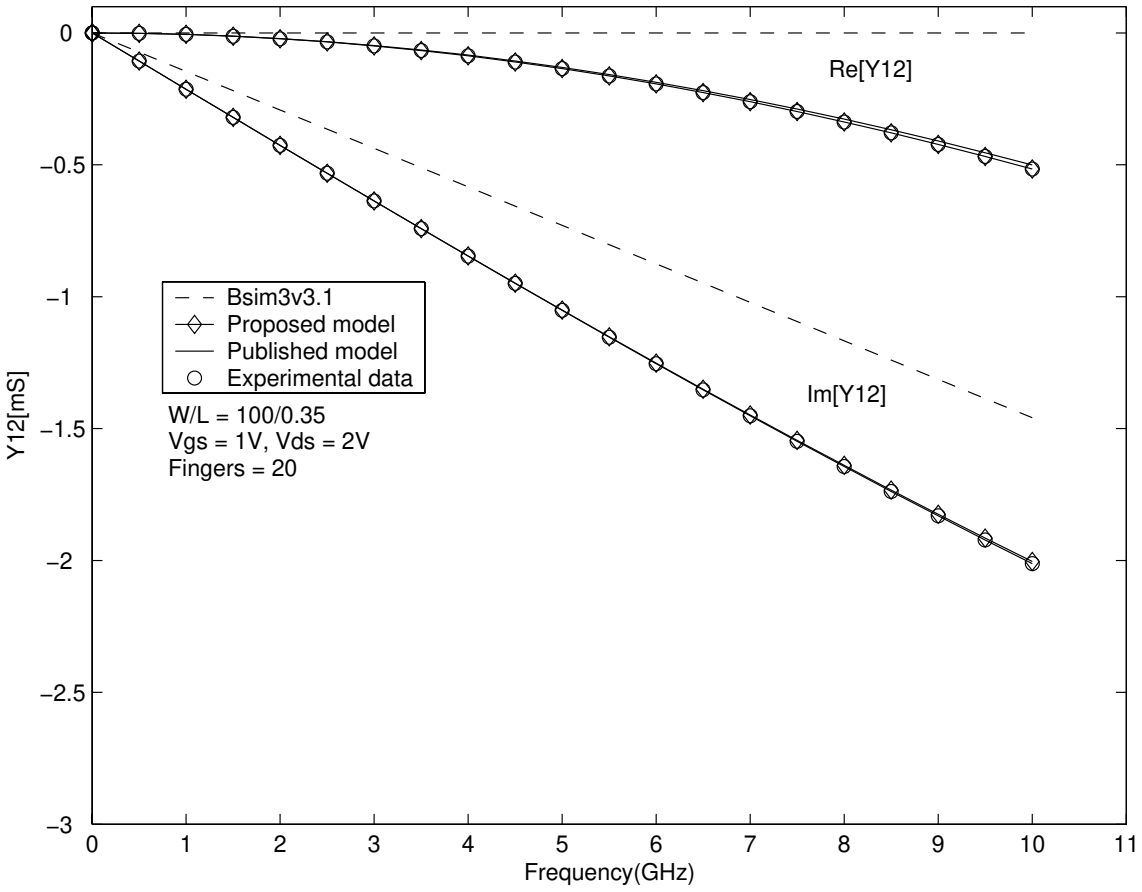

Figure 34 Y_{21} parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated Y_{21} parameter of the proposed model matches the measured data very well and is more accurate than the Bsim3v3.1 model.

