Characterization and modeling of nanoscale MOSFET for ultra-low power RF IC design

THÈSE N^{O} 7030 (2016)

PRÉSENTÉE LE 27 MAI 2016

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE CIRCUITS INTÉGRÉS
PROGRAMME DOCTORAL EN MICROSYSTÈMES ET MICROÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Maria-Anna CHALKIADAKI

acceptée sur proposition du jury:

Prof. P.-A. Farine, président du jury
Prof. C. Enz, directeur de thèse
Prof. A. Vladimirescu, rapporteur
Dr A. Juge, rapporteur
Dr G. Boero, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2016

The most that can be expected from any model is that
it can supply a useful approximation to reality:
All models are wrong; some models are useful.

— George Edward Pelham Box

To the future reader

Acknowledgements

During these PhD years, I had the chance to collaborate with and meet a lot of remarkable people, who I would like to thank in the following few lines.

First of all, I would like to thank my thesis advisor, Prof. Christian Enz, for his support and guidance during my PhD research. Not only did he transfer to me a lot of his knowledge on device modeling, but he also taught me how to be methodical and insist on the detail, and I am genuinely grateful for that.

I need to acknowledge that it would have been impossible for me to carry out this research and write this thesis without the continuous support of Dr. André Juge and Dr. Patrick Scheer. I would especially like to express my sincere gratitude for their kindness, availability and insightful advices.

A big part of this work is the outcome of the teamwork with the BSIM team. I would like to thank all of its members, in both UC Berkeley and IIT Kanpur, for their fruitful collaboration and particularly Prof. Yogesh Chauhan with whom I worked closely a lot in my first PhD steps.

Besides my advisor and Dr. André Juge, I must also thank the rest of the members of my thesis jury, Prof. Pierre-André Farine, Prof. Andrei Vladimirescu and Dr. Giovanni Boero for their comments and questions on my research and the PhD manuscript.

A big thanks to my colleagues Anurag, Farzan, Lucian, Omid, Raghav, Jérémy, Vladimir, Vincent, Claudio, François and Gabrielle for making the time spent at work pleasant and amusing, and especially to those with whom we ended up being real friends enjoying life outside the office as well. I would like to give special thanks to Jérémy who kindly accepted to translate in French the abstract of this thesis. I am also thankful to Dr. Wladek Grabinski and Dr. Jean-Michel Sallese for all the discussions we had and the encouragement they offered. Many thanks also to my Greek (and not) friends with whom I shared a lot of experiences during my stay in Switzerland.

I must mention here how much I have appreciated all the administrative help that I have received during my PhD years from Lysiane Bourquin, Lucie Auberson, Marie Halm, Nicky Taveira, Catherine Falik, Christina Govoni and Isabelle Buzzi.

This PhD has been the next phase of a device modeling "journey" that started almost a decade ago in Technical University of Crete (TUC). So, I would also like thank my Diploma and Master thesis supervisor, Prof. Matthias Bucher, for the guidance he offered during my first research

steps, and my labmates at that time Angelos, Nikos and Nikos.

Of course, I cannot say enough for my parents who have been constantly supporting me and caring only for my happiness. I am also grateful to my amazing friend Eva for always providing me with a different perspective on any matter.

But, above all, I would like to thank my husband Antonios and our little star Chloe, for standing by me on all occasions, filling my life with love and joy. Their smile is just enough for me!

Maria-Anna Chalkiadaki
April 25, 2016

Abstract

The evolution and characteristics of the electronics is directly linked to the technological and societal progress. Today, there is a huge variety of electronic solutions offered, with the RF low-power systems, such as wireless sensor networks, wireless body area networks, self-powered and energy-harvesting systems and the Internet of Things (IoT), to gain more and more ground. However, these RF low-power applications set stringent constraints on the power consumption, which complicate even more the already difficult task of the RF IC design. This can be addressed by exploiting the phenomenal RF performance offered by the state-of-the-art nanoscale CMOS technologies, with impressive peak transit frequency at the order of hundreds of GHz, and sub-1 dB minimum noise figure. More specifically, most of the RF applications operate at the low GHz range, so the cut-off frequency surplus, achieved typically in strong-inversion (SI), can be traded-off with a lower power consumption by shifting the operating point to moderate- (MI) or weak-inversion (WI), while keeping the RF performance within the desired specifications.

There is an extensive work by the semiconductor community on characterization and modeling the MOS transistor at RF. Nonetheless, most of such studies focus either on the high-performance SI or on rather mature processes with respect to the contemporary state-of-the-art. In this thesis, an extensive and multi-faceted work on detailed characterization and accurate modeling of nanoscale MOSFETs for low-power operation, focusing therefore on subthreshold operation, is presented and discussed. The analysis is always performed under the perspective of the ultra-low power RF IC design. After all, the reliability of the RF IC simulation tools in this high-end range of frequencies and at very low current densities, which constitute the two extreme conditions in terms of operation of the transistor, strongly depends on the accuracy of the model used.

The dissertation follows a dual course. First, a simple, yet thorough, small-signal RF model is elaborated in order, to describe analytically the RF performance of nanoscale MOSFETs from SI down to the deep WI region, including its noise behavior. This aspect of the work allows the characterization of the RF performance and underlines the particularities of this mode of operation of the device. The study of the device characteristics under all levels of inversion reveals that WI displays different trends than SI region. Further, the analytical expressions are also used in order to form a step-by-step parameter extraction methodology. Especially, for the extraction of the RF noise model parameters, an innovative step-by-step procedure,

which is applied directly on measurements, is developed. This whole analysis is consistent with the existing advanced standard MOSFET compact models which makes it even worth for the designer.

Then, a state-of-the-art physics-based compact model (BSIM6) is used. Within this part of the work, a set of novel advancements and contributions are introduced in order for the model to be able to capture the complexity of the behavior of modern advanced CMOS technologies. The results show excellent agreement regarding all different aspects, across all modes of operation (CV, DC, RF performance), even at very low bias conditions.

The evaluation of both the modeling approaches is done in detail and uses design oriented tools and metrics, such as the $G_{\mathrm{m}}/I_{\mathrm{D}}$, the Y-parameters, the four RF noise parameters and a wide range of figures-of-merit (FoMs). Finally, a discussion around inversion coefficient (*IC*) design methodology is carried out, where several FoMs based on *IC* are modeled with the use of very simple analytical expressions requiring only few parameters. Measurements of advanced 40 nm and 28 nm CMOS technologies, with the latter to be the ultimate process for conventional bulk CMOS, are used throughout the thesis to validate all the different modeling approaches.

Key words: Advanced CMOS, nanoscale bulk MOSFET, low-power, analytical modeling, compact modeling, BSIM6, RF small-signal, RF noise, parameter extraction, model evaluation, geometrical scaling, *IC* design methodology, RF FoMs.

Résumé

Les caractéristiques et évolutions de l'électronique sont directement liées aux progrès technologiques et sociétaux. De nos jours, une immense variété de solutions électronique sont disponible avec des systèmes RF basse-consommation, tels que les réseaux de capteurs sans fil, les réseaux corporels sans fil, les systèmes à récupération d'énergie et auto-suffisants, et l'internet des objets gagnant de l'importance. Cependant, ces applications RF basse consommation imposent des contraintes strictes sur la consommation énergétique, compliquant d'autant plus la conception de circuits intégrés RF. Ce problème peut être résolu en exploitant les performances RF phénoménales des technologies CMOS actuelles, avec notamment des pics de fréquences de transit de l'ordre de centaines de GHz, et des figures de bruits minimum inférieures à 1dB. Plus particulièrement, la plupart des applications RF fonctionnent dans la gamme basse des fréquences GHz permettant d'échanger le surplus de fréquence de coupure contre une consommation plus basse en déplaçant le point d'opération de l'inversion forte à l'inversion modérée ou l'inversion faible, tout en maintenant les performances RF dans les spécifications requises.

La communauté des semi-conducteurs étudie de manière approfondie la caractérisation et la modélisation des transistors MOS fonctionnant en RF. Toutefois, la plupart de ces études se concentrent soit sur les hautes performances en inversion forte ou sur des procédés matures par rapport à l'état de l'art contemporain. Cette thèse présente un travail approfondi sur la caractérisation détaillée et la modélisation précise de MOSFETs nanométriques faible consommation, en se concentrant sur le fonctionnement sous la tension de seuil. L'analyse entière est faite dans la perspective de conception de circuits intégrés RF ultra basse consommation. Finalement, la fiabilité des outils de simulation de circuits intégrés RF dans la gamme haute de fréquences, et pour des densités de courant très faibles, qui constituent les deux conditions de fonctionnement extrêmes des transistors, dépend fortement de la précision du model utilisé.

Cette dissertation est séparée en deux. Dans un premier temps, un model RF petit signal simple mais complet est élaboré afin de décrire de façon analytique les performances RF des MOSFETs nanométriques, de l'inversion forte jusqu'à l'inversion très faible tout en incluant le comportement de bruit. Cet aspect du travail permet de caractériser les performances RF et souligne les particularités de ce mode de fonctionnent. L'étude des caractéristiques du composant sous tous les niveaux d'inversions révèle que l'inversion faible affiche une tendance différente de celle de l'inversion forte. De plus, les expressions analytiques sont

également utilisées pour créer une méthodologie d'extraction de paramètres étape par étape. Plus particulièrement, une procédure étape par étape d'extraction des paramètres du model de bruit RF appliquée directement aux mesures est développée. Toute l'analyse est cohérente avec les modèles compact standard du MOSFET, ce qui la rend d'autant plus intéressante pour le concepteur.

Deuxièmement, le modèle compact BSIM6 basé sur la physique est utilisé. Dans cette partie du travail, un ensemble de nouvelles avancées et contributions sont introduites pour permettre au modèle de couvrir la complexité du comportement des technologies CMOS modernes. Les résultats montrent un accord excellent sur tous les aspects et pour tous les modes de fonctionnement (CV, DC et performances RF), même à polarisations très faible.

L'évaluation des deux approches de modélisation est faite de façon détaillée et utilise des outils et métriques orientés design tels que $G_{\mathrm{m}}/I_{\mathrm{D}}$, les paramètres Y, les quatre paramètres de bruit RF ainsi qu'un large panel de figures de mérite. Finalement, la méthodologie de conception basée sur le coefficient d'inversion (*IC*) est discutée avec plusieurs figures de mérite basées sur le coefficient d'inversion, tout en utilisant des expressions analytiques très simples nécessitants très peu de paramètres. Des mesures dans les technologies CMOS avancées en 40 nm et 28 nm, la seconde étant le procédé ultime pour les technologies CMOS conventionnelles à substrat, sont présentées dans cette thèse pour valider les différentes approches de modélisation.

Mots-clés : CMOS avancé, MOSFET Nanométrique à Substrat, Basse-Consommation, Modélisation Analytique, Modèle Compact, BSIM6, Petit-Signal RF, Bruit RF, Extraction de Paramètres, Evaluation de Modèle, Mise à l'échelle Géométrique, Méthodologie de Conception de Circuits Intégrés, Figure de Mérite RF.

Περίληψη

Τα χαρακτηριστικά της σύγχρονης ηλεκτρονικής είναι άμεσα συνδεδεμένα με την τεχνολογική και την κοινωνική εξέλιξη. Σήμερα, διατίθεται μία πληθώρα ηλεκτρονικών εφαρμογών, εκ των οποίων αξίζει να σημειωθούν τα χαμηλής ισχύος υψηλών συχνοτήτων (RF) συστήματα, όπως τα ασύρματα δίκτυα αισθητήρων (wireless sensor networks), τα αυτοτροφοδοτούμενα συστήματα (self-powered systems), τα συστήματα άντλησης ενέργειας (energy-harvesting systems) και το "διαδίκτυο των πραγμάτων" (IoT), που κερδίζουν όλο και περισσότερο έδαφος. Αυτές οι εφαρμογές, εξάλλου, θέτουν πιο αυστηρούς περιορισμούς όσον αφορά στην κατανάλωση ενέργειας, δυσκολεύοντας το ήδη πολύπλοκο έργο της σχεδίασης των ολοκληρωμένων κυκλωμάτων (IC) σε υψηλές συχνότητες. Υπό αυτό το πλαίσιο καθίσταται δυνατό να εκμεταλλευτούμε την εξαιρετική RF απόδοση των σύγχρονων, υπομικρομετρικών CMOS τεχνολογιών που παρουσιάζουν συχνότητα αποκοπής της τάξης των εκατοντάδων GHz και δείκτη θορύβου χαμηλότερο του ενός dB. Πιο συγκεκριμένα, ένα σημαντικό μέρος από τις RF εφαρμογές λειτουργούν στο χαμηλότερο εύρος των RF συχνοτήτων, οπότε η μεγάλη τιμή της συχνότητας αποκοπής που επιτυγχάνεται σε ισχυρή αναστροφή (SI), μπορεί να ανταλλαχθεί με μία χαμηλότερη κατανάλωση μετακινώντας το σημείο λειτουργίας στην μέτρια (MI) ή την ασθενή αναστροφή (WI), ενώ ταυτόχρονα διατηρείται η RF απόδοση εντός των προδιαγραφών.

Στη βιβλιογραφία υπάρχει αρκετό υλικό που αφορά στον χαρακτηρισμό και τη μοντελοποιήση του MOS τρανζίστορ σε RF συχνότητες. Παρόλα αυτά, οι περισσότερες έρευνες εστιάζουν είτε στην υψηλής απόδοσης ισχυρή αναστροφή είτε σε πιο ώριμες, σε σύγκριση με τις σύγχρονες, τεχνολογίες. Σε αυτήν τη διδακτορική διατριβή, παρουσιάζεται μια εκτεταμένη και πολύπλευρη δουλειά πάνω στο λεπτομερή χαρακτηρισμό και στην ακριβή μοντελοποίηση του υπομικρομετρικού MOS τρανζίστορ για λειτουργία σε χαμηλή ισχύ, και συνεπώς επικεντρωμένη σε συνθήκες λειτουργίας στη χαμηλή αναστροφή. Όλες οι αναλύσεις γίνονται υπό το πρίσμα της σχεδίασης RF IC κυκλωμάτων με χαμηλή κατανάλωση. Εξάλλου, η αξιοπιστία των IC εργαλείων προσομοίωσης σε RF συχνότητες και σε χαμηλά ρεύματα, συνθήκες σχετικά ακραίες όσον αφορά στη λειτουργία του τρανζίστορ, βασίζεται στην ακρίβεια των μοντέλων που χρησιμοποιούνται.

Το διδακτορικό αυτό δομείται πάνω δύο βασικούς άξονες. Αρχικά χρησιμοποιείται, ένα απλό, αλλά πλήρες, RF μοντέλο μικρού σήματος προκειμένου να περιγραφεί αναλυτικά η RF συμπεριφορά του υπομικρομετρικού MOS τρανζίστορ, συμπεριλαμβανομένου του θορύβου,

σε όλα τα επίπεδα αναστροφής, από την ασθενή έως την ισχυρή. Αυτό το κομμάτι της δουλειάς επιτρέπει επαρκώς το χαρακτηρισμό της RF λειτουργίας και τονίζει τις ιδιαιτερότητες που παρουσιάζονται κατά τη λειτουργία σε ασθενή αναστροφή. Οι αναλυτικές εκφράσεις χρησιμοποιούνται προκειμένου να προταθεί μία καινοτόμα διαδικασία εξαγωγής παραμέτρων, η οποία μπορεί να εφαρμοστεί κατευθείαν πάνω σε μετρήσεις. Καθώς η όλη ανάλυση γίνεται σε συνέπεια με τα ήδη υπάρχοντα σύγχρονα MOSFET compact μοντέλα καθίσταται επίσης χρήσιμη και από την μεριά του σχεδιαστή.

Έπειτα χρησιμοποιείται, το πιο σύγχρονο, πρότυπο φυσικό compact μοντέλο BSIM6. Σε αυτό το μέρος της δουλειάς, ένα σύνολο από βελτιώσεις εισάγονται στο μοντέλο ώστε να μπορεί να αναπαριστά την πολύπλοκη συμπεριφορά των σύγχρονων CMOS τεχνολογιών με ακρίβεια. Τα αποτελέσματα της σύγκρισης του μοντέλου με μετρήσεις δείχνουν μία εξαιρετική συμφωνία σε όλους τους τρόπους λειτουργίας (CV, DC, RF), ακόμα και σε πολύ χαμηλές συνθήκες πόλωσης.

Η αξιολόγηση και των δύο προσεγγίσεων μοντελοποίησης γίνεται λεπτομερώς, χρησιμοποιώντας εργαλεία και μεγέθη κατάλληλα για σχεδίαση, όπως το $G_{\mathrm{m}}/I_{\mathrm{D}}$, οι Y-παράμετροι, οι τέσσερις RF παράμετροι θορύβου καθώς και ένα σύνολο από αξιολογικούς δείκτες (FoMs). Στο τέλος, γίνεται μία ανάλυση που χρησιμοποιεί το συντελεστή αναστροφής (*IC*) ως εργαλείο σχεδίασης, όπου διαφορετικά FoMs μοντελοποιούνται ως προς *IC* με τη χρήση πολύ απλών αναλυτικών εκφράσεων που απαιτούν για τον υπολογισμό τους μόνο ένα μικρό αριθμό παραμέτρων. Μετρήσεις από δύο σύγχρονες CMOS τεχνολογίες, συγκεκριμένα 40 nm και 28 nm, με την 28 nm να είναι και η τελευταία τεχνολογία που αναπτύσσεται για τα συμβατικά MOS τρανζίστορ, χρησιμοποιούνται προκειμένου να αξιολογηθούν οι διαφορετικές μέθοδοι μοντελοποίησης.

Λέξεις κλειδιά: Σύγχρονες CMOS τεχνολογίες, υπομικρομετρικό MOS τρανζίστορ, χαμηλή ισχύς, αναλυτική μοντελοποίηση, BSIM6, υψηλές συχνότητες, χαμηλό σήμα, θόρυβος υψηλών συχνοτήτων, εξαγωγή παραμέτρων, αξιολόγηση μοντέλου, γεωμετρική υποκλιμάκωση, μεθοδολογία σχεδίασης, συντελεστής αναστροφής (*IC*), αξιολογικοί δείκτες υψηλών συχνοτήτων (RF FoMs).