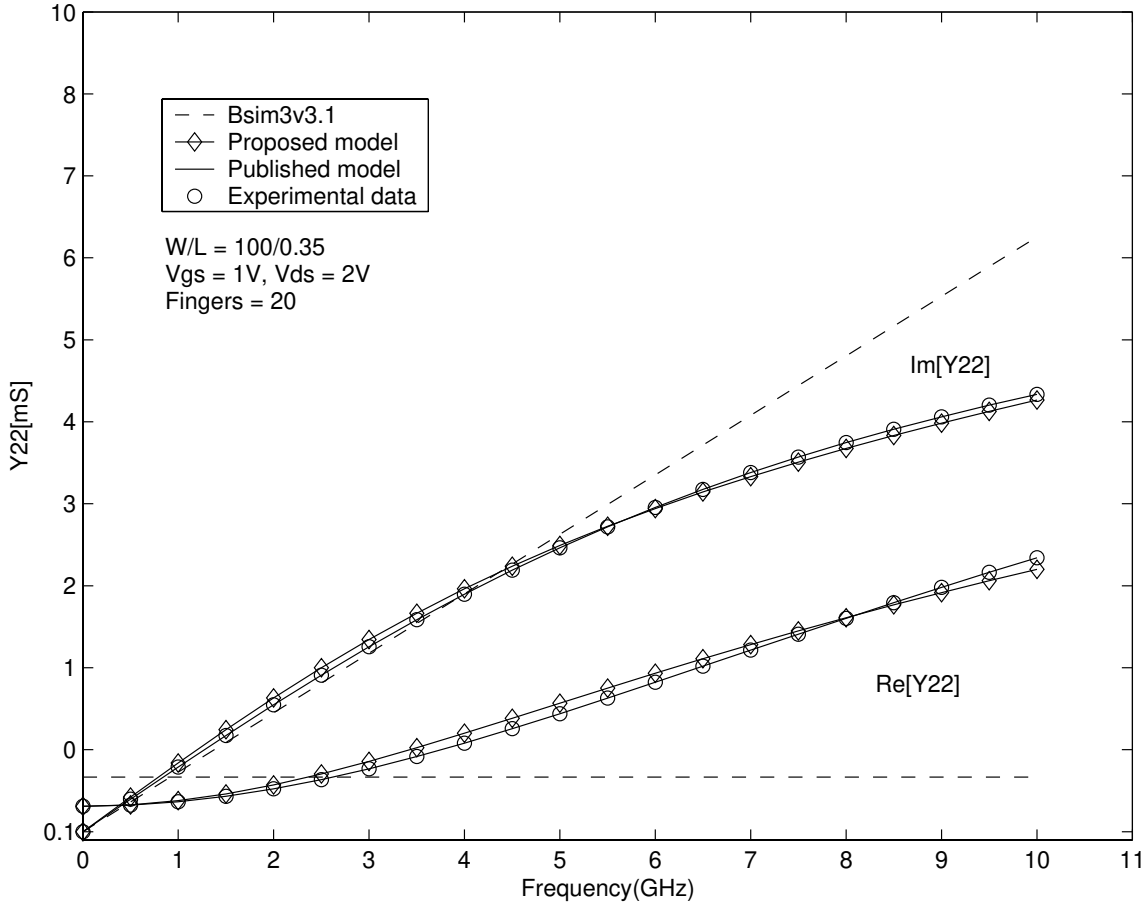

Figure 35 Y_{22} parameter of measured data, proposed model, Bsim3v3.1 model, and published model. The simulated Y_{22} parameter of the proposed model matches the measured data very well and is more accurate than the Bsim3v3.1 model.

4.4 CONCLUSION

This chapter presented a radio-frequency (RF) small signal MOSFET model together with a simple parameter extraction algorithm. The intrinsic part of the equivalent circuit (EC) is based on a first order non-quasi-static (NQS) formulation. The intrinsic and extrinsic components are extracted by performing Y-parameter analysis on the proposed model. Simple analytical expressions for extracting the effective gate-to-source and gate-to-drain channel resistances are developed for the first time. The distributed nature of the substrate resistance at high frequency is modeled using a novel lumped three-resistor T-network. Substrate signal coupling through the intrinsic body node and its effect on the output admittance are carefully examined. Simplified new expressions for the real and imaginary part of the output admittance are developed that prove essential in extracting the substrate related parameters. The physical validity of the model and the accuracy of the extraction method are verified by comparing Pspice simulation results of the EC to experimental data up to 10GHz.

5.0 CONCLUSIONS AND FUTURE WORK

The primary objective of this work was to address the Radio Frequency (RF) small-signal and large-signal models for the MOS transistor. The quasi-static (QS) and non-quasi-static (NQS) models were discussed and the assumptions used in their development were examined. The various charge components were briefly introduced and the source/drain charge partitioning was presented. The limitation of the QS approach at high frequency was investigated using the Bsim3v3.1 model. The development of a first order NQS small-signal model was briefly presented and its suitability for RF applications was indicated. The effect of the distributed gate, channel, and substrate resistances on the high frequency characteristics of the MOS transistor was examined. We proposed a Radio Frequency small-signal equivalent circuit (EC) together with an efficient parameter extraction algorithm that is necessary for the device optimization and the development of accurate large-signal models. The validity of the proposed model and the accuracy of the extraction method are verified by comparing Pspice simulation results of the EC to experimental data and the Bsim3v3.1 model up to 10GHz. It was demonstrated that the proposed model is more accurate than the Bsim3v3.1 since it is based on NQS formulation. The rest of the chapter discusses several improvements that could be made in future revisions of this work as well as potential future work that could be done in the area of RF CMOS modeling.

5.1 POTENTIAL IMPROVEMENT

Considering the effect of the gate-to-bulk capacitance and source/drain resistances on the high frequency performance of the MOS transistor can enhance the accuracy and extend the region of validity of the proposed equivalent circuit. For Lightly doped source/drain diffusion regions (LDD), the series resistances are bias-dependant. A simple extraction of these resistances from high frequency measurements is possible by biasing the transistor at $V_{DS} = 0$ [47]. At this bias, the transistor is symmetric in terms of drain and source. Consequently, the effects of the gate/bulk transconductance as well as the transcapacitances are negligible and can be neglected. The experimental data used in this work is for a transistor biased in the saturation region. Thus, it is not possible to utilize the extraction method [47]. The gate-to-bulk capacitance is needed for accurate modeling beyond 10GHz [55].

The polysilicon gate resistance (R_{gpoly}) given by equation (4-8) decreases quadratically with the number of fingers. Since the test device used in this work has 20 fingers, R_{gpoly} was neglected during parameter extraction for simplicity. A more robust parameter extraction for an arbitrary number of fingers can be developed by including the effect of R_{gpoly} on the Y-parameters.

5.2 FUTURE WORK

The model developed in this work can be implemented in circuit design tools either as a Table model or as a physical model. In the case of the table model, the model parameters can be extracted at different bias conditions using the proposed parameter extraction method. The extracted parameters are then stored as tables and are used during circuit simulation. The set of measurements conditions should be carefully considered to cover the whole range in which the device is expected to operate. An appropriate interpolation scheme must be developed that can produce the values of model parameters when the device does not operate at measurement points. In the case of the physical model, physical equations describing the behavior of each circuit element should be developed. The subtle physical mechanisms in submicrometer devices prevent the development of simple analytical expressions of the model elements without considering several simplifying assumptions. As a result, the developed equations will contain constants that have a physical origin but are usually used as fitting parameters to be extracted from the experimental data. The proposed model and the parameter extraction serve as a first stage in developing a physical model. A provocative future work is to develop analytical physical equations for each model parameter and apply the proposed extraction procedure to extract the constants and coefficients of the analytical equations.

APPENDIX

Y-PARAMETER ANALYSIS OF THE PROPOSED RF SMALL-SIGNAL MODEL

In Chapter 4, section 4.2, the simplified Y-parameters of the proposed RF small-signal EC were presented. In this appendix, a detailed analysis arriving to equations (4-14)-(4-17) will be carried out. For convenience, the proposed EC displayed is repeated below in Figure 36. Due to their small value, the polysilicon gate resistance R_{gpoly} and the gate-to-bulk capacitance C_{gb} are neglected for simplicity.

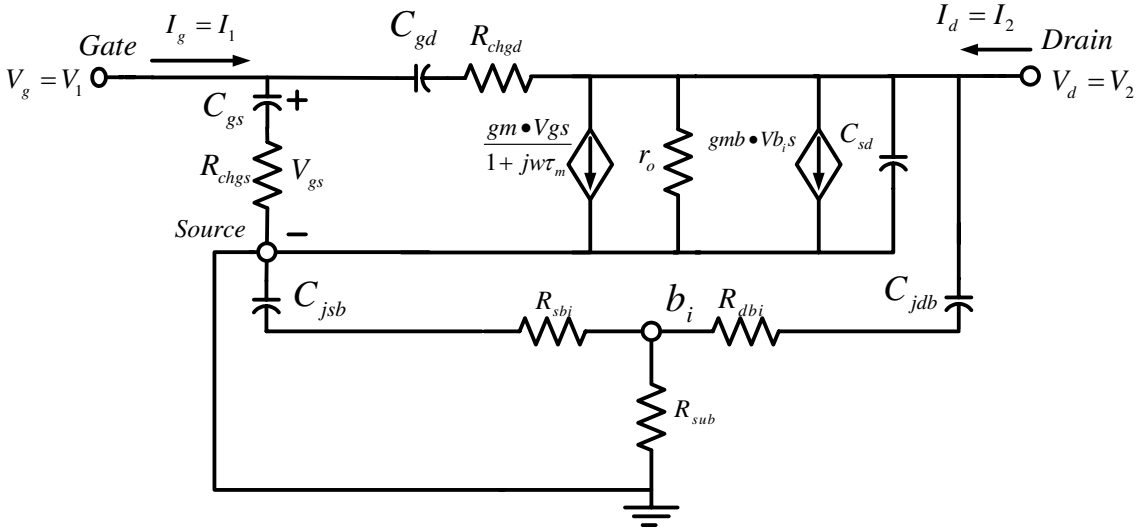

Figure 36 An RF small-signal MOSFET model based on the first order non-quasi-static effect. The effective gate-to-source and gate-to-drain channel resistance is modeled by R_{chgs} and R_{chgd} respectively. Substrate signal coupling through the intrinsic body node and junction capacitances is examined by introducing a new 3-resistor T-network.