Contents

List of Figures

List of Tables

Symbols

Physical Parameters

k	Boltzmann's Constant
T	Absolute Temperature in Kelvin
μ_0	Low-field Mobility Constant

Voltages

U_T	Thermodynamic Voltage
V_D	Drain Voltage
V_G	Gate Voltage
V_S	Source Voltage
V_B	Bulk Voltage
V_T0	Threshold Voltage at $V_\mathrm{SB} = 0\ V$
V_TB	Threshold Voltage at $V_\mathrm{SB} \neq 0\ V$

Currents

I_D	Drain Current
$I_{\mathrm{D}_\mathrm{sat}}$	Drain Current in Saturation
i_d	Normalized Drain Current
IC	Inversion Coefficient

Conductances and Transconductances

G_m	Gate Transconductance
G_ms	Source Transconductance
G_ds	Output Conductance
g_m	Normalized Gate Transconductance
g_ms	Normalized Source Transconductance
g_ds	Normalized Output Conductance

Capacitances and Transcapacitances

C_{ox}	Oxide Capacitance per Unit Area
C_{OX}	Total Oxide Capacitance
C_{GG}	Total Gate Capacitance
C_{GG_i}	Intrinsic Part of the Total Gate Capacitance
C_{GG_e}	Extrinsic Part of the Total Gate Capacitance
$C_{GG_{eW}}$	Extrinsic Part of the Total Gate Capacitance per Width
C_{GS}	Total Gate-to-Source Capacitance
C_{GS_i}	Intrinsic Part of the Total Gate-to-Source Capacitance
C_{GS_e}	Extrinsic Part of the Total Gate-to-Source Capacitance
C_{GD}	Total Gate-to-Drain Capacitance
$C_{GD_{WL}}$	Total Gate-to-Drain Capacitance per Unit Area
C_{GD_i}	Intrinsic Part of the Total Gate-to-Drain Capacitance
C_{GD_e}	Extrinsic Part of the Total Gate-to-Drain Capacitance
C_{GB}	Total Gate-to-Bulk Capacitance
C_{GB_i}	Intrinsic Part of the Total Gate-to-Bulk Capacitance
C_{GB_e}	Extrinsic Part of the Total Gate-to-Bulk Capacitance
C_{SB}	Total Source-to-Bulk Capacitance
C_{SB_i}	Intrinsic Part of the Total Source-to-Bulk Capacitance
C_{SB_j}	Source-to-Bulk Junction Capacitance
C_{DB}	Total Drain-to-Bulk Capacitance
C_{DB_i}	Intrinsic Part of the Total Drain-to-Bulk Capacitance
C_{DB_j}	Drain-to-Bulk Junction Capacitance
C_{BB}	Total Bulk Capacitance
C_m	Intrinsic Gate Transcapacitance
C_{ms}	Intrinsic Source Transcapacitance

Resistances

R_G	Gate Series Resistance
R_S	Source Series Resistance
R_D	Drain Series Resistance
R_B	Bulk Series Resistance

Frequencies

F_t	Transit or Cut-off Frequency
f_t	Normalized Transit or Cut-off Frequency
F_{max}	Maximum Oscillation Frequency
F_c	3-dB Corner Frequency

Noise

F	Noise Factor
F_{min}	Minimum Noise Factor
NF	Noise Figure
NF_{min}	Minimum Noise Figure
R_{n}	Input-referred Thermal Noise Voltage Resistance
G_{s}	Source Conductance
B_{s}	Source Susceptance
Y_{opt}	Optimum Source Admittance
G_{opt}	Optimum Source Conductance
B_{opt}	Optimum Source Susceptance
γ_{nD}	Thermal Noise Excess Factor at the Drain
δ_{nG}	Thermal Noise Parameter at the Gate
c_{g}	Gate-Drain Thermal Noise Correlation Parameter
$\alpha_{\gamma_{\text{nD}}}$	γ_{nD} Factor Parameter
I_{nD}	Channel Thermal Noise
$S_{\text{I}^2_{\text{nD}}}$	Power Spectral Density of the Channel Thermal Noise
I_{nG}	Induced Gate Noise
$S_{\text{I}^2_{\text{nG}}}$	Power Spectral Density of the Induced Gate Noise

Other

E_{g}	Energy Band-Gap
E_{x}	Longitudinal Electric Field
E_{c}	Critical Electric Field
n	Slope Factor
v_{drift}	Drift Velocity of the Carriers
v_{sat}	Saturation Velocity of the Carriers
λ_{c}	Velocity Saturation Parameter
L_{sat}	Channel Length Demonstrating Velocity Saturation
α_{dibl}	DIBL Parameter
$\lambda_{\text{sat\_gds}}$	G_{ds} Saturation Parameter
A_{v}	Voltage Gain
$A_{\text{v}_{\text{i}}}$	Intrinsic Voltage Gain
L_{crit}	Critical Length below which the $A_{\text{v}_{\text{i}}}$ is affected by SCE
$L_{\text{max\_BW}}$	Maximum Length to achieve a specific BW
$L_{\text{max\_GBW}}$	Maximum Length to achieve a specific GBW
H_{21}	Current Gain
U	Unilateral Gain

Normalization Factors

$I_{\text{spec}_\square}$	Specific Current per Square
I_{spec}	Specific Current
G_{spec}	Specific Conductance
R_{spec}	Specific Resistance
$F_{\text{t}_{\text{spec}}}$	Specific Transit or Cut-off Frequency

Main Acronyms and Abbreviations

BJT	Bipolar Junction Transistor
BSIM	Berkeley Short-channel IGFET Model
BW	Bandwidth
GBW	Gain-Bandwidth
CLM	Channel Length Modulation
CMOS	Complementary Metal-Oxide-Semiconductor
CS	Common-Source
DIBL	Drain Induced Barrier Lowering
DITS	Drain Induced Threshold Shift
DUT	Device Under Test
ECB	Electron tunneling from Conduction Band
FD-SOI	Fully-Depleted Silicon-on-Insulator
FET	Field Effect Transistor
FinFET	Fin Field-Effect-Transistor
FoM	Figure-of-Merit
GIDL	Gate Induced Drain Leakage
GISL	Gate Induced Source Leakage
IC	Integrated Circuit
LUT	Look-Up Table
MI	Moderate-Inversion
MOS	Metal-Oxide-Semiconductor
MOSFET	Metal-Oxide-Semiconductor Field-Effect-Transistor
nMOS	n-type Metal-Oxide-Semiconductor
NQS	Non-Quasi-Static
pMOS	p-type Metal-Oxide-Semiconductor
PSD	Power Spectral Density
QS	Quasi-Static
RF	Radio-Frequencies
SCE	Sort-Channel Effects
SEM	Scanning Electron Microscope
SI	Strong-Inversion

T	Temperature
TCAD	Technology Computer Aided Design
VSAT	Velocity Saturation
WI	Weak-Inversion

1 Introduction

1.1 Motivation

Over the past decades, the microelectronics industry has managed to follow the Moore's law [1] by continuously scaling down the transistor's dimensions in order to improve its performance and reduce the power consumption and cost of the system. In recent years, the state-of-the-art downscaled CMOS processes provide smaller and faster devices, in the order of a few tens of nanometers in terms of the gate length. Although increasing the speed and density of integration of advanced Systems-on-Chip (SoCs) is important, reducing the power consumption remains a high priority for ultra-low-power applications like Internet-of-Things (IoT), Wireless-Sensor-Networks (WSN), Wireless-Body-Area-Networks (WBAN), self-powered and energy-harvesting systems, etc. These kinds of applications pose stringent constraints on power consumption, becoming the driving force towards low-power analog and RF circuit design.

Advanced nanoscale CMOS devices display very high peak transit frequency F_t of several hundreds of GHz and sub-1 dB minimum noise figure NF_{min} [2–4]. Nonetheless, the operating frequency of numerous RF applications remains significantly lower. Thanks to this very high F_t and the low NF_{min} the design of ultra-low-power RF applications has become feasible, with the nanoscale MOSFETs constituting the most attractive solution. The high transit frequency can be traded-off with lower power consumption, by shifting the operating point towards lower levels of inversion. In moderate-inversion, for advanced nanoscale devices, the transistor benefits from a minimum value of the NF_{min} while the F_t remains in the GHz range [5]. An even more aggressive policy, that would bring the operating point in weak-inversion, would further minimize the power consumption at the expense of a considerable degradation of the F_t and NF_{min}. Depending on the constraints of the application and the capabilities of the technologies, the designer is to search for the best trade-off in this design hunt.

In such extreme operating conditions, i.e. low current density and RF, an accurate and reliable description of the behavior of the MOS transistor is an indispensable tool for a successful

design. Compact models, thus, have to play their crucial role in facilitating the complex task of RF IC design. A lot of effort has already been put into characterization and modeling of the MOS transistor in RF, however, most of the published work is focused either on the high-performance strong-inversion region or on more mature processes with respect to the state-of-the-art.

The main motivation behind this research work, has been to explore whether the aggressive downscaling of MOSFET, along with the new fabrication techniques, has resulted in physical phenomena that cause a deviation from the "well-known" behavior of the transistor in state-of-the-art technologies, with a strong focus on the sub-threshold region. Towards this direction, different modeling aspects were investigated. First, there was the BSIM6 compact model which was evaluated thoroughly and enhanced where this was necessary. Then, a set of analytical expressions that allowed us to study the RF behavior of the nanoscale devices from weak- to strong-inversion, including the RF noise, were developed. Finally, we employed and further developed simple analytical expressions that can be used by designers to have a first insight in different Figures-of-Merit (FoMs) of advanced MOS transistors for the whole range of the inversion coefficient (*IC*), based on the *IC* design methodology. To fulfill this task, measurements of real devices of two commercial advanced CMOS technologies, namely 40 nm and 28 nm, were used.

1.2 The MOS Transistor

Modern electronics is the natural evolution of a field the seeds of which may be found even millenia before today [6, 7]. However, the development of electronics boosted the last century and it has been exponentially progressing the last decades. Societal evolution, in all aspects and science included, is directed by the needs that emerge and have to be taken care of. More particularly, electronics have been used in order to support the needs of mankind in different ways like sensing, communicating and reproducing information over long distances or processing and storing information in heavy workloads. For the first kind of applications, it is mainly the analogue electronics that have found room of manifesting their capabilities, while for the latter, those are mainly performed by using digital electronics, a major spin-off of the electronics industry, which is big enough to deserve the title of a field on its own. Even though there is a wide common area between analogue and digital electronics, there are also important characteristics which differentiate the two fields and which result in their distinction.

One major development that drove significantly the advancement of electronics was the invention of a three-terminal device at the beginning of the 20^{th} century [8, 9]. This element was later called triode because of the fact that it was adding a third electrode to the already available diode devices. Practically, this third electrode was connected to a grid placed inside the electron tube and between the first two electrodes, and in this way was able to control the

conduction between the other two electrodes.

This evolution was revolutionary for the field. One could say that it practically gave birth to the electronics field in the form that we know it today. In any case, everyone would agree that this step brought the field to a totally new level with much more possibilities than before. The limitation of the two terminal devices limited the whole circuit design to passive topologies. The introduction of the third terminal primed the device with a pair of input and output ports. Subsequently, the operational capabilities were drastically multiplied giving space for amplification techniques of an input signal to the output, new possibilities for sensing of an event at the input and translating this electrically at the output of the device and processing signals and information.

Even though, qualitatively the triode, as a device, is a might element that did allow electronics to go into a wide range of applications, it also contained certain limitations that where not allowing the field of electronics to reveal its maximum potential. These issues were mostly related to fabrication characteristics, such as reliability and cost. The next step that expanded further the horizons of electronics was the introduction of solid-state semi-conducting materials, such as the Silicon ($_{14}$Si), as the core element of the electron devices [10], that came out at the 1950's. Progressively with time, this novel platform of materials was accompanied by advanced fabrication techniques that crucially mitigated reliability issues and drove the implementation cost to lower and lower levels.

The first successful transistor in solid-state electronics was the Bipolar Junction Transistor (BJT) [11]. The advantages that it brought to electronics were unforeseeable. Compared to its predecessor, the triode vacuum tube, it was characterized by a much smaller size, much more reliable performance, lower cost, and more convenient fabrication technique both in terms of a single device and in the perspective of a whole electronic system as well. Roughly speaking, the BJT is a current-controlled current source. Similarly to the electron tube based triode, its active nature of a three terminal device, where one port controls the conductive state of another, kept open all the topologies that were evolved with previous technologies. This time, the limit was set by a finer detail of the performance of the element, that was mainly the nature of its input gate. Between the base and the emitter the device is essentially a forward bias diode, which, even though can be used in order to control the much higher current of the collector, its power consumption is not negligible. On the course of time, its energy footprint became its Achilles' heal that needed to be dealt with.

At that time, the Field-Effect-Transistor, came to the rescue. The MOSFET as an architecture of a device had already been envisioned, in parallel to the evolution of the vacuum tubes [12, 13]. Yet, it was not till the 1970's that the device was introduced in a wide scale in the field of electronics. The major improvement that is characterizing the MOSFET is that it transforms the transistor from a current-controlled current source, such as the BJT is, to a voltage-controlled current source. This is done by introducing an oxide layer at the input node of the device,

which is the Gate in the MOSFET and the Base in the BJT. Apart from this addition, there is an important degree of similarities between the operation of the two transistors, even though, it must be underlined that the oxide layer reconstructs from the scratch all the physical mechanisms that take place in the transistor.

The MOSFET contained a key characteristic that allowed it to be the vessel that formed the electronics in the shape that we know it today and that is the vehicle that has brought the field into the modern age. The lithographic way that the crucial dimension of the device is designed, i.e. the gate length, allowed the fabrication processes to embark on a down-scaling quest that started from the very beginning of the CMOS technologies and which is still going on even today [1]. The observation that the amount of transistors per chip will follow a certain exponential trend with time was the first instantiation of the Moore's Law, which later on was updated to a certain trend of the exponential decrease of the gate length of the MOSFET with time. This trip allowed the technologies to shrink from a minimum gate length in the order of tens of micrometers during the first CMOS technologies in the 1960's to the deep sub-micrometer technologies of todays, where the minimum gate length is in the order of few tens of nanometers, if not even lower [14, 15].

In term of performance, the reduction of the gate length of the device has resulted into a major evolution. The speed of the transistor as a switch and its maximum operational frequency of the device as an analogue element is inversely proportional to the square of this characteristic dimension [16]. The aggressive downscaling has revealed a wide spectrum of capabilities and application on which the CMOS technologies can be an ideal tool to provide high-performance solutions. This benefits both the fields of digital and analogue electronics.

Nonetheless, this miniaturization has come neither gratuitously nor effortlessly. A lengthy series of hurdles needed to be surpassed in order to be able to shrink the minimum gate length of the transistor in a way that each newer technology will be fully functional and an improvement with respect to the predecessor [17]. Subsequently, the ideal and simplified sketch of a MOSFET that one can find in a textbook of an introductory course of electronics in university level is profoundly different from an SEM picture of a cross-section of an actual device of a state-of-the-art CMOS technology. These complicated fabrication procedures result into devices whose performance is affected by an extended list of physical phenomena. In turn, the pile of the challenges that any modeling effort of such transistor has to face gets higher and higher, and the effort needed in order to end up with adequate compact models becomes increasingly demanding [18].

1.3 Compact and Analytical Modeling

Compact modeling is an essential element in modern electronics. The electron devices and circuits operate within the general framework of physics and more particularly under the

laws of electromagnetism and semiconductor's physics. Based on these laws, it is possible to study the internal operation of the electron devices and how this depends on the outside conditions. Using such laws it is possible to extract mathematical models that match the actual performance of the fabricated devices.

There is a series of different models that can be used in the framework of the electronics. The range of these tools varies mainly in terms of complexity and accuracy. The highest the complexity the finest the maximum accuracy one can get with such a tool. Nonetheless, the complexity comes with the penalty of increased computational needs and, subsequently, it is inversely connected with the speed of the model. [19]

On the one end of the simulation toolbox there are the numerical simulations (Technology CAD - TCAD) tools. With such tools the user can fully represent the whole structure of an electron device, in either a two- or a three-dimensional frame. The information provided to such tools is a full description of the architecture of the device containing all the parameters that are connected with the materials used for the device, together with detailed information on the geometry of the structure. Numerical simulators are responsible for solving the known or selected laws of physics within the frame of a specific architecture. The discretization of the whole area or volume of the device, depending on the dimensionality of the given problem, results into a grid of points in the device that are used in order to find the solution of the differential problem that is to be solved. The density of this grid is directly connected to the complexity of the system and the accuracy of the solution at the same time. A wide range of academic and commercial implementations of such tools exist and are available to the electronics community [20].

Such tools offer a unique and indispensable insight in the overall behavior of the electron device under study. The solution found by the numerical simulators contains information that does not only cover the whole electron device as an entity alone, but also it provides a full internal profile of the physical status of the device under operation. This insight provides invaluable information about the internal state of the structure and it can assist the modeling engineer to understand the operation of the device in a clearer way. On the other hand, the complexity of this simulation method is so high that even with the use of the modern computational technology it is required to spend minutes or hours in order to solve a single device. This restriction makes such tools very important for device level analysis, despite their heavy workload, yet they are inadequate for circuit level analysis.

At the other end of the spectrum of simulation modeling tools one can find a particular type of models, called look-up table (LUT) models [21, 22]. The particularity of such models is their maximum speed and utmost simplicity. The content of such LUT models is limited to a group of tables that associate certain outer and environmental conditions of the corresponding electron device, such as the bias and the temperature, with the corresponding internal and performance properties, such as the static current flow and their dynamic behavior. The

information for such models may be acquired either by performing measurements directly on the devices under study or by using TCAD numerical simulators, as described above. The LUT model contain in the tables specific bias and outer conditions. When the actual conditions of a simulation differ from the available set, then an interpolation is followed using the closest available points.

The benefits of this approach is the almost zero computational needs of the model in order to provide an answer on the behavior of a device under certain circumstances. On the other hand, the simplicity of such a modeling approach do not allow the models to reach high levels of accuracy. The first problem comes from the fact that interpolation schemes are needed when the provided bias conditions are not identical to the available ones. Practically, this covers a very wide range of cases, since the limitation on the size of the LUT do not allow matrices of infinite size to be provided. Further, such models are not able to provide accurate information on differential aspects of the performance of the device, and more particularly on higher order effects such as inter-modulation which depends heavily on the 3^{rd} and the 5^{th} derivative of the current. In general, the non-linear behavior is not accurately described with such a LUT approach.

Compact modeling lays between the two aforementioned approaches and combines the advantages of the two extremes. The development of a compact model has its roots on the physics that take place inside the electron device. Based on the physical laws and, with the aid of a series of approximations, certain formulas can be extracted that connect the internal state of the device and its electrical behavior with the outer conditions of bias, temperature and signal application. It is imperative though that these formulas are able to be solved in a way that they will provide analytical relations between the quantities of interest and that no numerical calculations and iterations are needed. For this reason, a second wave of approximations might be needed in order to remove derivatives and integrals that do not allow such formulation. This series of approximations dissociate gradually the model from the accuracy of the pure expression of the physical laws. Nevertheless, the benefit from resulting into a set of compact formulas is crucial enough and thus any additional inaccuracy of the model is fully justified. However, it must be underlined that the smarter the way these approximations are made, the higher the accuracy that the model will be able to preserve.

Compact models employ, as a general rule, a powerful characteristic. During the development of their formulation, certain physical properties of the device are associated with specific model parameters. These parameters are mostly connected with material properties, fabrication characteristics and geometrical details. Frequently, in the real device such quantities, e.g. the doping of the substrate, cannot be easily represented by one single number. However, part of the flexibility and a major advantage of the compact models is the fact that a single value for each model parameter has to be used, extracted under the criterion of the optimal behavior of the model with respect to the real devices and technology and not based on the nominal characteristics of the physical manifestation of the model parameter itself. This flexibility over

the values of the model parameters makes the model adaptable enough in order to fit a wide range of technologies and devices. This approach is mostly followed for the model parameters which are connected with various physical properties. For the geometrical characteristics of the device, especially the ones who define the design of the structure, the nominal values are used typically.

Around the core of the compact model, which is based on the physics of the behavior of the device, it is imperative, in order to expand the capabilities of a compact model to add a shell of empirical or semi-empirical formulations. This shell is important in order to cover the gaps where the pure physical approach is so complicated that it would be impossible to be simplified into a compact and analytical formulation. It is to the benefit of the compact model, such semi-empirical additions to have a limited extension with respect to the whole size of the model. Each empirical addition removes from the compact model a piece of its physical foundation, which further affects the predictability of the model and its ability to accurately describe all the multiple sides of any single phenomenon. Nonetheless, it is of high importance the fact that such additions can translate a computationally heavy, iterative part of a model to a fast formulation, or that they can fill-in holes of the model, where the physical approach is not capable of offering a compact solution.