The assumptions used in simplifying the Y-parameters are verified at the end of the Appendix. The Y-parameters of the EC shown in Figure A.1 can be derived as follows:

$$Y_{11} = \frac{I_1}{V_1}\Big|_{V_2=0} = \frac{I_g}{V_g}\Big|_{V_d=0} = \frac{jwC_{gs}}{1+jwR_{chgs}C_{gs}} + \frac{jwC_{gd}}{1+jwR_{chgd}C_{gd}}$$

$$= \frac{jwC_{gs}\left(1-jwR_{chgs}C_{gs}\right)}{1+\left(wR_{chgs}C_{gs}\right)^2} + \frac{jwC_{gd}\left(1-jwR_{chgd}C_{gd}\right)}{1+\left(wR_{chgd}C_{gd}\right)^2}. \qquad \text{(A.1)}$$

For a frequency range up to 10GHz, the following assumptions can be made: $\left(wR_{chgs}C_{gs}\right)^2 << 1$ and $\left(wR_{chgd}C_{gd}\right)^2 << 1$. As a result, the input admittance can be expressed as follows:

$$Y_{11} \approx w^2\left(R_{chgs}C_{gs}^2 + R_{chgd}C_{gd}^2\right) + jw\left(C_{gs}+C_{gd}\right). \qquad \text{(A.2)}$$

$$Y_{21} = \frac{I_2}{V_1}\Big|_{V_2=0} = \frac{I_d}{V_g}\Big|_{V_d=0} = \frac{g_m}{1+jw\tau_m} - \frac{jwC_{gd}}{1+jwR_{chgd}C_{gd}}$$

$$= \frac{g_m\left(1-jw\tau_m\right)}{1+\left(w\tau_m\right)^2} - \frac{jwC_{gd}\left(1-jwR_{chgd}C_{gd}\right)}{1+\left(wR_{chgd}C_{gd}\right)^2}$$

$$\approx g_m\left(1-jw\frac{C_m}{g_m}\right) - jwC_{gd} - w^2R_{chgd}C_{gd}^2$$

$$= g_m - w^2R_{chgd}C_{gd}^2 - jw\left(C_{dg}-C_{gd}\right) - jwC_{gd}$$

$$= g_m - w^2R_{chgd}C_{gd}^2 - jwC_{dg}, \qquad \text{(A.3)}$$

where $\tau_m = \frac{C_m}{g_m} = \frac{C_{dg} - C_{gd}}{g_m}$. In simplifying the Y_{21} parameter the following assumptions have been made: $(w\tau_m)^2 << 1$ and $(wR_{chgd}C_{gd})^2 << 1$.

$$Y_{12} = \frac{I_1}{V_2}\Big|_{V_1=0} = \frac{I_g}{V_d}\Big|_{V_g=0} = \frac{-jwC_{gd}}{1 + jwR_{chgd}C_{gd}}$$

$$\approx -w^2 R_{chgd} C_{gd}^2 - jwC_{gd}\,. \qquad \text{(A.4)}$$

Deriving an expression for the output admittance is a tedious algebraic exercise. To simplify the analysis, the EC circuit in the Y_{22} measurement configuration is shown in Figure 37.

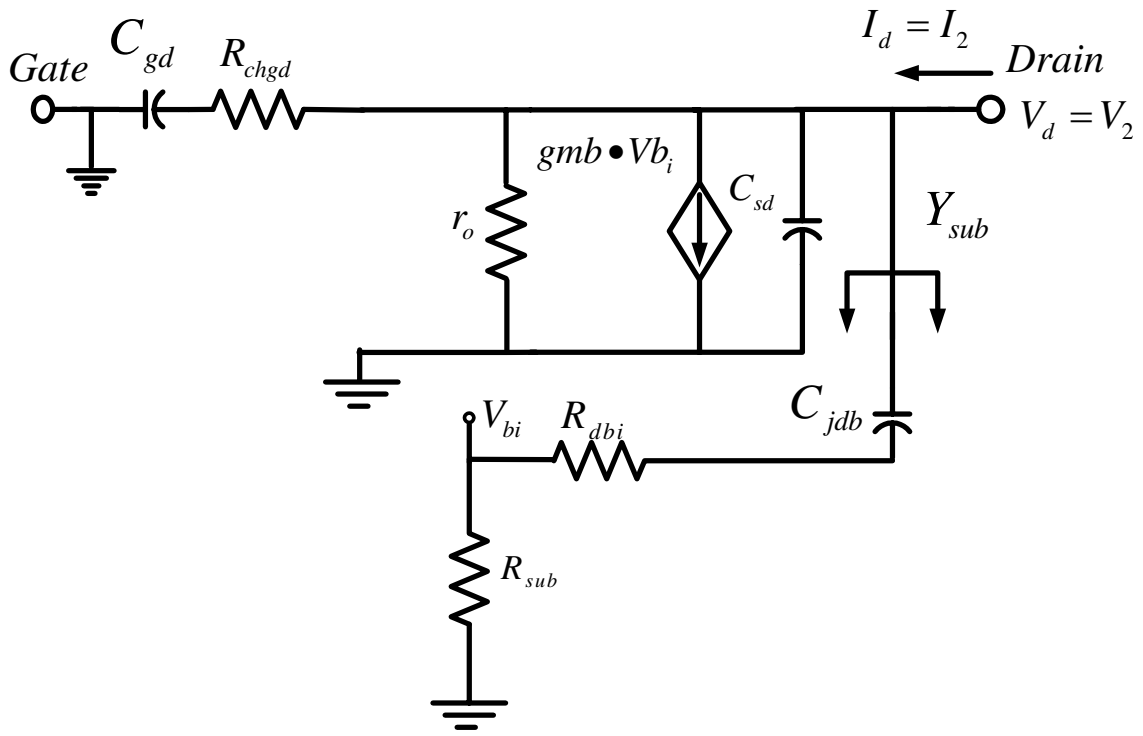

Figure 37 The proposed EC in the Y_{22} measurement setup. The effect of the source-to-bulk junction capacitance C_{jsb} and resistance R_{sbi} on the output admittance is neglected for simplicity.

The contribution of the source-to-bulk network to the output admittance is neglected for simplicity [51]. From the circuit shown in Figure 37, the output admittance Y_{22} can be derived as follows:

$$Y_{22} = \frac{I_2}{V_2}\Big|_{V_1=0} = \frac{I_d}{V_d}\Big|_{V_g=0} = \frac{jwC_{gd}}{1+jwR_{chgd}C_{gd}} + g_{sd} + jwC_{sd} + g_{mb}\frac{V_{bi}}{V_2} + Y_{sub} \tag{A.5}$$

where Y_{sub} is the admittance of the substrate looking into the drain terminal. From Figure 37, we get:

$$Y_{sub} = \frac{jwC_{jdb}}{1+jw\left(R_{dbi}+R_{sub}\right)C_{jdb}}. \tag{A.6}$$

The potential of the intrinsic body node can be expressed as function of V_2 as follows:

$$V_{bi} = \frac{jwR_{sub}C_{jdb}}{1+jw\left(R_{sub}+R_{dbi}\right)C_{jdb}}V_2. \tag{A.7}$$

From (A.6) and (A.7), the output admittance is given by:

$$\begin{aligned} Y_{22} &= \frac{jwC_{gd}}{1+jwR_{chgd}C_{gd}} + g_{sd} + jwC_{sd} + g_{mb}\frac{jwR_{sub}C_{jdb}}{1+jw\left(R_{sub}+R_{dbi}\right)C_{jdb}} + \frac{jwC_{jdb}}{1+jw\left(R_{dbi}+R_{sub}\right)C_{jdb}} \\ &= \frac{jwC_{gd}\left(1-jwR_{chgd}C_{gd}\right)}{1+\left(wR_{chgd}C_{gd}\right)^2} + g_{sd} + jwC_{sd} + g_{mb}\frac{jwR_{sub}C_{jdb}\left(1-jw\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)}{1+\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2} \\ &\quad + \frac{jwC_{jdb}\left(1-jw\left(R_{dbi}+R_{sub}\right)C_{jdb}\right)}{1+\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2} \end{aligned}$$

$$\approx jwC_{gd} + w^2R_{chgd}C_{gd}^2 + g_{sd} + jwC_{sd} + \frac{jwg_{mb}R_{sub}C_{jdb}}{1+\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2} + \frac{w^2g_{mb}R_{sub}\left(R_{sub}+R_{dbi}\right)C_{jbd}^2}{1+\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2}$$

$$+\frac{jwC_{jdb}}{1+\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2}+\frac{w^2\left(R_{sub}+R_{dbi}\right)C_{jbd}^2}{1+\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2}. \tag{A.8}$$

By curve fitting the experimental data for both the real and imaginary parts of Y_{22} to a polynomial of an arbitrary degree, it is found that the real and imaginary parts can be fitted with great accuracy to a second order and third order polynomials respectively of the following form:

$$\mathrm{Re}\left[Y_{22}\right]\approx g_{sd}+Aw^2 \tag{A.9}$$

$$\mathrm{Im}\left[Y_{22}\right]\approx Bw+Cw^3. \tag{A.10}$$

For a frequency up to 10 GHz we assume that $\left(w\left(R_{sub}+R_{dbi}\right)C_{jd}\right)^2 << 1$. Taking into account the functional form of the real and imaginary part of Y_{22}, the following simplification is considered:

$$\frac{1}{1+\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2}\approx\begin{cases}1 \text{ for } \mathrm{Re}[Y_{22}]\\ 1-\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2 \text{ for } \mathrm{Im}[Y_{22}]\end{cases} \tag{A.11}$$