As an outcome, the compact model contains all the essence of the electrical behavior of the electron device. This information is passed into the circuit simulators, and for this reason, it is vital that the the model is compact enough that allows the simulator to find circuit level solutions even for complicated topologies and under a wide range of environmental conditions, i.e. temperature. The compact model, at this point makes clear its status as the hidden link between the device and the circuit. Any level of circuit design requires models that can operate fast and with high levels of convergence under all types of simulation, e.g. from static to high frequency analysis and noise. Practically, the compact model starts from the device architecture and the technology itself and brings the details, via an avenue of simplifications and approximations, to the circuit design environment and makes it possible for the electron device to be integrated into a circuit. This hidden link that keeps the whole chain of electronics together and connects the two ends of the field, technology and circuit design, is one of the most important elements in the field and any advancement and improvement helps in strengthening the backbone of the electronics today.

Compact models might be a necessary tool during the design procedure, however there are other modeling approaches that can be exploited and whose value should not be underestimated. In this category lie the modeling through the use of analytical expressions. These analytical formulas are usually dedicated to model a specific part of the behavior of a device. They might not show the completeness of a compact model and the conditions under which they can be used is limited, however they are important to help towards the in-depth understanding of the working mechanisms of a specific aspect in the device behavior. Studying the device operation using simulation tools like numerical simulators or compact models

can be quite confusing, since these tools demonstrate a complete picture of the device behavior in which all the physical phenomena interact with each other, and thus it is difficult to separate the contribution of each one. The use of analytical expressions is simple and the model developer can focus on and investigate exclusively an aspect of interest. The analytical expressions can then be imported in a compact model and add a new or an improved feature to it. Furthermore, some simple analytical expressions can be utilized from engineers outside the modeling community, like circuit designers, who would like to get a first insight into the device behavior, but do not master all the different physical mechanisms.

1.4 State-of-the-Art

1.4.1 Compact Modeling

The evolution of the MOSFET compact modeling follows closely the development of the MOSFET device itself. The most important criterion for the evaluation of the compact model is whether the model includes the state-of-the-art characteristics to cover each generation of the device. The earliest implementations were able to describe accurately enough the strong inversion region of the MOSFET in static and low-frequency aspects and these points were enough in order to consider these models adequately good for the needs of that period [23]. Modern technologies include MOSFET with high end performance even in weak- and moderate-inversion and with operational frequency range that goes to tens of GHz and the compact models should also be evaluated under this framework.

At the core of the state-of-the-art MOSFET compact model of contemporary technologies, either for industrial or for academic and research needs, there are three main types of modeling approaches which differ in terms of the basic quantities that they are using in order to calculate and express the basic electrical behavior of the device. These three types constitute the threshold-voltage-based models, the surface-potential-based models and the charge-based models. Each type has widely known advanced representatives which attest the solid foundation of each approach and which does not allow any category to claim for an unquestionable predominance in the field. Historically speaking, the threshold-voltage-based models were exclusively used till the beginning of the century for industrial needs. These models have a stronger empirical side which has allowed them to be more flexible, adjustable and easier to be developed.

The surface-potential-based models and the charge-based models have a stronger physical side with fewer empirical contributions in their structure. Their development was mostly connected with the research and academic activities, or they were receiving only limited industrial share. Nevertheless, their capabilities in the commercial environment has been strongly appreciated the last years. This has brought more attention and support to the corresponding groups that lead such physics based compact MOSFET modeling activities and

has allowed them to bring their implementation at a high maturity level, capable to support the state-of-the-art IC design needs of today [24].

Threshold-Voltage-based Compact Models

The history of electron device compact modeling is strongly connected with the evolution of the circuit simulation platforms. One of the milestones in circuit simulation tools is the SPICE circuit simulator, which started in the 70's [25, 26]. Comparing this tool with other similar, contemporary tools the major and key differentiation that led to the de facto standardization of the SPICE simulator was the inclusion of the available compact models of that time directly into the software. This allowed the user to have direct access to a complete environment which contains both the simulator itself and the corresponding algorithms, and the models that can be used for the circuit simulation as well. The MOSFET technology of that period was mostly, if not exclusively, used under high bias conditions and in strong inversion. For these needs, threshold voltage based models and the square law were enough in order to describe the electronic devices.

The high-end representative of this category of models is developed and maintained by the Department of Electrical Engineering and Computer Sciences (EECS) at the University of California, Berkeley, and is used under the name Berkeley Short-channel IGFET Model (BSIM). The first version of the model was released in the 80's, while its third revision became a standard for industrial application in the 90's. However, the model was limited by its rather empirical foundations and was restricted in certain aspects of high order behavior [27]. A major evolution of the model was released under the name BSIM4 which included all the advancements needed in order ot be able to follow the state-of-the-art contemporary technologies [28]. The model has been widely used at industrial level and it is still used for a wide range technologies, even nanoscale ones.

The turning point that resulted the emerging of the physic-based models, and which brought them from the limited range needs of the research activities to the full size industrial application, was the open contest from the Compact Model Council (CMC) of the "Next-Generation Standard MOSFET Model" launched in 2004. The goal of the contest was to identify a newer model that would be able to outperform the standard model of the era, which was the BSIM4 model, and which would be able to fill in the shortcomings of this, mostly empirical, MOSFET compact model.

Surface-Potential-based Compact Models

The surface-potential based models place in their core the calculation of the surface potential of the channel after the gate bias. As it name suggests, all the important electrical properties of the device, such as the channel current, are calculated from the surface potential. The

major difficulty that this approach has to surpass is the fact that the surface potential cannot be expressed analytically explicitly after the bias of the device. In order to circumvent this obstacle one has to use numerical techniques that, even though they do not remove value from the physical approach of the problem, they add a computational load that undermines the compactness of the model [18]

Currently, there are two models in the semiconductor community which belong to this category. One is the PSP model [29] which was the outcome of the collaboration of two modeling groups, the SP model, from Pennsylvania State University [30], and the MM11 model, from Philips [31]. The combination of the two models was based on using the core calculation of the SP model for the surface potential, together with the extrinsic modeling pieces of code that the industrial MM11 had developed. The result was the PSP model, which managed to win the aforementioned CMC competition in 2006 and became a new standard model for the semiconductor community. Since 2012, the model is supported jointly by the Delft University and the NXP Research [32].

In parallel, Hirosima University has also been working on a MOSFET compact model, entitled HiSIM, which belongs also to the category of the surface-potential-based models. The team is supporting strongly the model development and the tight collaborations with industrial partners have allowed the HiSIM model to develop, improve and also be acknowledged as a standard model [33].

Charge-based Compact Models

A third way of handling the MOSFET physics is by targeting the charges in the device and more particularly focusing on the inversion charge in the channel. This approach calculates all the quantities of importance in the structure, including static currents, dynamic behavior and intrinsic noise, based on their dependence on the charges and on the charge distribution along the channel [34, 35]. To this category belongs another standardized MOSFET compact model which is a later version developed by the BSIM group, which is entitled BSIM6. To the same category belonged also the BSIM5 version of the models developed by the same group, which is now discontinued and given room to the current BSIM6 model [36].

Furthermore, there are two more models maintained by universities which rely their core on the charge-based modeling. One is maintained by a group located in Brazil under the acronym ACM, which stands for the Advanced Compact Model [37], and the other has been developed by the EPFL and later on jointly with the TUC, Greece, and is known under the name of the EKV model [38–40]. Actually, the BSIM6 model was developed using at its core the long-channel EKV model to which the BSIM4 expressions that describe the multiple physical phenomena that appear in a real, non-ideal, devices were added.

This work has been developed on the foundation of such a modeling approach and tries to

proceed further into improving such tools and bringing them to the challenging environment of the lower node CMOS technology available in semiconductor industry today.

1.4.2 Analytical RF MOSFET modeling

The above modeling works focus on the description of the physical behavior of the intrinsic part of the transistor. Primarily this includes the calculation of the static currents, which flow through the device, and the charges that are associated with the nodes of the device. However, this level of modeling hardly provides a sufficient tool for the estimation of the performance of the MOS device far from the static analysis and the low frequency regime. As the operating frequency of the simulations climbs up into the RF range it is the extrinsic part of the MOSFET that strongly influences and sets the pace of the device behavior.

The most common approach in order to extend a model into the RF regime is the careful representation of the physical enclosure of the MOSFET by a sub-circuit consisting of resistors, capacitors and diodes. The parasitic junction diodes typically have already been included for the low frequency and static representation of the device as they affect even the leakage currents at the static regime, together with the onset of the AC performance of the transistor. Nonetheless, their role in terms of the RF performance is further upscaled and their accurate modeling becomes of higher importance. The resistors that are incorporated are responsible for describing the resistive paths of the extrinsic part of the device covering mostly the resistance of the gate material and the resistance that the semiconductor has within the substrate. The series resistances of the drain and source nodes are as well influencing also the static behavior of the device, as they lay upon the path of the current of the channel. Similarly, with the diodes, the role of the resistances in RF becomes more important and their exact value changes more drastically the overall performance of the device [41]. The capacitances beyond the junction interfaces, such as the fringing and the overlap parasitic capacitances manifest their existence primarily in the RF regime. Together with the resistances, and with the added nodes added from the development of the RF sub-circuit, allow the macro-model to capture the delay and high frequency effects that are not possible to be described by a plain and first order, static model.

The components that are included in the extrinsic sub-circuit should be chosen carefully. Although it can be tempting to account for all the physical components that exist in a MOSFET in detail, this would lead to a very complex equivalent sub-circuit that would not only increase the simulation speed of the model due to the additional nodes that are introduced, but also it would be very difficult, if not impossible, to extract the values of the components from measured data. Furthermore, the sub-circuit, especially if it is to be implemented in a compact model, should maintain its validity no matter the technology or the device layout that it is used for. A universal model should surpass the need for a specific solution for each process and geometry. It is therefore, once again, an issue of finding a balanced trade-off between

complexity and accuracy.

A suitable representation of the extrinsic parasitic network around a MOSFET, considering at the same time accuracy and simplicity, is the outcome of an extensive characterization of the device RF behavior with the use of simple analytical expressions. Although a lot of effort has already been put into characterizing and modeling the MOS transistor at RF, including RF noise, most of the published work in the literature is focused either on the high performance strong inversion region or on more mature processes with respect to the state-of-the-art or the modeling part is not expressed in consistency with the standard compact MOSFET models [3, 42–55].

Based on the valuable work that has already been published, this research focuses on proposing RF analytical expressions that are valid for nanoscale devices, in the whole *IC* range and are compatible with the standard compact MOSFET models.

2 Technology, DUT and Measurements Description

2.1 Introduction

In this thesis, there are two basic modeling "tools" that are used and evaluated. On one hand there is the analytical model, whose benefits are its versatility and its efficiency in providing a deep insight into the behavior of the device, and on the other hand the BSIM6 model, whose key features are its completeness and its feasibility to be integrated into circuit simulators. In this chapter we provide a short description regarding the processes, DUTs and measurements that were used in order to evaluate the different features and capabilities of the two modeling approaches. At this point it should be clarified that the measurements used and shown in this thesis were performed, following the standard industrial procedures, in the fabrication facilities by the production company. Due to reasons of confidentiality the name of the company cannot be provided.

2.2 Technology Details

Two state-of-the-art and highly advanced CMOS technologies have been employed in order to assess the qualities of the models and their accuracy. The nominal lengths of these technologies are 40 nm and 28 nm. From both of these technologies the standard bulk CMOS devices have been used as the devices under test (DUT). Regarding the processes, it is worth mentioning that contrary to the polycrystalline silicon (or polysilicon) and silicon dioxide that is used in older processes, in the 28 nm technology the gate stack is fabricated using a high-k dielectric (HK) combined with a metallic gate (MG) in order to optimize its performance at this aggressive downscaled regime.

The high-k metal-gate (HKMG) approach is an emerging technological booster that is used in order to permit the continuation of the Moore's Law [1], by enabling faster switching speeds while reducing the device power consumption. The traditional oxide material of silicon dioxide has become too thin in advanced technologies and would become even thinner in order to be

able to maintain a good enough dynamic coupling between the gate and the channel. However, the miniaturization of the oxide in terms of its thickness naturally leads to the increase of the tunneling current through the oxide. It can be underlined here that this tunneling gate current, even though negligible in older technologies, is increasing exponentially with the decrease of the oxide thickness. For technologies at this low node the gate leakage has become a significant issue and the usage of high-k materials can strongly improve the performance of the device. High-k materials that are used as replacement of the silicon dioxide have a dielectric constant a few times larger than that of the silicon dioxide. This allows them to offer an equivalent capacitive coupling between the gate and the channel with a significantly higher oxide thickness. This characteristic allows the reduction of the parasitic gate leakage while at the same time there is room for improvement on the effective oxide thickness of the gate stack.

The metallic gate adds further to the improvement of the performance of the device since the polysilicon can no longer provide equally advanced results with respect to the ideal metal. The limited charge availability in the gate node results into a depleted layer at the interface between the oxide and the gate which effectively increases the thickness of the structure and deteriorates the maximum speed of the device. The return to the metallic gate, the material type which, historically speaking, has given the first letter to the name of the MOS device, has again become the most adequate choice for the gate stack.

On the other hand, this booster does not come cost-free. The silicon dioxide and the poly-crystalline silicon offered a natural simplicity in terms of materials used in the process since they are based on the silicon which is already available in the semiconductor fabrication process. In this direction the usage of new materials makes the whole fabrication process more challenging and increases its complexity. Nonetheless, the benefits obtained justify the technological step in terms of cost and complexity.

An important note that should be kept in the mind is that this particular node of the 28 nm is regarded as the last technology node for the conventional planar bulk CMOS devices. The classical planar MOSFET geometry has started its industrial life already decades ago and, partly due to its planar simplicity, its low cost and its two-dimensional straight forward geometrical down-scaling, has so far triumphed in the field of semiconductors. However, at this point an impassable barrier seems to exist. The short channel effects make the single sided planar bulk approach no longer sufficient for accurate current control. Thus, more advanced structures, are already being fabricated for the continuation of the downscaling journey of the MOSFET. The Fully-Depleted Silicon-on-Insulator (FD-SOI) [56] and the Fin-Field-Effect-Transistor (FinFET) [57] architectures are among the devices that can offer strong enough control of the gate over the channel in order to extend the shrink of the MOSFET even further.

A few more details can be mentioned regarding the CMOS technologies that are studied. For the 40 nm process the maximum supply voltage that can be applied is set to $V_{DD} = 1.1\ V$, while the actual minimum drawn length of the shortest possible MOSFET is the same with the

nominal length of the technology $L_{\text{min}} = 40$ nm. For the 28 nm process the maximum bias that can be applied over the channel is slightly lower and set to $V_{\text{DD}} = 1.0\ V$, while the actual drawn length of the shortest MOSFET of the process is higher than the nominal length of the technology and more precisely $L_{\text{min}} = 32$ nm.

2.3 DUT Details

The study of the models extends to two different types of devices. The basic analysis starts, as always, from the typical and most simple DC devices, while, on a next step, devices operating in the RF regime are studied. For the RF analysis, special RF geometries, that follow guidelines which optimize their dynamic behavior are used in order to be able to exploit the high frequency capabilities of the low node technology at the maximum degree. The main advantages that the RF structures can offer with respect to their simpler DC counterparts are in terms of higher gain at RF and higher cut-off frequencies. This optimized performance makes them more adjusted for mm-wave applications. These advantages come with the penalty of a larger footprint of the structure and a higher complexity of the layout of the device. However, this drawback is negligible compared with the tangible RF out-performance that they offer over the DC devices. Generally speaking, the total number of the RF devices needed in a circuit is not high enough and so their size is not the most important parameter for the optimization of the circuit.

For the technologies under study, each RF DUT consists of a series of multi-finger devices connected all in parallel, isolated from the rest of the die by a deep buried n-well layer (typically connected to the ground for nMOS devices). More specifically, the minimum length RF devices of each process have $M = 6$ (devices in parallel), $N_{\text{f}} = 10$ (number of fingers per device), $W = 2$ μm ($W_{\text{tot}} = M \cdot N_{\text{f}} \cdot W = 120$ μm) and $L = 40$ nm and 28 nm. In order to provide a clear picture to the reader, a visualization of the layout of such a device is depicted in Fig. 2.1. In this particular instance a rather small 2x2 structure ($M = 2$ and $N_{\text{f}} = 2$) is drawn.

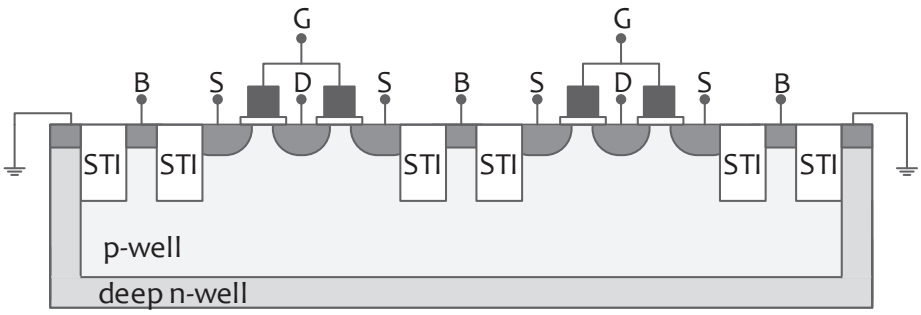


Figure 2.1 – Cross section of a 2x2 nMOS RF DUT across L.

2.4 Measurement Details

The validation of a model is achieved through the comparison between the predictive simulation that is based on the model and the actual measured behavior that is extracted from fabricated devices. A series of different types of measurements are required in order to be able to cover the full spectrum of the electrical behavior of the MOSFET device.

The most fundamental set of measurements that describe the electrical behavior of the MOS device is the static DC measurements. In this setup a certain set of biases is applied to the device and the currents that flow through the nodes of the device are recorded. At this type of measurements the series resistances that appear from the probes and the metalization used in order to bias the device should be kept to minimum as their impact on the measured values cannot be avoided. However, the minimization of these metal connections at a negligible level is not difficult to achieve. If needed, these resistances could be included in the simulations as well, however, their relatively much lower value than the DUT do not make this necessary. Typically, these measurements are used in order to study the behavior of the device from weak-inversion to strong-inversion and both in the linear mode and saturation. Furthermore, it is meaningful to extend the DC analysis in accumulation, in order to study the Gate-Induced-Drain-Leakage (GIDL), the Gate-Induced-Source-Leakage (GISL) whose influence is more prominently shown in that regions partly due to the absence of the channel current. DC measurements in accumulation region are also useful to study some of the gate tunneling current components. Different body bias conditions should be taken into account while also measuring the device under different temperatures is essential in order to obtain a picture of the behavior of the device in the full temperature range that appears in realistic applications.

Another major tool for the characterization of the MOSFET is the CV analysis, i.e. the capacitive measurements between its nodes. Such measurements typically cover the full range of the gate biasing from the positive (strong-inversion) to the negative (accumulation) side. The capacitive measurements are able to reveal more clearly certain information that is partly hidden in the DC behavior. These aspects include the body factor of the device, or the doping concentration of the substrate, while also it can show certain aspects of the behavior in accumulation, such as the Flat-Band Voltage (V_{FB}). Such measurements are typically done within a frequency range from tens of kHz up to a few MHz. For the capacitive measurements a single calibration step of the measurement setup is possible to be made where all the parasitic and extrinsic part of the setup is measured with no DUT being connected to the measuring equipment. The capacitive load measured is considered as the reference measurement level and all the additional capacitance which is measured after the contact with the DUT is the actual capacitive load that the DUT alone adds to the measurement.

The third measurement environment that brings into the discussion the information of the high frequency behavior are the S-parameters measurements. For the S-Parameter measurements the device is regarded as a two-port network where the input port (namely 1) is the port

between the Gate (G) and the Source (S), while the output port (namely 2) is the port between the Drain (D) and the Bulk (B). Typically, the Bulk node for these measurements is shorted to the Source node. The S-parameters are small-signal measurements which take place under a certain static bias condition. This bias should sweep over the whole inversion levels and covering both linear operation and saturation. The physical meaning of the S-parameters is based on the power transmission of the RF signal. For example, the S21 parameter shows how much of the power of the signal that is applied into the port 1 reaches the port 2. Complementary to this, the S11 parameter is how much of the energy transmitted into the port 1 is reflected back to the signal source. Similar definitions apply to the S12 and S22, respectively. The S-parameters are measured in a wide frequency range that extends much beyond the capabilities of the CV measurement systems and which reach, depending on the equipment the range of tens of GHz. In this thesis measurements up to 50 GHz are used.

The nature of the RF measurements requires a more advanced de-embedding technique in order to isolate the contribution of the DUT itself and remove all the influences of the parasitic environment. There is a series of de-embedding procedures which are all based on measuring reference (dummy) structures, which maintain the same parasitic environment as the DUT but without including the actual DUT. Regarding the RF measurements used in this thesis, the de-embedding of the S-parameters is based on 'OPEN' and 'SHORT' dummy test structures [58]. A 'SHORT' structure is created by replacing the DUT with a metal piece that short the ports with each other and to the ground, while an 'OPEN' structure is created by removing completely the DUT from the PAD and leaving the nodes of the ports floating. For the above device the 'SHORT' one is responsible for removing the influence of the in series parasitics at each port and the 'OPEN' device removes the parasitics that are in parallel connection to the device and works in a similar manner with the calibration in the CV measurements. The de-embedded measurements should then match the simulations which are not influenced by this parasitic environment of the measurement setup and thus not taken into account in the simulations.

Within the RF regime the study can be completed by performing noise measurements on the device. The noise measurements are performed again under a static bias profile in the GHz regime, which in this thesis reach up to 18 GHz. The resulted values of the measurement are the four noise parameters in the RF regime, namely: F_{min}, R_{n}, G_{opt} and B_{opt}. Similar to the S-parameters, a procedure based on 'OPEN' and 'SHORT' dummy test structures [59] can be applied for the de-embedding of the RF noise measurements. In order to extract the four RF noise parameters the NF50 method is used [60]. This measurement adds the last piece in a rather complete picture of the electrical performance of the device. Corresponding RF-noise simulations are performed in order to capture also the noise aspects of the electrical behavior of the MOSFET.

Analytical Small-Signal RF Modeling Part I

3 Analytical RF Modeling

3.1 Introduction

Nowadays, low-power wireless applications, such as wireless sensor networks, wireless body area networks and the Internet-of-Things (IoT), are setting stringent constraints on power consumption, in particular on the part of the radio operating at RF. This can be partly addressed by taking advantage of the phenomenal RF performance obtained from the continuous downscaling of CMOS technology [61]. Indeed, advanced nanoscale transistors achieve impressive peak transit frequency, reaching typically several hundreds of GHz, and extremely low thermal noise, sub-1 dB minimum noise figure, for a nanoscale CMOS process [2–4]. Since the above mentioned applications are mostly running in the low GHz frequency range, the very high peak transit frequencies achieved in strong-inversion can be traded-off with power consumption by shifting the operating point towards moderate- or even weak-inversion.

Although a lot of effort has already been put into characterizing and modeling the MOS transistor at RF, most of the published work in the literature is focused either on the high performance strong-inversion region [3, 43, 62–65] or on more mature processes [42, 47, 50, 66, 67] with respect to the state-of-the-art or usually the modeling part is not expressed in consistency with the standard compact MOSFET models [45, 46, 68, 69] or the result demonstration is limited [70, 71].

In this chapter, we present a thorough small-signal RF characterization and modeling of the advanced nanoscale transistor from strong-inversion down to deep weak-inversion region. We explore whether classical equivalent circuits and models can still provide an accurate representation of the small-signal RF behavior of the nanoscale MOSFET at low inversion levels. We show that a very basic RF equivalent schematic can lead to simple analytical expressions, providing an accurate description of the small-signal RF behavior.The analytical expressions for the Y-parameters and the related RF Figures-of-Merit (FoMs), H_{21}, U, F_{t} and F_{max}, are demonstrated. Additionally, we describe the procedure for the direct extraction from measurements of the components of the small-signal equivalent circuit. The advantage of

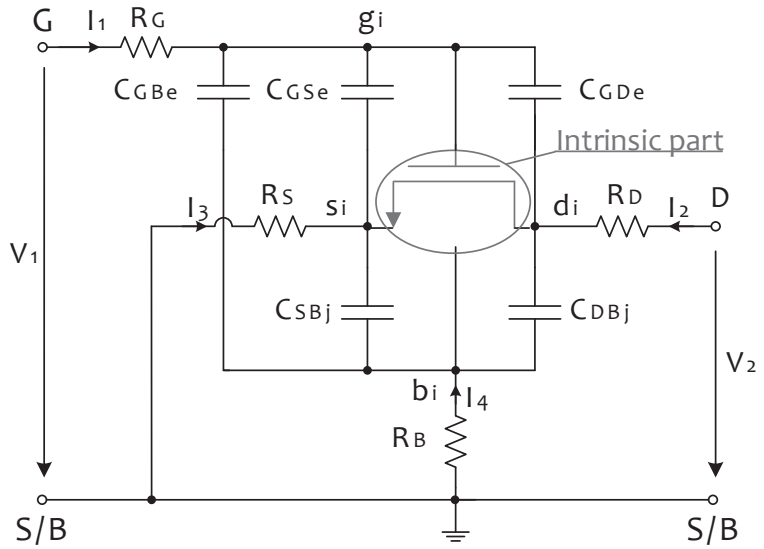
the proposed parameter extraction methodology is that it remains consistent with existing standard compact MOSFET models.The analytical expressions are validated against measurements of two commercial state-of-the-art CMOS processes, namely 40 nm and 28 nm, even in deep weak-inversion; a region where models are typically not validated. We should point out here, that although the analytical expressions are only validated for nMOS devices, the general study of the technologies shows that pMOS devices display similar behavior to nMOS [72], and therefore we expect the analytical expressions to be valid for pMOS devices as well [5].

3.2 MOSFET Equivalent Circuit at RF

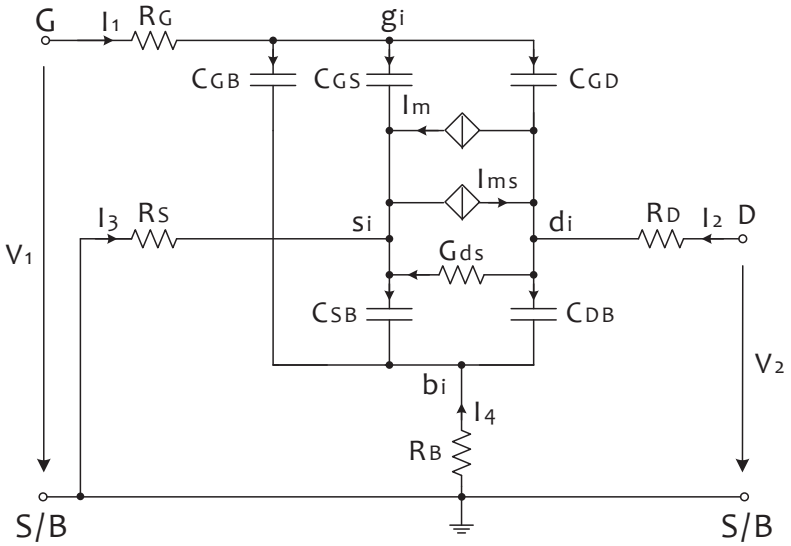
As the operating frequency increases into the GHz range, the contribution of the extrinsic part of the transistor (the part outside the channel region) dominates that of its intrinsic counterpart. Consequently, an RF model should account for both the intrinsic and the extrinsic components [73, 74]. It is certainly possible to model the RF MOS transistor in great detail accounting for all physical components, but this would lead to a very complex equivalent circuit whose components would be very difficult, if not impossible, to extract individually from measured data. Additionally, such an equivalent circuit would highly depend on the technology and the device layout, yet, it is mandatory to have a universal model surpassing the need for a specific solution for each process and geometry. It is therefore crucial to find a balanced trade-off between complexity and accuracy.

Fig. 3.1a shows a generic and simple equivalent circuit that can be used for different technologies and device layouts, while at the same time, it offers a good compromise between accuracy and computing efficiency and can be easily implemented as a spice sub-circuit. It consists of the intrinsic part of the transistor and the parasitics components that affect the behavior of MOSFET at RF, namely, the gate resistance R_G, the substrate resistance R_B, the source/drain series resistances R_S, R_D, the extrinsic capacitances C_{GSe}, C_{GDe}, C_{GBe}, which include both the overlap and the fringing capacitances, and the junction capacitances C_{BSj}, C_{BDj}. Note that, advanced nanoscale RF MOSFETs are usually large multi-finger devices designed to meet the RF requirements in terms of gain, and thus, using a single substrate resistance in the RF equivalent circuit [75] can be usually sufficient [39]. Normally, for RF measurements, MOSFETs are used in a typical two-port configuration, with S and B connected to the ground ($V_S = V_B = 0$ V). The same configuration is adopted in the equivalent schematic, but, the parasitic components that are related to the test structure, e.g., pad capacitances, lead series resistances/inductances etc., are not included, since all the data used in this work are de-embedded measurements.

In most RF applications, a MOSFET usually operates in the saturation region. The quasi-static (QS) RF small-signal equivalent circuit in saturation, corresponding to Fig. 3.1a, is shown in Fig. 3.1b. In this circuit the capacitances include the contribution of both the intrinsic and the



(a) Equivalent sub-circuit of an RF MOSFET



(b) Quasi-static small-signal equivalent circuit of an RF MOSFET in saturation

Figure 3.1 – (a) Equivalent sub-circuit of an RF MOSFET and (b) quasi-static small-signal equivalent circuit of an RF MOSFET valid in saturation.

extrinsic capacitances, given by:

$$C_{GS} = C_{GSi} + C_{GSe}, \tag{3.1a}$$
$$C_{GD} = C_{GDi} + C_{GDe}, \tag{3.1b}$$
$$C_{GB} = C_{GBi} + C_{GBe}, \tag{3.1c}$$
$$C_{SB} = C_{SBi} + C_{SBj}, \tag{3.1d}$$
$$C_{DB} = C_{DBi} + C_{DBj}. \tag{3.1e}$$

The currents generated by the voltage-controlled current sources (VCCSs) can be calculated using [39]:

$$I_m = Y_m \cdot [V(gi) - V(bi)], \tag{3.2a}$$
$$I_{ms} = Y_{ms} \cdot [V(si) - V(bi)], \tag{3.2b}$$

with

$$Y_m = G_m - j\omega C_m, \tag{3.3a}$$
$$Y_{ms} = G_{ms} - j\omega C_{ms}, \tag{3.3b}$$

being the gate transadmittance and source transadmittance, respectively. In Eq. (3.3a) and Eq. (3.3b), G_m and G_{ms} are the gate and source transconductances and C_m and C_{ms} the gate and source transcapacitances [39] .

3.3 Y-parameters Analysis

Using the RF small-signal circuit of Fig. 3.1b, we can carry out a Y-parameters analysis in order to determine the analytical expressions for the Y-parameters suitable to characterize the small-signal RF behavior of a MOSFET. The Y-parameters of a two-port network can be calculated from:

$$Y_{11} = \frac{I_1}{V_1}\Big|_{V_2=0}, \tag{3.4a}$$
$$Y_{12} = \frac{I_1}{V_2}\Big|_{V_1=0}, \tag{3.4b}$$
$$Y_{21} = \frac{I_2}{V_1}\Big|_{V_2=0}, \tag{3.4c}$$
$$Y_{22} = \frac{I_2}{V_2}\Big|_{V_1=0}, \tag{3.4d}$$

with I_x and V_x the current and voltage across port-x. In our case, port-1 is between the G and the S and port-2 between the D and the B. In order to simplify the analytical expressions for

the Y-parameters first we assumed that:

$$\omega^2\Big(C_{\mathrm{BB}}^2 R_{\mathrm{B}}{}^2 + R_{\mathrm{G}}\big(2C_{\mathrm{GB}}{}^2 R_{\mathrm{B}} + C_{\mathrm{GG}}{}^2 R_{\mathrm{G}}\big)\Big) + \omega^4 R_{\mathrm{G}}{}^2 R_{\mathrm{B}}{}^2\big(C_{\mathrm{GB}}{}^2 - C_{\mathrm{BB}}C_{\mathrm{GG}}\big)^2 \ll 1, \tag{3.5}$$

where

$$C_{\mathrm{BB}} = C_{\mathrm{DB}} + C_{\mathrm{SB}} + C_{\mathrm{GB}} \tag{3.6}$$

is the total bulk capacitance. Eq. (3.5) can be valid for operating frequencies up to the low THz range. Then we neglected: (i) all the higher than second-order terms, (ii) the least dominant terms and (iii) the NQS (Non-Quasi-Static) effects. Finally, the simplified expressions for the Y-parameters in saturation can be derived as:

$$Y_{11} \approx \omega^2\big(C_{\mathrm{GB}}{}^2 R_{\mathrm{B}} + C_{\mathrm{GG}}{}^2 R_{\mathrm{G}}\big) + j\omega C_{\mathrm{GG}}, \tag{3.7a}$$

$$Y_{12} \approx \omega^2\big(C_{\mathrm{DB}}C_{\mathrm{GB}}R_{\mathrm{B}} - C_{\mathrm{GD}}C_{\mathrm{GG}}R_{\mathrm{G}}\big) - j\omega C_{\mathrm{GD}}, \tag{3.7b}$$

$$Y_{21} \approx G_{\mathrm{m_{eff}}} + \omega^2\Big(C_{\mathrm{GB}}R_{\mathrm{B}}\big(C_{\mathrm{DB}} - C_{\mathrm{m}} + C_{\mathrm{ms}}\big) - C_{\mathrm{GG}}R_{\mathrm{G}}\big(C_{\mathrm{GD}} + C_{\mathrm{m}}\big)\Big) - j\omega\big(C_{\mathrm{GD}} + C_{\mathrm{m}}\big), \tag{3.7c}$$

$$Y_{22} \approx G_{\mathrm{ds}} + \omega^2\Big(C_{\mathrm{DB}}R_{\mathrm{B}}\big(C_{\mathrm{DB}} - C_{\mathrm{m}} + C_{\mathrm{ms}}\big) + C_{\mathrm{GD}}R_{\mathrm{G}}\big(C_{\mathrm{GD}} + C_{\mathrm{m}}\big)\Big) + j\omega\big(C_{\mathrm{DB}} + C_{\mathrm{GD}}\big), \tag{3.7d}$$

with

$$C_{\mathrm{GG}} = C_{\mathrm{GS}} + C_{\mathrm{GD}} + C_{\mathrm{GB}} \tag{3.8}$$

being the total gate capacitance. Note that the poles due to the S/D series resistances are located typically at much higher frequencies than the transit frequency, so R_{S} and R_{D} can be neglected in the calculations of the Y-parameters [42, 43]. Their contribution is accounted for by using the effective gate and source transconductances:

$$G_{\mathrm{m_{eff}}} = G_{\mathrm{m}}/D, \tag{3.9a}$$

$$G_{\mathrm{ms,eff}} = (G_{\mathrm{ms}} + G_{\mathrm{ds}})/D, \tag{3.9b}$$

where D in saturation is given by [39]:

$$D \approx 1 + G_{\mathrm{ms}}R_{\mathrm{s}}. \tag{3.10}$$

The values of the effective gate and source transconductances are directly extracted from the Y-parameters, without the need to individually calculate R_{S} and R_{D}. Nevertheless, if needed, different methods to extract the S/D series resistances either during a DC or an RF analysis exist [76–79].

We notice that G_{ms} and C_{SB} do not appear in the simplified expressions of the Y-parameters given by Eq. (3.7), which implies that they do not play an important role in a first-order analytical model of the Y-parameters of such a common-source (CS) circuit configuration.

This is mainly due to the small voltage drop $V_{si} - V_{bi}$, which is because S is shorted to ground and R_S/R_D are not accounted for when calculating the Y-parameters.