Therefore the simplified expression of the output admittance that is useful for parameter extraction is given by:

$$\begin{aligned}Y_{22}\approx{}& jwC_{gd}+w^2R_{chgd}C_{gd}^2+g_{sd}+jwC_{sd}+jwg_{mb}R_{sub}C_{jdb}-jw^3g_{mb}R_{sub}\left(R_{sub}+R_{dbi}\right)^2C_{jdb}^3\\ &+w^2g_{mb}R_{sub}\left(R_{sub}+R_{dbi}\right)C_{jdb}^2+jwC_{jdb}-jw^3\left(R_{sub}+R_{dbi}\right)^2C_{jdb}^3\\ &+w^2\left(R_{sub}+R_{dbi}\right)C_{jdb}^2.\\ ={}& g_{sd}+w^2\left[R_{chgd}C_{gd}^2+\left(R_{sub}+R_{dbi}\right)\left(1+g_{mb}R_{sub}\right)C_{jdb}^2\right]\\ &+jw\left[C_{gd}+C_{sd}+C_{jdb}\left(1+g_{mb}R_{sub}\right)\right]-jw^3\left[C_{jdb}^3\left(R_{sub}+R_{dbi}\right)^2\left(1+g_{mb}R_{sub}\right)\right]\end{aligned} \tag{A.12}$$

In deriving the simplified Y-parameters of the proposed model, the following assumptions have been made:

$$\left(wR_{chgs}C_{gs}\right)^2 << 1 \tag{A.13}$$

$$\left(wR_{chgd}C_{gd}\right)^2 << 1 \tag{A.14}$$

$$\left(w\tau_m\right)^2 << 1 \tag{A.15}$$

$$\left(w\left(R_{sub}+R_{dbi}\right)C_{jd}\right)^2 << 1 \tag{A.16}$$

For a frequency up to 10GHz and considering the extracted device parameters given in Table II, weget: $\left(wR_{chgs}C_{gs}\right)^2 \approx 0.046$, $\left(wR_{chgd}C_{gd}\right)^2 \approx 0.055$, $\left(w\tau_m\right)^2 \approx 0.19$, and $\left(w\left(R_{sub}+R_{dbi}\right)C_{jdb}\right)^2 \approx 0.34$. Although (A.15) and (A.16) are barely satisfied at 10GHz, these assumptions are inevitable for parameter extraction and are used in literature [51].

BIBLIOGRAPHY

1. D.E. Ward, "Charge-based modeling of capacitances in MOS transistors," Technical Report G201-11, Integrated Circuits Laboratory, Stanford University, California, June 1981.

2. BSIM3 Version 3.1 Manual, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, 1996.

3. S. Lui and L. W. Nagel, "Small-signal MOSFET models for analog circuit design," IEEE Journal of Solid-State Circuits, vol. SC-17, pp. 983-998, December 1982.

4. C. Turchetti, G. Masetti, and Y. Tsividis, "On the small-signal behavior of the MOS transistor in quasi-static operation," Solid-State Electronics, vol. 26, pp. 941-949, 1983.

5. Y. P. Tsividis, "Reaction between incremental intrinsic capacitances and transconductances in MOS transistors," IEEE Transactions on Electron Devices, vol. ED-27, pp. 946-948, May 1980.

6. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, Semiconductor Device Modeling for VLSI, Prentice-Hall, Englewood Cliffs, N.J., 1993.

7. Y. Tsividis and G. Masetti, "Problems in precision modeling of the MOS transistor for analog applications," IEEE Transactions on Computer-Aided Design, vol. CAD-3, pp. 72-79, January 1984.

8. M. Bagheri and Y. Tsividis, "A small-signal dc-to-high-frequency nonquasi-static model for the four-terminal MOSFET valid in all regions of operation," IEEE Transactions on Electron Devices, vol. ED-32, pp. 2383-2391, Novemver 1985.

9. K.A. Sakallah, Y. T. Yen, and S. S. Greenberg, "A first-order charge conserving MOS capacitance model," IEEE Transactions on computer-Aided Design, vol.9, pp. 99-108, January 1990.

10. Y. P. Tsividis, Operation and Modeling of the MOS Transistor. New York: McGraw-Hill, 1987.

11. J. R. Brews, "Physics of the MOS transistor," chap. 1 in Silicon Integrated Circuits, Part A, D. Kahng (editor), Applied Solid State Science Series, Academic Press, New York, 1981.

12. J.A. Greenfield and R. W. Dutton, "Nonplanar VLSI device analysis using the solution of Poisson's equation," IEEE Transactions on Electron Devices, vol. ED-27, pp. 1520-1532, 1980.

13. A. Yoshii, H. Kitazawa, M. Tomizawa, S. Horiguchi, and t. Sudo, "A three-dimensional analysis of semiconductor devices," IEEE Transactions on Electron Devices, vol. ED-29, pp. 184-189, 1982.