3.4 Expressions for the Direct Extraction of the RF Components

In order to model accurately the device characteristics, the RF components of the small-signal equivalent circuit should be extracted from measurements. The simplified equations for the Y-parameters defined in Eq. (3.7) form a system of ten equations ($\Re\{Y_{11}\}$, $\Im\{Y_{11}\}$, $\Re\{Y_{12}\}$, $\Im\{Y_{12}\}$, $\Re\{Y_{21}\}$, $\Re\{Y_{21}\}|_{\omega=0}$, $\Im\{Y_{21}\}$, $\Re\{Y_{22}\}$, $\Re\{Y_{22}\}|_{\omega=0}$, $\Im\{Y_{22}\}$) with ten unknowns and thus can be used to derive the expressions for the direct extraction of the RF components' values. As demonstrated below the extraction is performed in two phases:

Phase 1:

$$G_{m_{eff}} = \Re\{Y_{21}\}|_{\omega=0} \tag{3.11a}$$

$$G_{ds} = \Re\{Y_{22}\}|_{\omega=0} \tag{3.11b}$$

$$C_{GG} = \Im\{Y_{11}\}/\omega \tag{3.11c}$$

$$C_{GD} = -\Im\{Y_{12}\}/\omega \tag{3.11d}$$

$$C_m = (\Im\{Y_{12}\} - \Im\{Y_{21}\})/\omega \tag{3.11e}$$

$$C_{DB} = (\Im\{Y_{12}\} + \Im\{Y_{22}\})/\omega \tag{3.11f}$$

Phase 2:

$$R_G = \frac{b\cdot\Re\{Y_{11}\} - a\cdot\Re\{Y_{12}\}}{\omega^2\Big(b\cdot C_{GG}{}^2 - c\cdot C_{GD}\Re\{Y_{11}\} - C_{GG}\big(C_m\Re\{Y_{12}\} - d\cdot C_{GD}\big)\Big)} \tag{3.12a}$$

$$R_B = \frac{\big(c\cdot\Re\{Y_{12}\} - b\cdot C_{GG}\big)^2\big(C_{GD}\Re\{Y_{11}\} + C_{GG}\Re\{Y_{12}\}\big)}{\omega^2 C_{DB}{}^2\Big(a\cdot C_{GG} - c\cdot\Re\{Y_{11}\}\Big)\Big(b\cdot C_{GG}{}^2 - c\cdot C_{GD}\Re\{Y_{11}\} - C_{GG}\big(C_m\Re\{Y_{12}\} - d\cdot C_{GD}\big)\Big)} \tag{3.12b}$$

$$C_{GB} = \frac{C_{DB}\big(c\cdot\Re\{Y_{11}\} - a\cdot C_{GG}\big)}{c\cdot\Re\{Y_{12}\} - b\cdot C_{GG}} \tag{3.12c}$$

$$C_{ms} = \frac{\begin{array}{c} C_{DB}\Big(b\cdot e\cdot C_{GG}{}^2 - c\cdot e\cdot C_{GD}\Re\{Y_{11}\} + C_{GG}\Big(d\cdot e\cdot C_{GD} - C_m\big(G_{ds}\Re\{Y_{11}\} + \Re\{Y_{12}\}^2 \\ -\Re\{Y_{11}\}\Re\{Y_{22}\}\big)\Big)\Big) + C_m\big(c\cdot\Re\{Y_{12}\} - b\cdot C_{GG}\big)\big(C_{GD}\Re\{Y_{11}\} + C_{GG}\Re\{Y_{12}\}\big) \end{array}}{\big(c\cdot\Re\{Y_{22}\} - b\cdot C_{GG}\big)\big(C_{GD}\Re\{Y_{11}\} + C_{GG}\Re\{Y_{12}\}\big)} \tag{3.12d}$$

with

$$a = G_{m_{eff}} - \Re\{Y_{21}\},$$

$$b = G_{\mathrm{ds}} - \Re\{Y_{22}\},$$
$$c = C_{\mathrm{GD}} + C_{\mathrm{m}},$$
$$d = G_{\mathrm{m_{eff}}} - \Re\{Y_{21}\} - \Re\{Y_{12}\},$$
$$e = G_{\mathrm{m_{eff}}} - \Re\{Y_{21}\} + \Re\{Y_{12}\}.$$

Each phase consists of a group of calculations. The order of the calculations in each phase is of no importance, since they do not depend on each other, but, Phase 1 should precede Phase 2. Some of the expressions, e.g., R_{B}, C_{ms}, despite being long, consist of simple mathematical calculations. The RF model parameters should not depend on frequency, but if they display some variations across frequency due to low quality measurements, their average value can be used. On the other hand, the RF parameters are bias and geometry dependent, so the parameter extraction procedure should be performed for each operating point and device geometry. Extracting the values of the RF components from measurements can serve as a valuable tool in understanding and modeling the different dependencies, e.g., on bias or geometry. These dependencies can then be included in compact models, for which a global fit (across the whole bias and geometry range), with a single set of parameters, is desirable.

3.5 Parameter Extraction from Measurements

For the validation of the analytical model, the de-embedded S-parameters up to 50 GHz of two commercial state-of-the-art CMOS processes, were used. The minimum length RF multifinger nanoscale nMOS devices of each process having $M = 6$ (devices in parallel), $N_{\mathrm{f}} = 10$ (number of fingers per device), $W = 2$ µm ($W_{\mathrm{tot}} = M \cdot N_{\mathrm{f}} \cdot W = 120$ µm) and $L = 40$ nm and 30 nm, were measured. The layout of a single cell ($M = 1$) of the 40 nm device is shown in Fig. 3.2. The measurements were carried out using a standard small-signal RF measurement set-up. For the de-embedding of the S-parameters a procedure based on 'OPEN' and 'SHORT' dummy test structures [58], has been applied. The de-embedded S-parameters measurements were converted to Y-parameters, which can be directly compared to the analytical expressions provided by Eq. (3.7).

Following the procedure described in Eq. (3.11) and Eq. (3.12), the RF components of the RF DUTs were extracted in saturation ($V_{\mathrm{D}} = 1.1$ V for the 40 nm device and $V_{\mathrm{D}} = 1.0$ V for the 30 nm device) and from weak- to strong-inversion. The normalized values of the extracted RF components vs. the inversion coefficient (IC) are shown in Fig. 3.3 and Fig. 3.4. IC is calculated using:

$$IC = \frac{I_{\mathrm{D_{sat}}}}{I_{\mathrm{spec}}}, \tag{3.14}$$

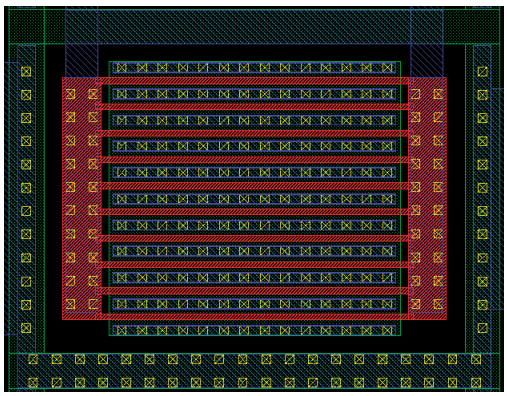


Figure 3.2 – Layout of a single cell ($N_f = 10$, $W = 2$ µm and $L = 40$ nm) of the measured RF DUT.

with I_{spec} being the specific current given by:

$$I_{spec} = 2n\beta U_T^2, \tag{3.15}$$

where n is the slope factor, $\beta = \mu C_{ox} W_{tot}/L$ the transfer parameter, $U_T = kT/q$ the thermodynamic voltage, μ the mobility of the carriers and C_{ox} the oxide capacitance per unit area. IC is especially useful for design optimization since it is a metric for the inversion level of a transistor, with $IC < 0.1$ defining weak-inversion (WI), $IC > 10$ strong-inversion (SI) and $0.1 < IC < 10$ moderate-inversion (MI) region [39]. The capacitances are normalized to:

$$C_{ox} \cdot W_{tot} \cdot L, \tag{3.16}$$

and the resistances to:

$$R_{spec} = \frac{1}{G_{spec}}, \tag{3.17}$$

with $G_{spec} = I_{spec}/U_T$ being the specific conductance. For example, the normalized total gate capacitance c_{gg} is found through:

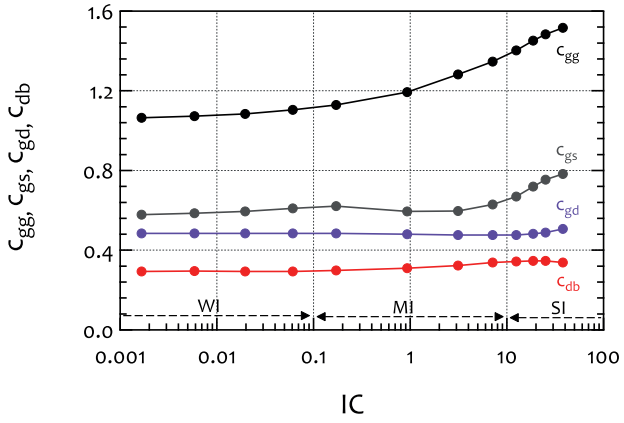
$$c_{gg} = \frac{C_{GG}}{C_{ox} \cdot W_{tot} \cdot L}, \tag{3.18}$$

the normalized gate resistance r_g through:

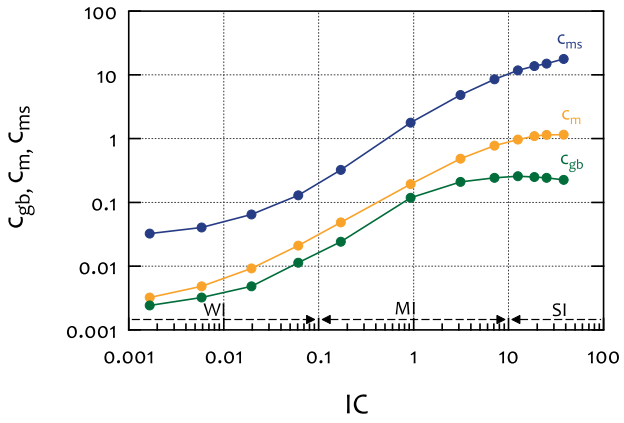
$$r_g = \frac{R_G}{R_{spec}}, \tag{3.19}$$

and so on. Note that, the normalization factor I_{spec} is extracted from DC measurements as explained in details in Fig. 9.4, while different methods have been proposed for the extraction of C_{ox} from CV measurements [80–82].

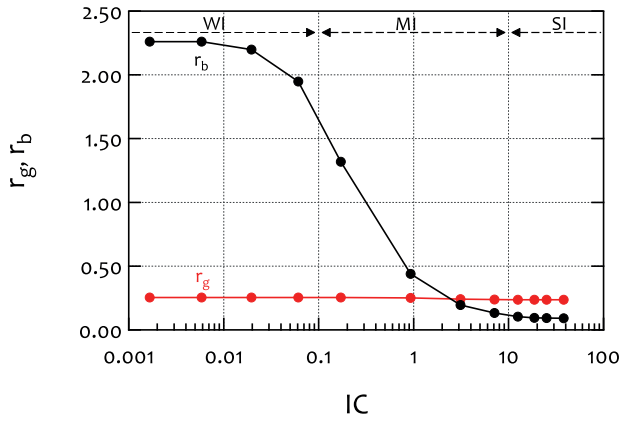
From Fig. 3.3 and Fig. 3.4, we observe that C_{GD}, C_{DB} and R_G present a weak V_G bias dependence



(a) c_{gg}, c_{gd}, c_{db}

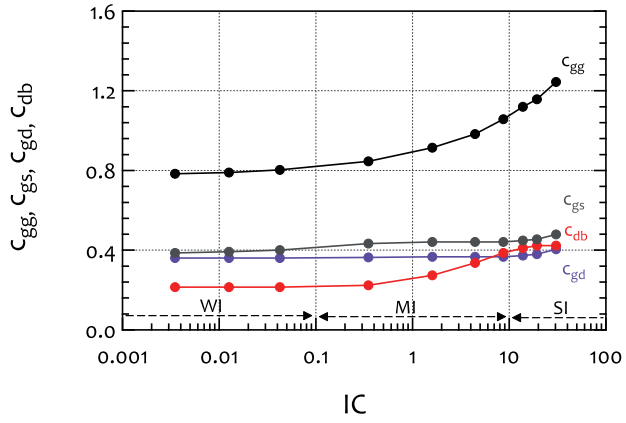


(b) c_{gb}, c_m, c_{ms}

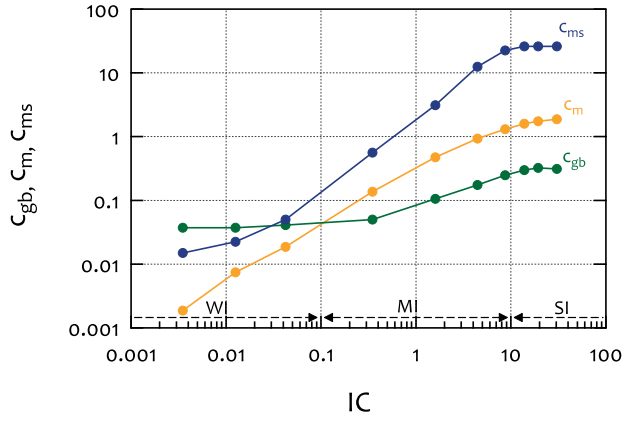


(c) r_g, r_b

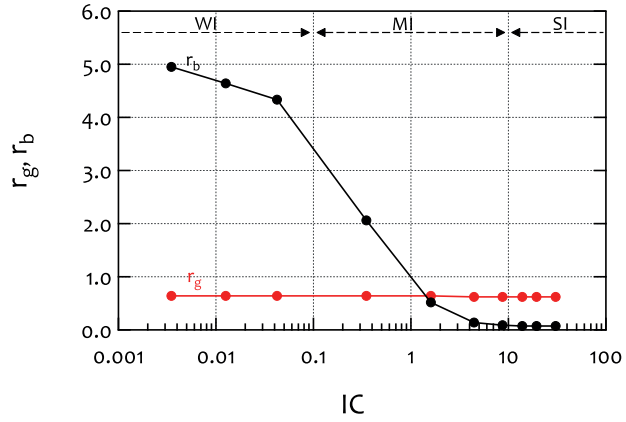
Figure 3.3 – Normalized extracted values of the RF components of the small-signal equivalent circuit vs. *IC* of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_D = 1.1$ V: (a) c_{gg}, c_{gs}, c_{gd}, c_{db}, (b) c_{gb}, c_m, c_{ms} and (b) r_g, r_b. The capacitances are normalized to $C_{ox} \cdot W_{tot} \cdot L = 62\ fF$ and the resistances to $R_{spec} = 15.92\ \Omega$.



(a) c_{gg}, c_{gd}, c_{db}



(b) c_{gb}, c_m, c_{ms}



(c) r_g, r_b

Figure 3.4 – Normalized extracted values of the RF components of the small-signal equivalent circuit vs. *IC* of a minimum channel length RF nMOS DUT of a 28 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 30$ nm, at $V_D = 1.0$ V: (a) c_{gg}, c_{gs}, c_{gd}, c_{db}, (b) c_{gb}, c_m, c_{ms} and (b) r_g, r_b. The capacitances are normalized to $C_{ox} \cdot W_{tot} \cdot L = 80\ fF$ and the resistances to $R_{spec} = 9.7\ \Omega$.

(regarding R_G similar results were obtained in [83]), whereas all the other RF model parameters show a much stronger one. Regarding R_B, we observe a sharp transition in moderate-inversion region, while in weak- and strong-inversion it remains relatively constant. This steep transition of R_B was also demonstrated in [71].

3.5.1 Validation of the Analytical RF Model

Using the extracted values of the RF components, we can compare the analytical expressions of the Y-parameters vs. frequency, against measurements, for different levels of inversion. The Y-parameters in Fig. 3.5, Fig. 3.6, Fig. 3.7 and Fig. 3.8 are normalized using the specific conductance G_{spec} following the expression:

$$y_{ij} = Y_{ij}/G_{spec}. \tag{3.20}$$

Even with the use of a single resistance for the substrate, the simple analytical model is able to predict accurately the RF behavior of both 40 nm and 28 nm devices, except for a discrepancy at the lower RF frequencies of $\Re\{Y_{22}\}$. R_B may have a strong impact on $\Re\{Y_{22}\}$ even at low frequencies, however the discrepancy is due to the isolation layer that expands below the RF Device-Under-Test (DUT) and is not accounted for in our calculations as this is a specific characteristic of these particular technologies and might not apply for all CMOS processes. For a perfect fit with measurements of this technology, it is important to take into account the impact of the isolation layer as already described in [70, 72]. Nevertheless, even without considering the impact of the isolation layer the discrepancy in $\Re\{Y_{22}\}$ is $< 15\%$ in weak-inversion and $< 5\%$ in strong-inversion.

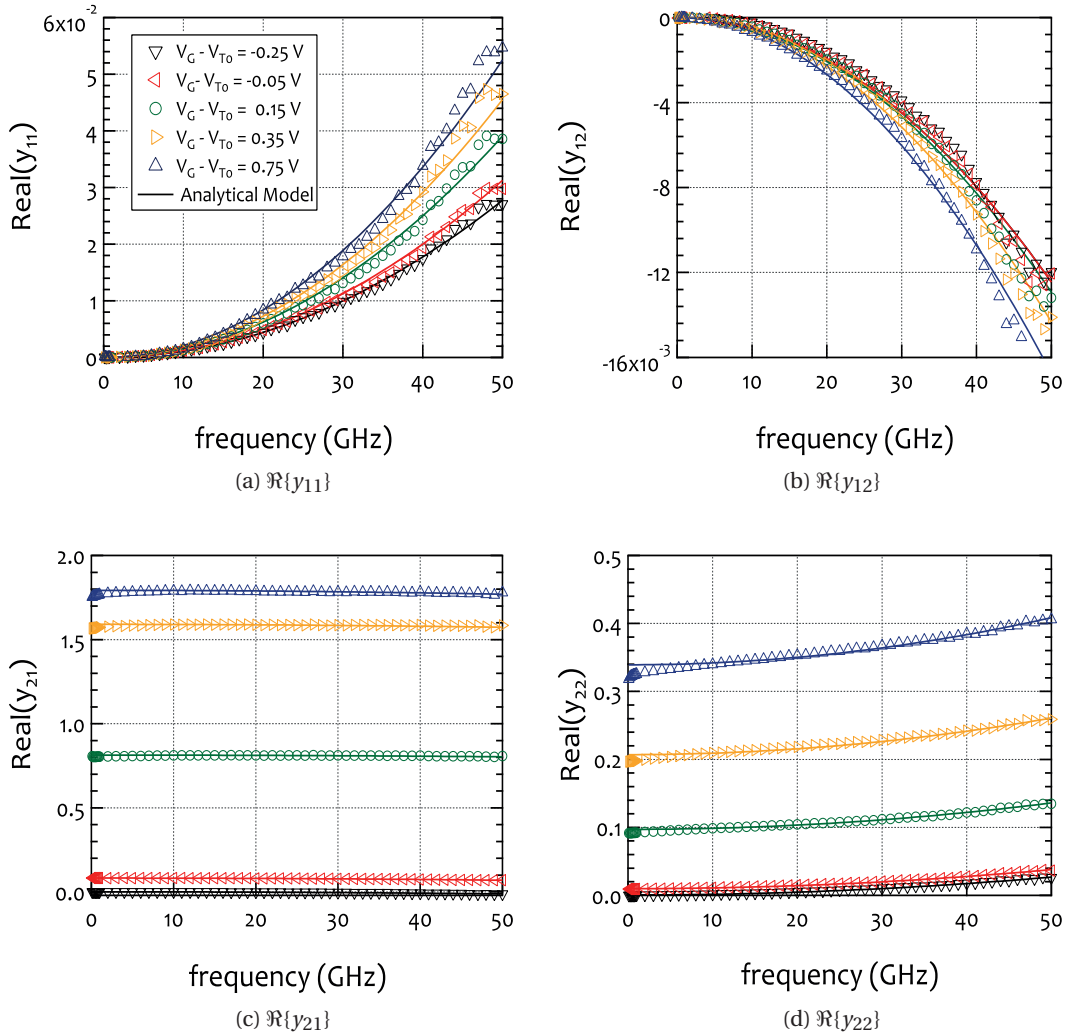
3.6 RF Figures-of-Merit

Using the simplified analytical expressions for the Y-parameters Eq. (3.7), we can model analytically different RF FoMs useful from a designer's point of view.

3.6.1 Current Gain - H_{21}

The current gain is defined as:

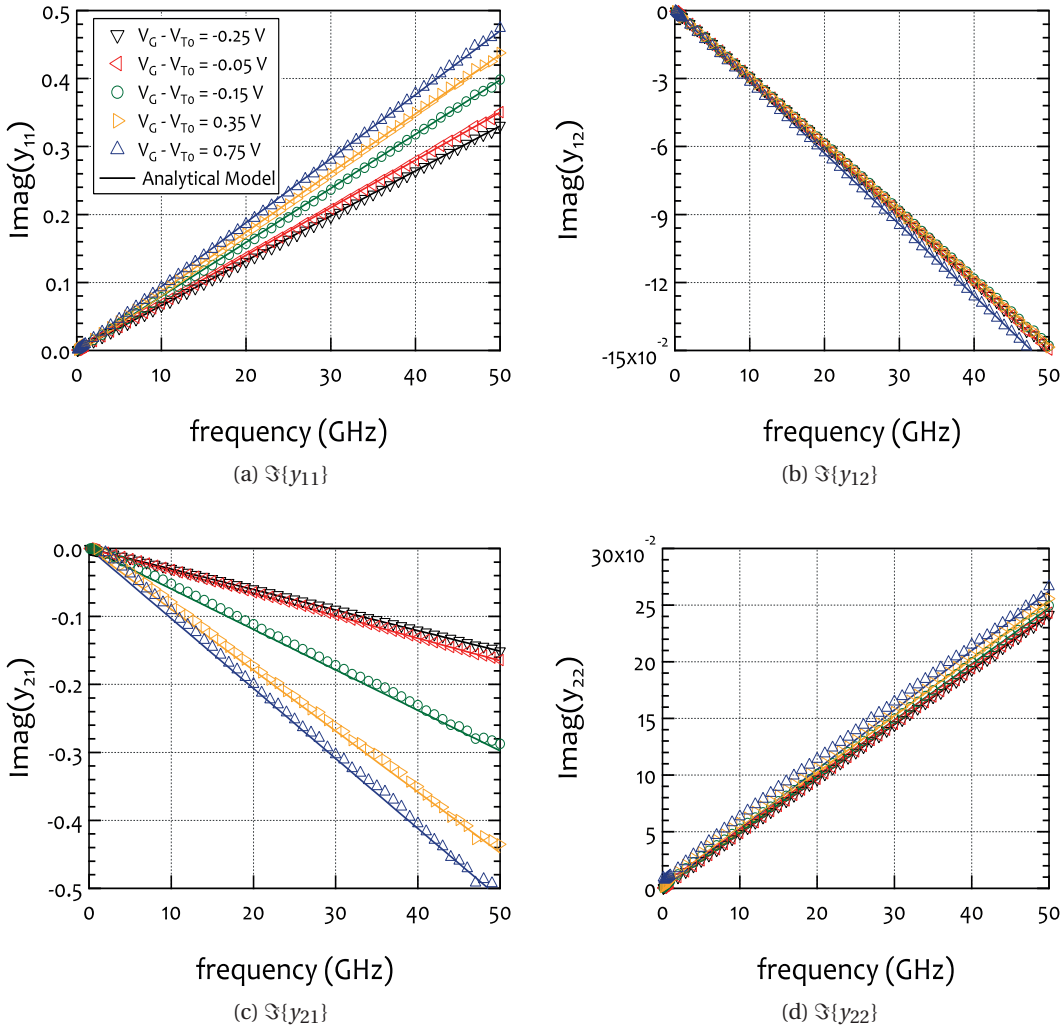
$$H_{21} = \left.\frac{I_2}{I_1}\right|_{V_2=0} = \frac{Y_{21}}{Y_{11}}. \tag{3.21}$$



(a) $\Re\{y_{11}\}$ (b) $\Re\{y_{12}\}$

(c) $\Re\{y_{21}\}$ (d) $\Re\{y_{22}\}$

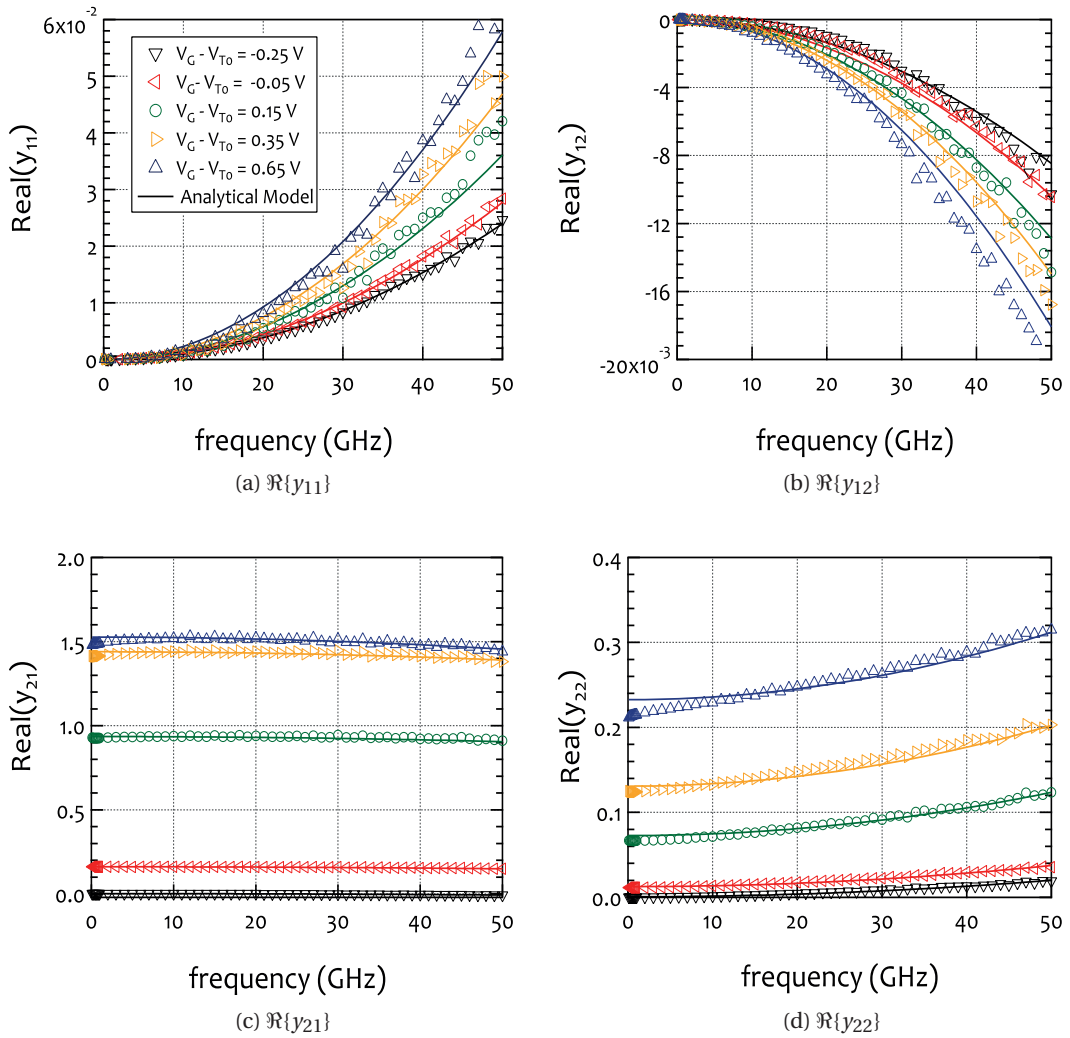
Figure 3.5 – Normalized real part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_G - V_{T0} = [-0.25, -0.05, 0.15, 0.35, 0.75]$ V and $V_D = 1.1$ V: (a) Real(y_{11}), (b) Real(y_{12}), (c) Real(y_{21}) and (d) Real(y_{22}). Note that at $V_G - V_{T0} = -0.25$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = [0.35, 0.75]$ V the device is in strong-inversion ($IC > 10$). The Y-parameters are normalized according to the formula: $y_{ij} = Y_{ij}/G_{spec}$, with $G_{spec} = 62.8\ mS$.



(a) $\Im\{y_{11}\}$ (b) $\Im\{y_{12}\}$

(c) $\Im\{y_{21}\}$ (d) $\Im\{y_{22}\}$

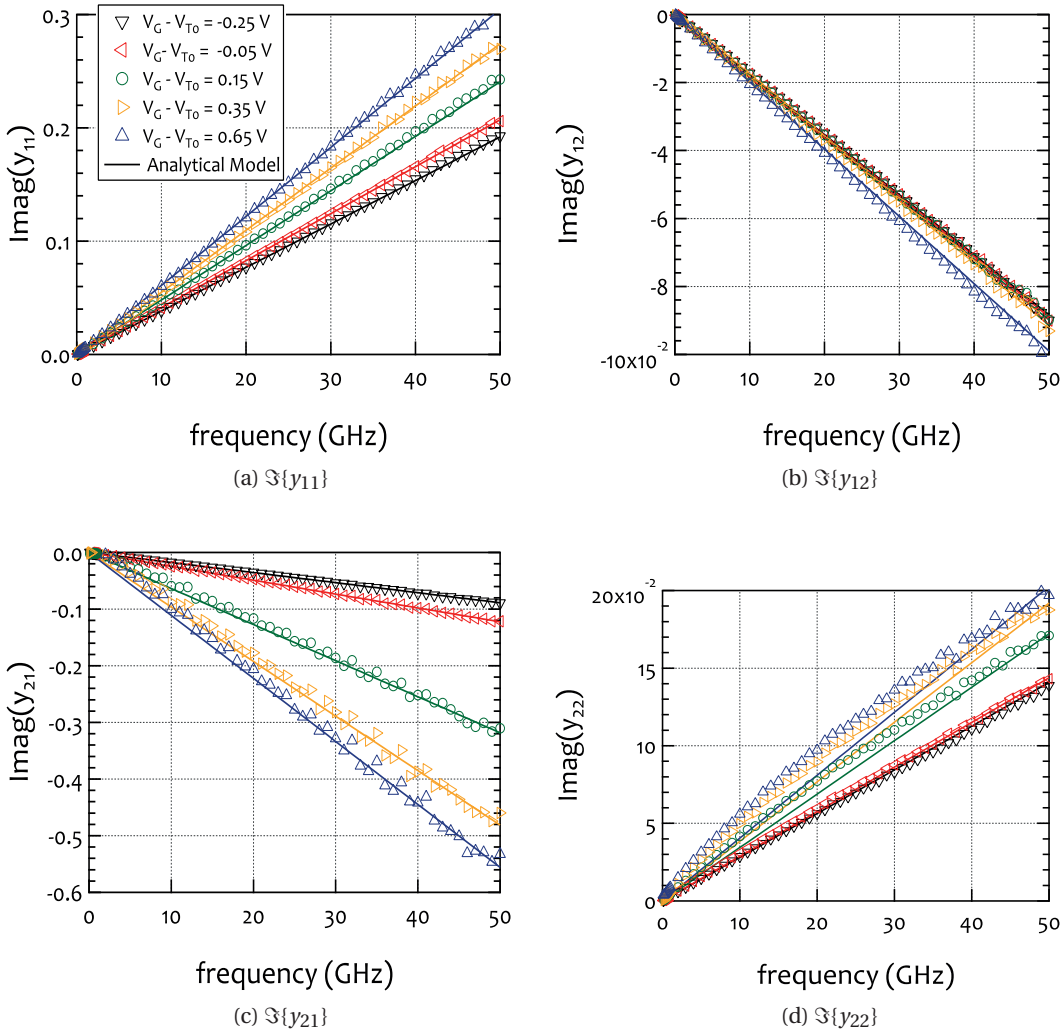
Figure 3.6 – Normalized imaginary part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_G - V_{T0} = [-0.25, -0.05, 0.15, 0.35, 0.75]$ V and $V_D = 1.1$ V: (a) Imag(y_{11}), (b) Imag(y_{12}), (c) Imag(y_{21}) and (d) Imag(y_{22}). Note that at $V_G - V_{T0} = -0.25$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = [0.35, 0.75]$ V the device is in strong-inversion ($IC > 10$). The Y-parameters are normalized according to the formula: $y_{ij} = Y_{ij}/G_{spec}$, with $G_{spec} = 62.8\ mS$.



(a) $\Re\{y_{11}\}$ (b) $\Re\{y_{12}\}$

(c) $\Re\{y_{21}\}$ (d) $\Re\{y_{22}\}$

Figure 3.7 – Normalized real part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 28 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 30$ nm, at $V_G - V_{T0} = [-0.25, -0.05, 0.15, 0.35, 0.65]$ V and $V_D = 1.0$ V: (a) Real(y_{11}), (b) Real(y_{12}), (c) Real(y_{21}) and (d) Real(y_{22}). Note that at $V_G - V_{T0} = -0.25$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = [0.35, 0.65]$ V the device is in strong-inversion ($IC > 10$). The Y-parameters are normalized according to the formula: $y_{ij} = Y_{ij}/G_{spec}$, with $G_{spec} = 103.1\ mS$.



(a) $\Im\{y_{11}\}$ (b) $\Im\{y_{12}\}$

(c) $\Im\{y_{21}\}$ (d) $\Im\{y_{22}\}$

Figure 3.8 – Normalized imaginary part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 28 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 30$ nm, at $V_G - V_{T0} = [-0.25, -0.05, 0.15, 0.35, 0.65]$ V and $V_D = 1.0$ V: (a) Imag(y_{11}), (b) Imag(y_{12}), (c) Imag(y_{21}) and (d) Imag(y_{22}). Note that at $V_G - V_{T0} = -0.25$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = [0.35, 0.65]$ V the device is in strong-inversion ($IC > 10$). The Y-parameters are normalized according to the formula: $y_{ij} = Y_{ij}/G_{spec}$, with $G_{spec} = 103.1\ mS$.

Using Eq. (3.7a) and Eq. (3.7c) in Eq. (3.21), we can calculate the complete expression for the current gain:

$$H_{21} = \frac{G_{m_{eff}} + \omega^2\Big(C_{GB}R_B\big(C_{DB} - C_m + C_{ms}\big) - C_{GG}R_G\big(C_{GD} + C_m\big)\Big) - j\omega\big(C_{GD} + C_m\big)}{\omega^2\big({C_{GB}}^2 R_B + {C_{GG}}^2 R_G\big) + j\omega C_{GG}}. \quad (3.22)$$

However, the above expression is rather complex, displaying two zeros at:

$$\omega_{z1} = -\frac{C_{GD} + C_m - \sqrt{\begin{array}{c}{C_{GD}}^2 + {C_m}^2 + 4\cdot G_{m_{eff}}\Big(C_{GB}R_B\big(C_{DB} + C_{ms}\big) \\ -C_m\big(C_{GB}R_B + C_{GG}R_G\big)\Big) + 2\cdot C_{GD}\big(C_m - 2\cdot C_{GG}G_{m_{eff}}R_G\big)\end{array}}}{2\cdot\Big(C_{GB}R_B\big(C_{DB} - C_m + C_{ms}\big) - C_{GG}R_G\big(C_{GD} + C_m\big)\Big)}, \quad (3.23a)$$

$$\omega_{z1} = \frac{2\cdot G_{m_{eff}}}{C_{GD} + C_m - \sqrt{\begin{array}{c}\big(C_{GD} + C_m\big)^2 + 4\cdot G_{m_{eff}}\Big(C_{GB}R_B\big(C_{DB} - C_m + C_{ms}\big) \\ -C_{GG}R_G\big(C_{GD} + C_m\big)\Big)\end{array}}}, \quad (3.23b)$$

and two poles at:

$$\omega_{p1} = 0, \quad (3.24a)$$

$$\omega_{p2} = \frac{C_{GG}}{\big({C_{GB}}^2 R_B + {C_{GG}}^2 R_G\big)}, \quad (3.24b)$$

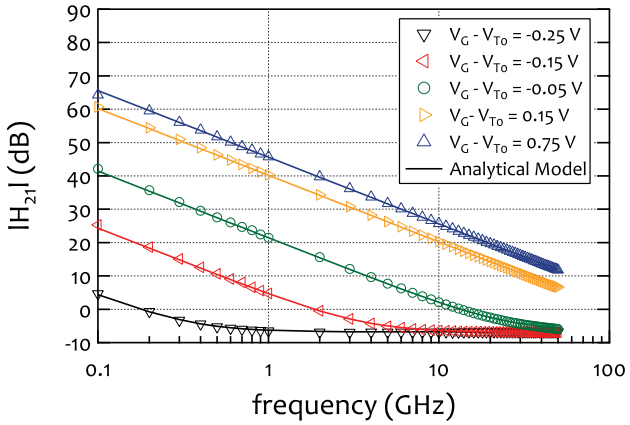
whereas the target is to have the simplest expressions possible. In order to achieve that, we can ignore the 2^{nd}-order terms in Eq. (3.22) ending up with:

$$H_{21} \approx \frac{G_{m_{eff}} - j\omega\big(C_{GD} + C_m\big)}{j\omega C_{GG}}. \quad (3.25)$$

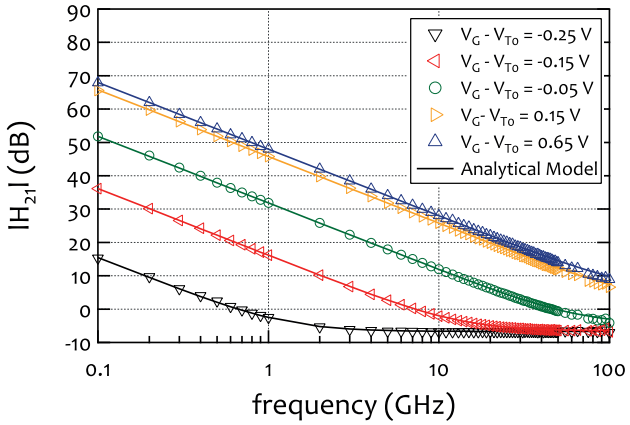
The above simplified expression is valid for $\omega \ll \omega_{p2}$, considering that ω_{p2} is located at lower frequencies than the two zeros.

In Fig. 3.9 the magnitude of the current gain $\big|H_{21}\big|$ vs. frequency is shown. The simple analytical model Eq. (3.25) is compared against measurements. We see that the model is able to capture $\big|H_{21}\big|$ accurately. The model is valid from weak- to strong-inversion, even for $f \gg G_{m_{eff}}/\Big(2\pi\cdot\big(C_{GD} + C_m\big)\Big)$, for which at lower inversion levels, e.g., at $V_G - V_{T0} = [-0.25, -0.15, -0.05]$ V, $\big|H_{21}\big|$ becomes independent of the frequency and simplifies to just a ratio of capacitances equal to:

$$\big|H_{21}\big| = \frac{\big(C_{GD} + C_m\big)}{C_{GG}}. \quad (3.26)$$



(a) Minimum Length RF DUT of a 40 nm CMOS process



(b) Minimum Length RF DUT of 28 nm a CMOS process

Figure 3.9 – $|H_{21}|$ vs. frequency of minimum length RF nMOS DUTs of 40 nm and 28 nm CMOS processes, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = [40, 30]$ nm. The bias conditions for the 40 nm DUT are $V_G - V_{T0} = [-0.25...0.75]$ V and $V_D = 1.1$ V, whereas for the 30 nm DUT are $V_G - V_{T0} = [-0.25...0.65]$ V and $V_D = 1.0$ V. Note that at $V_G - V_{T0} = [-0.25, -0.15]$ V the devices are in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the devices are in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = 0.75$ or 0.65 V the devices are in strong-inversion ($IC > 10$).

3.6.2 Transit Frequency - F_t

The transit frequency is defined as the frequency at which the magnitude of the current gain becomes equal to unity (or 0 dB). Using Eq. (3.25) and solving for $|H_{21}| = 1$ we can calculate F_t [1] as:

$$F_t = \frac{G_{m_{eff}}}{2\pi\sqrt{{C_{GG}}^2 - \left(C_{GD} + C_m\right)^2}}. \tag{3.27}$$

[1] The transit frequency and the maximum oscillation frequency are mostly denoted by f_t and f_{max}, respectively. However for consistency in the current thesis, in order to distinguish them from their normalized form, F_t and F_{max} will be used for the non-normalized quantities.