14. A. Husain and S. G. Chanmverlain, "Three-dimensinal simulation VKSO MOSFET's: the three-dimensional simulation program WATMOS," IEEE Transactions on Electron Devices, vol. ED-29, pp. 631-638, 1982.

15. Y. A. El-Mansy and A. R. Boothroyd, "A simple two-dimensional model of IGFET operation in the saturation region," IEEE Transactions on Electron Devices, vol. ED-24, pp. 254-262, 1997.

16. R. R. Troutman and A.G. Fortino, "Simple model for threshold voltage in short-channel IGFETs," IEEE Transactions on Electron Devices, vol. ED-24, pp. 1266-1268, October 1977.

17. S. Y. Oh, D. E. Ward, and R. W. Dutton, "Transient analysis of MOS transistors," IEEE J. Solid-State Circuits, vol. SC-15, no. 4, pp. 636-643, 1980.

18. J. J. Paulous and D. A. Antoniadis, "Limitations of quasistatic capacitance models for the MOS transistors," IEEE Electron Device Letters, vol. EDL-4, pp. 221-224, 1983.

19. Y. P. Tsividis and K. Suyama, "MOSFET modeling for analog circuit CAD: problems and prospects," IEEE journal of Solid-State Circuits, vol. 29, no. 3, pp. 210-216, Mar. 1994.

20. F.-L. Ng Allen, K. Ko Ping, and C. Mansun, "Determining the Onset Frequency of Nonquasistatic Effects of the MOSFET in AC Simulation," IEEE Electron Device Letters, vol. 23, NO. 1, January 2002.

21. P. Yang and P. K. Chatterjee, "SPICE modeling for small geometry MOSFET circuits," IEEE Trans. Computer-Aided Design, vol. CAD-1, pp. 169-182, Oct. 1982.

22. M. Anderson and P. Kuivalainen, "Transit-time model for short-channel MOSFETs," IEEE Transaction on Electron Devices, vol. 40, p. 830, April 1993.

23. H. J. Park, P. K. Ko, and C. Hu, “A nonquasi-static MOSFET model for SPICE-Transient analysis,” IEEE Transactions on Electron Devices, vol. 36, pp. 561-576, March 1989.

24. C. Hu, “ Hot-carriers effects,” pp. 119-160, in Advanced MOS Device Physics, N. G. Einspruch and G. Gildenblat (editors), VLSI Electronics, vol. 18, Academic Press, New York, 1989.

25. R. M. D. A. Velghe, D. B. M. Klaassen, and F. M. Klaassen, “Compact MOS modeling for analog circuit simulation,” Digest, International Electron Device Meeting, pp. 485-488, December 1993.

26. S. Cserveny, “MOS small-signal conductances in the weak avalanche multiplication region,” IEEE Transactions on Electron Devices, vol. 39, pp. 1233-1235, May 1992.

27. R. Singh, A. Juge, R. Joly, and G. Morin, “An investigation into the nonquasi-static effects in MOS devices with on-wafer s-parameter techniques,” Proceedings of the international Conference on Microelectronic Test Structures, pp. 21-25, 1993.

28. L. –J. Pu and Y. Tsividis, “Small-signal parameters and thermal noise of the four-terminal MOSFET in non-quasi-static operation,” Solid-State Electronics, vol.33, p. 521, 1990.

29. T. Smedes, and F. M. Klaassen, “An analytical model for the nonquasi-static small-signal behavior of submicron MOSFETs,” Solid-State Electronics, vol. 38, pp. 121-130, 1995.

30. M. Chan, K. Y. Hui, C. Hu, P. K. Ko, “A Robust and Physical BSIM3 Non-Quasi-Static Transient and AC Small-Signal Model for Circuit Simulation,” IEEE Transactions on Electron Devices, vol. 45, no. 4, April 1998.

31. B. Razavi, R.-H. Yan, and K.F. Lee, “Impact of distributed gate resistance on the performance of MOS devices,” IEEE Transactions on Circuits and Systems ---I, vol. 41, pp.750-754, November 1994.

32. X. Jin, J.-J. Ou, C.-H. Chen, W. Liu, M. J. Deen, P. R. Gray, and C. Hu,“An effective gate resistance model for CMOS RF and noise modeling,” in IEDM Tech. Dig., Dec. 1998, p.981.

33. J. A. Kitti, “Salicides and alternative technologies for future ICs : Part1,” Solid State Technolo. J., pp. 81-92, June 1999.

34. I. Bahl and P. Bhartia, Microwave Solid-State Circuit Design, Wiley, New York, 1988.

35. E. Morifuji, H.S. Momose, T. Ohguro, T. Yoshitomi, and H. Kimijima, “Future perspective and scaling down roadmap for RF CMOS,” Symp. VLSI Technology, 1999, pp. 163-164.