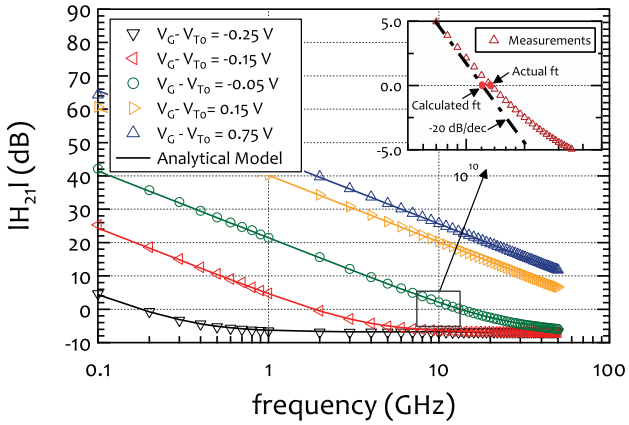
The above expression is valid up to frequencies where $f \ll \omega_{p2}/(2\pi) = C_{GG}/\left(2\pi \cdot \left({C_{GB}}^2 R_B + {C_{GG}}^2 R_G\right)\right)$. However, the transit frequency at high levels of inversion is located above this limit so Eq. (3.27) cannot be used across all the different inversion levels. In order to provide a consistent approach that is valid from weak to strong-inversion, for the calculation of F_t we further simplify Eq. (3.27) using the assumption that ${G_{m_{eff}}}^2 \gg \omega^2 \left(C_{GD} + C_m\right)^2$, which results in the well-know expression [39]:

$$F_t = \frac{G_{m_{eff}}}{2\pi C_{GG}}. \tag{3.28}$$

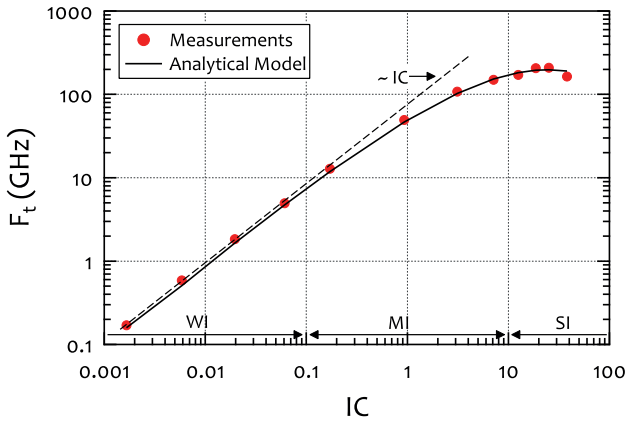
In Fig. 3.10b and Fig. 3.11b, the transit frequency F_t vs. IC is displayed. The analytical expression Eq. (3.28) is compared against measurements. To calculate F_t from measurements, the extrapolated value of $|H_{21}|$ at 0 dB, assuming a -20 dB/dec slope, is calculated, using $F_t = f_{spot} \cdot |H_{21}(f_{spot})|$, where f_{spot} is the frequency at which F_t is calculated. However, for the correct estimation of F_t, especially at low levels of inversion, f_{spot} must be carefully chosen before the point at which $|H_{21}|$ levels off.

From Eq. (3.28) we see that $F_t \propto G_{m_{eff}}$, so at low levels of inversion where $G_{m_{eff}} \propto IC$ [39], $F_t \propto IC$, whereas at higher levels of inversion F_t starts to saturate and even slightly decreases, due to the effect of velocity saturation and the increase of C_{GG} in strong-inversion, as shown in Fig. 3.10b and Fig. 3.11b. Combining also the facts that $C_{GG} \propto L$ and $IC \propto 1/L$, we find that $F_t \propto 1/L^2$ in weak-inversion, while $F_t \propto 1/L$ in strong-inversion [39]. This means that contrary to strong-inversion, in weak-inversion we can take full advantage of scaling, as there are no limitations posed by the short channel effects.

Note that, if the zeros and the poles of the complete expression for H_{21} are rather close to F_t, they affect $|H_{21}|$ in a frequency range close to F_t. Thus, $|H_{21}|$ does not display a -20 dB/dec slope close to the transit frequency and using Eq. (3.28) for the calculation of F_t might underestimate the transit frequency by $\sim 10\% - 20\%$. This change in slope is obvious in moderate-inversion levels as can be seen inside the inset of Fig. 3.10a and Fig. 3.11a, where a zoom in the region close to F_t of the $|H_{21}|$ curve, which corresponds to moderate-inversion for the 40 nm device and to weak-inversion for the 30 nm device, shows where we calculate F_t using Eq. (3.28) and where it is actually located according to measurements. In such a case, Eq. (3.27) would be accurate. Nonetheless, Eq. (3.28) provides a very simple approximation to estimate F_t. At this point it is also worth mentioning, that the actual F_t of the devices could be even higher if RF pulsed measurements were carried out so that the degradation of the RF characteristics resulting from the electro-thermal phenomena would be avoided [84].



(a) $|H_{21}|$



(b) F_t

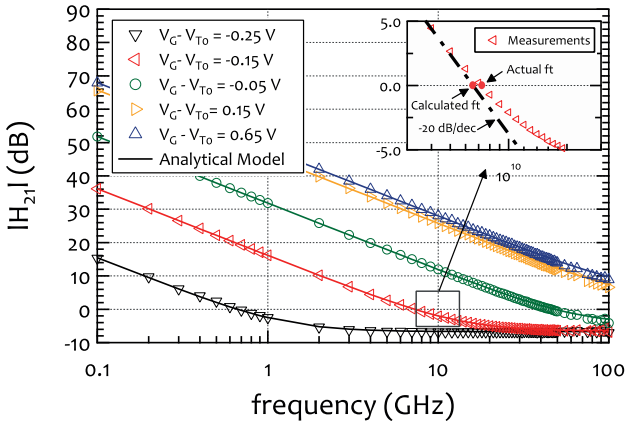
Figure 3.10 – (a) $|H_{21}|$ vs. frequency at $V_G - V_{T0} = [-0.25...0.75]$ V and (b) F_t vs. IC (at $f = 0.1$ GHz for measurements), of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_D = 1.1$ V. The inset inside (a) zooms in a region close to F_t of the $|H_{21}|$ curve in moderate-inversion, and shows a 10% – 20% difference between the calculated F_t and the actual one. Note that at $V_G - V_{T0} = [-0.25, -0.15]$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = 0.75$ V the device is in strong-inversion ($IC > 10$).

3.6.3 Unilateral Gain - *U*

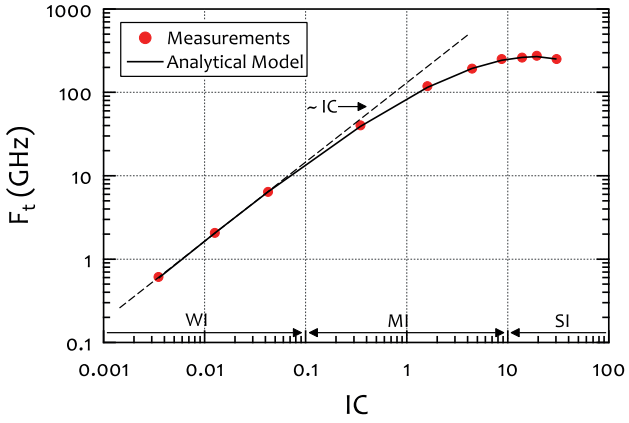
Mason's Unilateral Gain can be calculated as [85]:

$$U = \frac{\left|Y_{21} - Y_{12}\right|^2}{4 \cdot \left(\Re\{Y_{11}\}\Re\{Y_{22}\} - \Re\{Y_{12}\}\Re\{Y_{21}\}\right)}. \tag{3.29}$$

Using the simplified analytical expressions for the Y-parameters Eq. (3.7) in Eq. (3.29), we can



(a) $|H_{21}|$



(b) F_t

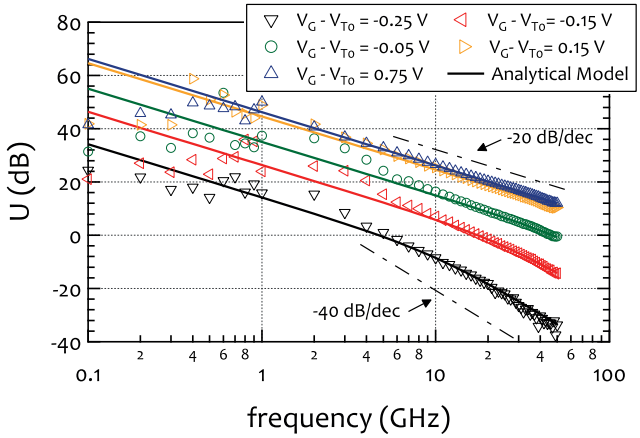
Figure 3.11 – (a) $|H_{21}|$ vs. frequency at $V_G - V_{T0} = [-0.25...0.65]$ V and (b) F_t vs. IC (at $f = 0.2$ GHz for measurements), of a minimum channel length RF nMOS DUT of a 28 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 30$ nm, at $V_D = 1.0$ V. The inset inside (a) zooms in a region close to F_t of the $|H_{21}|$ curve in weak-inversion, and shows a 10% – 20% difference between the calculated F_t and the actual one. Note that at $V_G - V_{T0} = [-0.25, -0.15]$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = 0.65$ V the device is in strong-inversion ($IC > 10$).

define U as:

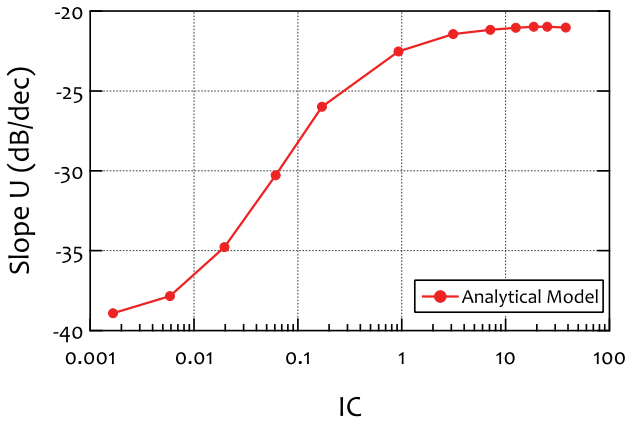
$$U = \frac{K}{\omega^2\left(1 + \left(\frac{\omega}{\omega_{p2}}\right)^2\right)}, \tag{3.30}$$

where K is given by:

$$K = \frac{{G_{m_{eff}}}^2}{4 \cdot \left(C_{GB}R_B\left(C_{GB}G_{ds} - C_{DB}G_{m_{eff}}\right) + C_{GG}R_G\left(C_{GD}G_{m_{eff}} + C_{GG}G_{ds}\right)\right)}, \tag{3.31}$$



(a) U



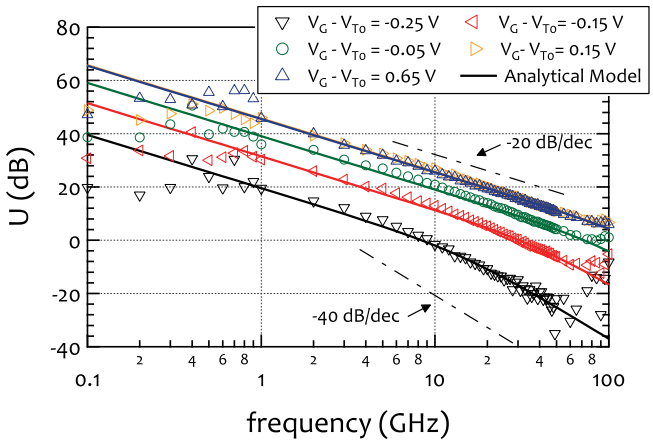
(b) Slope of U at high f

Figure 3.12 – (a) U vs. frequency at $V_G - V_{T0} = [-0.25...0.75]$ V and (b) the slope of U at high frequencies, of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_D = 1.1$ V. In (b) the slope of U that it is predicted by the analytical model is shown and clearly displays the transition from −40 dB/dec slope in weak-inversion to −20 dB/dec in strong-inversion. Note that at $V_G - V_{T0} = [-0.25, -0.15]$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = 0.75$ V the device is in strong-inversion ($IC > 10$).

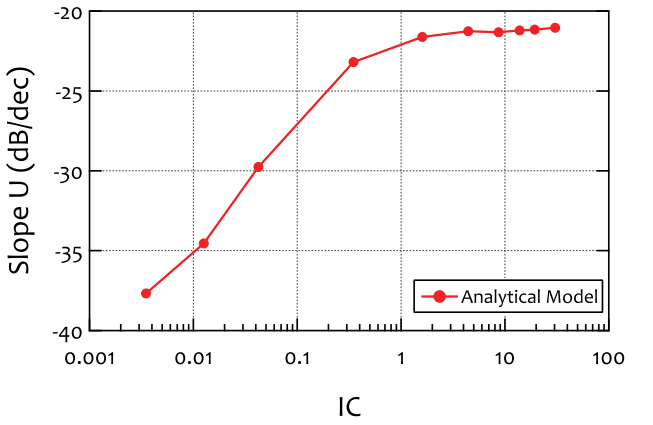
ω_{p2} is a real double pole that is calculated from:

$$\omega_{p2} = \sqrt{\frac{C_{GB}R_B\left(C_{GB}G_{ds} - C_{DB}G_{m_{eff}}\right) + C_{GG}R_G\left(C_{GD}G_{m_{eff}} + C_{GG}G_{ds}\right)}{R_B R_G\left(C_{DB}C_{GG} + C_{GB}C_{GD}\right)\left(C_{GG}\left(C_{DB} - C_m + C_{ms}\right) + C_{GB}\left(C_{GD} + C_m\right)\right)}}, \tag{3.32}$$

while the zeros are neglected as they are located at very high frequencies. The ω_{p2} pole cannot be neglected, since it has a prominent effect on U especially at low levels of inversion. This is obvious in Fig. 3.12a and Fig. 3.13a where the analytical model Eq. (3.30) is compared against



(a) U



(b) Slope of U at high f

Figure 3.13 – (a) U vs. frequency at $V_G - V_{T0} = [-0.25...0.65]$ V and (b) the slope of U at high frequencies, of a minimum channel length RF nMOS DUT of a 28 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 30$ nm, at $V_D = 1.0$ V. In (b) the slope of U that it is predicted by the analytical model is shown and clearly displays the transition from −40 dB/dec slope in weak-inversion to −20 dB/dec in strong-inversion. Note that at $V_G - V_{T0} = [-0.25, -0.15]$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = 0.65$ V the device is in strong-inversion ($IC > 10$).

measurement. We see that the slope of U with respect to frequency changes as we move towards lower levels of inversion, as already discussed in [70, 86]. In strong-inversion, e.g., at $V_G - V_{T0} = 0.75$ V or 0.65 V, U has a slope of ~ −20 dB/dec, while as the current density decreases the slope gradually becomes steeper, till it reaches a value of ~ −40 dB/dec in deep weak-inversion, e.g., $V_G - V_{T0} = -0.25$ V. In Fig. 3.12b and Fig. 3.13b the slope of U that is predicted by the analytical model at the higher RF frequencies, namely at $f \gg 10$ GHz, is shown. The analytical model is able to capture accurately the slope change in the unilateral gain, and this is achieved when ω_{p2} is accounted for.

3.6.4 Maximum Oscillation Frequency - $F_{\max}$

The maximum oscillation frequency is defined as the frequency at which the unilateral gain becomes equal to unity. Using Eq. (3.30) and solving for $U = 1$ results in:

$$F_{\max} = \frac{\sqrt{\omega_{\mathrm{p2}} \cdot \left(\sqrt{4 \cdot K + \omega_{\mathrm{p2}}^2} - \omega_{\mathrm{p2}}\right)}}{2\sqrt{2} \cdot \pi}. \tag{3.33}$$

In Fig. 3.14 the maximum oscillation frequency $F_{\max}$ vs. IC is displayed. The analytical model Eq. (3.33) is compared against measurements. The observation made above concerning the slope change of U implies that $F_{\max}$ can no longer be calculated using the traditional method for which:

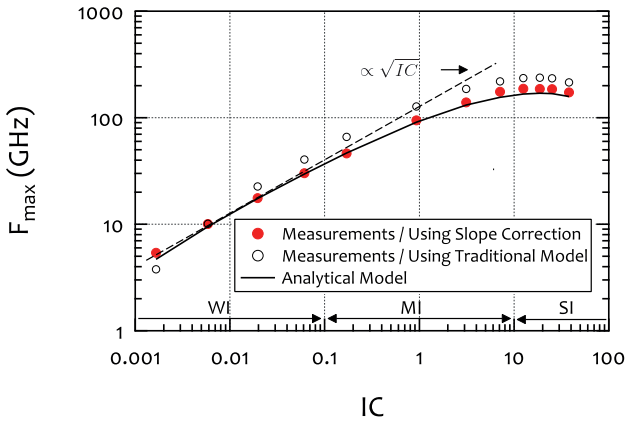
$$F_{\max} = f_{\mathrm{spot}} \cdot \sqrt{U(f_{\mathrm{spot}})}, \tag{3.34}$$

with f_{spot} being the frequency at which $F_{\max}$ is calculated [85]. Indeed, the previous approximation for $F_{\max}$ assumes a constant slope of -20 dB/dec and results in an overestimation of the value of the maximum oscillation frequency especially at lower inversion levels. Instead, $F_{\max}$ must be calculated as the frequency at which $U = 1$ for the operating points for which this is possible or as the frequency at which U, extrapolated with the correct slope, becomes equal to one. This method was used to obtain $F_{\max}$ from measurements and simulations in Fig. 3.14. We see that the analytical model is able to correctly predict $F_{\max}$, even at low current densities where U displays a steeper slope at higher RF frequencies. Since $F_{\max} \propto \sqrt{G_{\mathrm{m_{eff}}}}$, for the same reasons as explained for F_{t}, at low inversion levels $F_{\max} \propto \sqrt{IC}$, while it saturates and even slightly decreases at higher levels of inversion (due to the effect of velocity saturation and the increase of C_{GG} in strong-inversion). In Fig. 3.14 the $F_{\max}$, using the traditional method of calculation $\left(F_{\max} = f_{\mathrm{spot}} \cdot \sqrt{U(f_{\mathrm{spot}})}\right)$, is also shown. It is obvious in that case that the maximum oscillation frequency is overestimated compared to $F_{\max}$ calculated by all the other methods that account for the slope change in U.

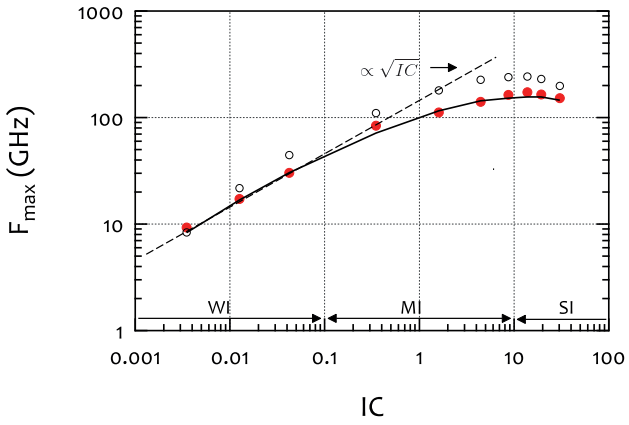
3.7 Conclusions

In order to achieve low-power consumption for GHz wireless applications, the parts of the radio operating at RF, can employ nanoscale technologies, for which the very high transit frequency can be traded-off with lower power, by shifting the operating point towards moderate- and even weak-inversion region. During the last years various metrics, e.g., [5, 54, 86–88], confirm how advantageous moderate inversion can be, as it offers a well-balanced trade-off in terms of gain, power consumption, noise and linearity.