36. G. Dambrine et al., "What are the Limiting Parameters of Deep-Submicron MOSFETs for High Frequency Applications?," IEEE Transactions on Electron Device Letter, vol. 24, No. 3, March 2003.

37. W. Liu, et al., "RF MOSFET modeling accounting for distributed substrate and channel resistance with emphasis on the BSIM3v3 Spice model," in IEDM Tech. Dig., pp. 309-312, December 1997.

38. J.-J. Ou, et al., "CMOS RF modeling for GHz communication IC's," VLSI Symp. on Tech., Dig. Of Tech. Papers, pp. 94-95, June 1998.

39. X. Jin, J.-J. Ou, C.-H. Chen, W. Liu, M. J. Deen, P. R. Gray, and C. Hu,"An effective gate resistance model for CMOS RF and noise modeling," in IEDM Tech. Dig., Dec. 1998, p.981.

40. Y. Cheng and M. Matloubian, "High Frequency Characterization of Gate Resistance in RF MOSFETs," IEEE Electron Device Letters, vol.22, no.2, Feb 2001.

41. Understanding the Fundamental Principles of Vector Network Analysis, Hewlett-Packard Application Note 1287-1.

42. J.N. Burghartz, M. Hargrove, C. S. Webster, R. A. Groves, M. Keene, K. A. Jenkins, R. Logan, and E. Nowak, "RF potential of a 0.18-μm CMOS logic device technology," IEEE Trans. Electron Devices, vol. 47, no.4, Apr. 2000, pp. 864-870.

43. J. Rudell, J. Ou, T. Cho, G. Chien, F. Brianti, J. Weldon, and P. Gray,"A 1.9 GHz Wide-Band IF Double Conversion CMOS Integrated Receiver for Cordless Telephone Applications," Proc. 1997 IEEE ISSCC, pp. 304-305.

44. A. Abidi, A. Rofougaran, G. Chang, and J. Rael, "The Future of CMOS Wireless Transceivers," in IEEE ISSCC Dig., Feb. 1997, pp. 118-119.

45. B.L. Ooi, M. S. Leong, and P. S. Kooi, "A novel approach for determining the GaAs MESFET small-signal equivalent-circuit element," IEEE Trans. Microwave Theory Tech., vol. 45, pp. 2084-2088, Oct. 1993.

46. G. Dambrine, A. Cappy, F. Helidore, and E. Palyze, "A new method for determining the FET small-signal equivalent circuit," IEEE Trans. Microwave Theory Tech., vol. 36, pp. 173-176, July 1998.

47. S. H.-M. Jen, C. C. Enz, D. R. Pehlke, M. Shroter, and B. J. Sheu, "Accurate modeling and parameter extraction for MOS transistors valid up to 10 GHz," IEEE Trans. Electron Devices, vol. 46, pp. 2217-2227, Nov. 1999.

48. I. Kwon, M. Je, K. Lee, and H. Shin, “A simple and analytical parameter extraction method of MOSFET for microwave modeling,” IEEE Trans. Microwave Theory Tech., vol. 50, pp. 1503-1509, June 2002.

49. Minkyu Je, and Hyungcheol Shin, “Gate Bias Dependence of the Substrate Signal Coupling Effect in RF MOSFETs,” IEEE Transactions on Electron Device Letters, vol. 24, no. 3, March 2003.

50. R. Sung, P. Bendix, and M. B. Das, “Extraction of High-frequency equivalent circuit parameters of submicron gate-length MOSFET’s,” IEEE Trans. Electron Devices, vol. 45, no. 8, Aug. 1998, pp. 1769-1775.

51. Y. Cheng and M. Matloubian, “On the high-frequency characteristics of substrate resistance of RF MOSFETs,” IEEE Electron Device Lett., vol. 21, pp. 604-606, Dec. 2000.

52. J. Han, M. Je, and H. Shin, “Simple and accurate method for extracting substrate resistance of RF MOSFETs,” IEEE Electron Device Lett., vol. 21, pp. 434-436, July 2002.

53. C Enz, and Y Cheng, “MOS Transistor Modeling for RF IC Design,” IEEE Journal of Solid-State Circuits, vol. 35, no. 2, pp. 186-201, 2000.

54. S. F. Tin and K. Mayaram, “Substrate Network Modeling for CMOS RF Circuit Simulation,” Proc. IEEE Custom Integrated Circuits Conference, pp. 583-586, May 1999.

55. S. F. Tin, A. A. Osman, K. Mayaram, and C. Hu, “A simple Subcircuit Extention of the BSIM3v3 Model for CMOS RF Design,” IEEE Journal of Solid-State Circuits, vol. 35, no. 4, April 2000.