This chapter was dedicated to the analytical modeling of the small-signal RF behavior of nanoscale MOSFETs, i.e., the Y-parameters and different FoMs. It was demonstrated that



(a) Minimum Length RF DUT of a 40 nm CMOS process



(b) Minimum Length RF DUT of 28 nm a CMOS process

Figure 3.14 – Fmax vs. IC (at $f = 10$ GHz for measurements using the traditional method) of minimum length RF nMOS DUTs of 40 nm and 28 nm CMOS processes, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = [40,30]$ nm. The bias conditions for the 40 nm DUT are $V_G - V_{T0} = [-0.25...0.75]$ V and $V_D = 1.1$ V, whereas for the 30 nm DUT are $V_G - V_{T0} = [-0.25...0.65]$ V and $V_D = 1.0$ V. Note that at $V_G - V_{T0} = [-0.25, -0.15]$ V the devices are in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the devices are in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = 0.75$ or 0.65 V the devices are in strong-inversion ($IC > 10$). It should be highlighted that F_{max} is layout dependent and thus, higher values can be obtained for a specific process through layout optimization (e.g., W reduction).

a generic, and simple RF small-signal equivalent circuit of the transistor can still provide accurate expressions even for very short channel devices. Special attention was given so that the expressions were valid even at low levels of inversion. A methodology for the direct extraction of the components of the RF small-signal equivalent circuit from measurements, was also presented. The V_G dependence of all the components was also shown, which can be useful for implementation in compact models, for which a global fit is desirable. The advantage of the proposed parameter extraction methodology is that it remains consistent with existing standard compact MOSFET models.

Furthermore, the study of the device characteristics from weak- to strong-inversion demonstrated that lower levels of inversion display different trends than strong-inversion region, and thus they must be modeled carefully. For example, the small-signal current gain H_{21} and the Mason unilateral gain U clearly illustrate the strong impact of the extrinsic part of the transistor in weak inversion region, which governs the overall behavior of the device, e.g., H_{21} levels-off and the slope of U becomes steeper above a specific frequency. For the first time, analytical expressions able to capture different FoMs from deep weak-inversion to strong-inversion, were presented.

4 Analytical RF Noise Modeling

4.1 Introduction

During the last decade, RF CMOS integrated circuits are strongly present in the commercial world [61]. The development of RF applications is strongly related to the continuous downscaling of MOSFET, thanks to the impressive RF performance that advanced nanoscale CMOS processes can provide. Nanoscale transistors constitute now a viable option for RF applications and RF Systems-on-Chip (SoCs). For realistic RF designs however, a model should be able to predict accurately the RF noise characteristics of the transistor, especially at low levels of inversion where the generated noise becomes significantly larger. In the GHz range, the thermal noise, generated in the channel and in the parasitic resistances, is the dominant noise source.

The thermal noise in the channel results from the random thermal motion of the current carriers. At high frequencies, the potential fluctuations within the channel are coupled with the gate terminal through the gate–oxide capacitance resulting in a noise current i_g flowing to the gate [89] called induced gate noise. Since, the physical source of the thermal channel noise and the induce gate noise is common, the terminal noise currents at the drain and at the gate are correlated (at least partially) [39, 89, 90].

In this chapter, we carry out an RF noise analysis and derive the analytical expressions that describe the four RF noise parameters, namely, $F_{\min}$, R_n, G_{opt} and B_{opt} [39]. We demonstrate a step-by-step procedure for the extraction of the RF noise model parameters directly from measurements, which can be linked to the parameters of standard compact MOSFET models. The analytical expressions are validated against measurements of a commercial state-of-the-art 40 nm CMOS process from moderate- to strong-inversion region. Note that the reason for not demonstrating the analytical RF noise model in weak-inversion, is the difficulty to carry out RF noise measurements in this region and as a result there were no available data [5].

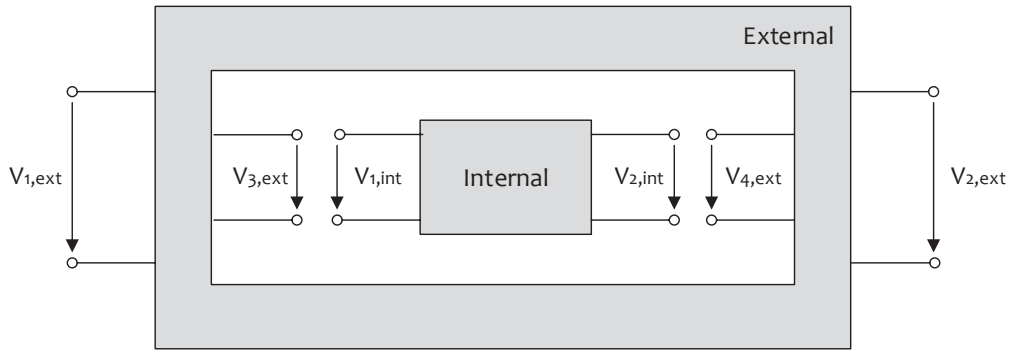


Figure 4.1 – For the RF noise analysis, a two-port DUT is divided in two parts, one external and one internal. Note that after the division the external part is a four-port and the internal part a two-port.

4.2 RF Noise Analysis

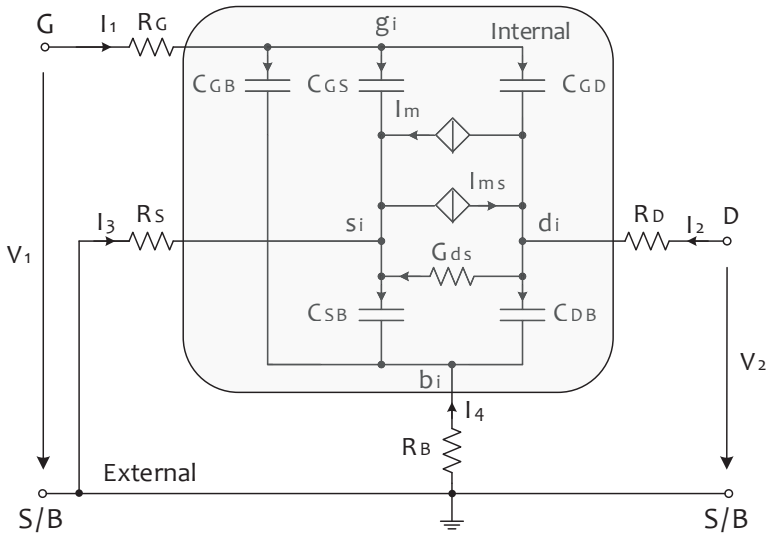
The target of this section is to derive the analytical expressions for the four RF noise parameters, namely, the minimum noise factor $F_{\min}$ (or minimum noise figure $NF_{\min} = 10 \cdot \log(F_{\min})$), the input referred noise resistance R_n and the real and imaginary part of the optimum source admittance Y_{opt} ($= G_{opt} + j \cdot B_{opt}$), that can be compared directly to the actual RF noise measurements [39]. The analysis is based on the noise correlation matrix of two-port networks [91] and on the multiport noise theory [92–94]. But, in contrast to all the work on the characterization of the RF noise found in literature, where noise de-embedding is first carried out in order to calculate the channel and the induced gate noise, we perform the reverse procedure. In the following few paragraphs, the concept of the reverse RF noise analysis is further elaborated.

The noise de-embedding is a bit tedious but it is also a well defined procedure [93, 94]. In short, in the noise de-embedding process the two-port MOS device is divided in two parts, one external and one internal. As shown graphically in Fig. 4.1, after the division the external part is a 4-port network, while the internal part is a 2-port network. Knowing the four RF noise parameters of the DUT, the equivalent circuit of the internal part and the four-port Y-parameters of the external part, the method allows one to predict the noise properties of the external part of the device. The noise of the external part can then be de-embbeded from the total DUT noise, leading to the four RF noise parameters of the internal part. From the four RF noise parameters of the internal part, the power spectral densities of the channel noise $S_{I_{nD}^2}$, the induced gate noise $S_{I_{nG}^2}$ and their correlation $S_{I_{nG},I_{nD}}$ can be calculated from [94]:

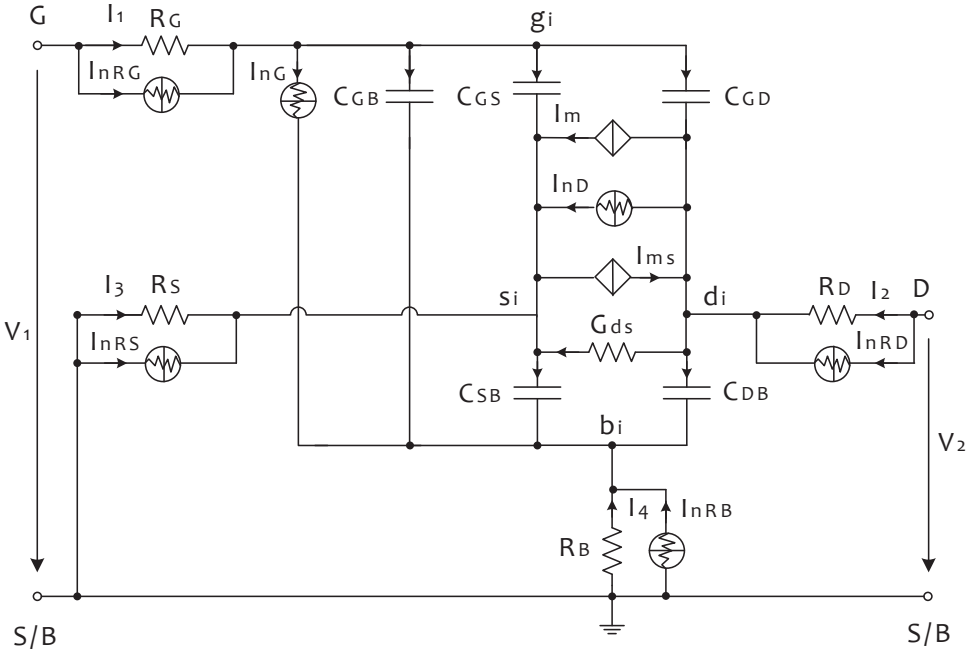
$$S_{I_{nD}^2} = 4kT \cdot R_n \cdot \left|Y_{21,int}\right|^2, \tag{4.1a}$$

$$S_{I_{nD}^2} = 4kT \cdot R_n \cdot \left(\left|Y_{opt}\right|^2 - \left|Y_{11,int}\right|^2 + 2 \cdot \Re\left(\left(Y_{11,int} - Y_{cor}\right) \cdot Y_{11,int}^* \right) \right), \tag{4.1b}$$

$$S_{I_{nG},I_{nD}} = 4kT \cdot R_n \cdot \left(Y_{11,int} - Y_{cor}\right) \cdot Y_{21,int}^*, \tag{4.1c}$$



(a) Quasi-static small-signal equivalent circuit of an RF MOSFET in saturation divided in two parts needed for the RF noise analysis



(b) Quasi-static small-signal equivalent circuit of an RF MOSFET in saturation including its noise sources

Figure 4.2 – (a) Quasi-static small-signal equivalent circuit of an RF MOSFET valid in saturation divided in two parts for the RF noise analysis. The internal part, includes the components inside the gray rectangular region, and the external part consists of the components outside the gray rectangular region. (b) Quasi-static small-signal equivalent circuit of Fig. 3.1b together with its noise sources of an RF MOS transistor.

where:

$$Y_{\text{cor}} = \frac{F_{\text{min}} - 1}{2 \cdot R_{\text{n}}} - Y_{\text{opt}}, \tag{4.2}$$

k is the Boltzmann's constant and T is the absolute temperature in Kelvin. Note that in the above expressions the noise parameter R_{n}, F_{min} and Y_{opt} refer to the internal part.

In our method, we exploit the well-known channel and induced gate noise models, then we embed all the other components that contribute to the thermal RF noise of the device and finally, we derive analytically the four RF noise parameters that should fit to the actual RF noise measurements. To do so, we also divide the RF MOS transistor equivalent circuit in two parts, using the same RF equivalent schematic as in Sec. 3.2, as shown in Fig. 4.2a. The quasi-static small signal equivalent schematic of an RF MOSFET including its noise sources is shown in Fig. 4.2b. The channel thermal noise is modeled by the noise current source I_{nD}, the induced gate noise by the noise current source I_{nG}, and the thermal noise of the resistances with noisy current sources parallel to each resistance.

4.2.1 Step-by-step Derivation of the Analytical Expressions of the Four RF Noise Parameters

Below the steps to derive the analytical expressions of the four RF noise parameters, performing a noise embedding method are described. The S/D series resistances can be neglected not only in the RF analysis but in the noise analysis as well since their typical values in MOSFETs are only a few Ohms and therefore their noise contribution is small [95]. Note that the correlation matrices throughout the analysis are normalized by the factor $2kTB$, where B is the noise bandwidth. This factor is canceled out in the final noise parameter expressions.

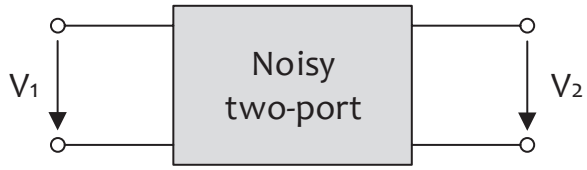
▷ **Step 1: Calculation of Y_{int}**

A Y-parameter analysis is carried out in order to calculate the two-port admittance matrix $\mathbf{Y_{int}}$ of the internal part, leading to:

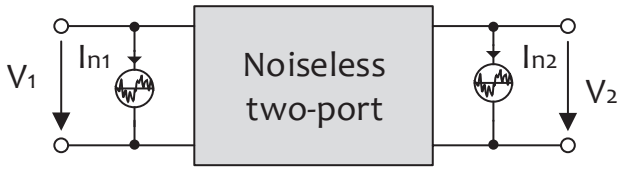
$$\mathbf{Y_{int}} = \begin{bmatrix} Y_{11,\text{int}} & Y_{12,\text{int}} \\ Y_{21,\text{int}} & Y_{22,\text{int}} \end{bmatrix} = \begin{bmatrix} j\omega C_{\text{GG}} & -j\omega C_{\text{GD}} \\ G_{\text{m}_{\text{eff}}} - j\omega\big(C_{\text{GD}} + C_{\text{m}}\big) & G_{\text{ds}} + j\omega\big(C_{\text{DB}} + C_{\text{GD}}\big) \end{bmatrix}. \tag{4.3}$$

▷ **Step 2: Calculation of $R_{\text{n,int}}$, $G_{\text{i,int}}$ and $Y_{\text{c,int}}$**

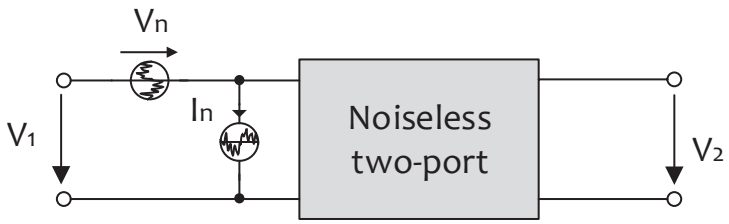
In this step, the input-referred thermal noise resistance $R_{\text{n,int}}$, the input-referred thermal noise conductance $G_{\text{i,int}}$ and the noise correlation admittance $Y_{\text{c,int}}$ of the internal part are calculated. For their calculation we will use the theory of noise analysis of multiport networks [92] according to which, any noisy two-port, shown in Fig. 4.3a, can be represented by its Norton equivalent, that is, its noiseless counterpart and a noise current source connected across each port (admittance representation), shown in Fig. 4.3b. However, for



(a) Noisy two-port network



(b) Admittance representation of the noisy two-port network



(c) Chain representation of the noisy two-port network

Figure 4.3 – Different representations of a linear noisy two-port. (a) Noisy two-port network. (b) Admittance representation of the noisy two port network. (c) Chain representation of the noisy two port network. [91]

noise calculation it is more convenient to refer both noise current sources to the input (chain representation) as shown in Fig. 4.3c. The two noise sources V_{n} and I_{n} of Fig. 4.3c are related to the noise sources I_{n1} and I_{n2} of Fig. 4.3b through the relations [92]:

$$V_{\mathrm{n}} = -\frac{I_{\mathrm{n2}}}{Y_{21}}, \tag{4.4a}$$

$$I_{\mathrm{n}} = I_{\mathrm{n1}} - \frac{Y_{11}}{Y_{21}} I_{\mathrm{n2}}, \tag{4.4b}$$

In a case of a MOS transistor the two noise current sources I_{n1} and I_{n2} of Fig. 4.3b are actually equivalent to I_{nG} and I_{nD} respectively.

The parameters $R_{\mathrm{n,int}}$, $G_{\mathrm{i,int}}$ and $Y_{\mathrm{c,int}}$ are given by [39,91]:

$$R_{\mathrm{n,int}} = \frac{\overline{\left|V_{\mathrm{n}}\right|^2}}{4kTB}, \tag{4.5a}$$

$$G_{\mathrm{i,int}} = \frac{\overline{\left|I_{\mathrm{n}}\right|^2}}{4kTB}, \tag{4.5b}$$

$$Y_{\mathrm{c,int}} = \frac{\overline{I_{\mathrm{n}} V_{\mathrm{n}}^*}}{\overline{\left|V_{\mathrm{n}}\right|^2}}, \tag{4.5c}$$

so knowing V_{n} and I_{n} from Eq. (4.4) and after some math calculations we end up with:

$$R_{\mathrm{n,int}} = \frac{\overline{\left|I_{\mathrm{nD}}\right|^2}}{4kTB \cdot \left|Y_{21,\mathrm{int}}\right|^2}, \tag{4.6a}$$

$$G_{\mathrm{i,int}} = \frac{\overline{\left|I_{\mathrm{nG}}\right|^2} + \frac{\left|Y_{11,\mathrm{int}}\right|^2}{\left|Y_{21,\mathrm{int}}\right|^2}\overline{\left|I_{\mathrm{nD}}\right|^2} - \frac{\left(Y_{21,\mathrm{int}} Y_{11,\mathrm{int}}^* \overline{I_{\mathrm{nG}} I_{\mathrm{nD}}^*} + Y_{21,\mathrm{int}}^* Y_{11,\mathrm{int}} \overline{I_{\mathrm{nG}}^* I_{\mathrm{nD}}}\right)}{\left|Y_{21,\mathrm{int}}\right|^2}}{4kTB}, \tag{4.6b}$$

$$Y_{\mathrm{c,int}} = Y_{11,\mathrm{int}} - Y_{21,\mathrm{int}} \frac{\overline{I_{\mathrm{nG}} I_{\mathrm{nD}}^*}}{\overline{\left|I_{\mathrm{nD}}\right|^2}}. \tag{4.6c}$$

In order to complete the calculation of the parameters $R_{\mathrm{n,int}}$, $G_{\mathrm{i,int}}$ and $Y_{\mathrm{c,int}}$, we need to know the mean-square values $\overline{\left|I_{\mathrm{nD}}\right|^2}$, $\overline{\left|I_{\mathrm{nG}}\right|^2}$, $\overline{I_{\mathrm{nG}} I_{\mathrm{nD}}^*}$ and $\overline{I_{\mathrm{nG}}^* I_{\mathrm{nD}}}$. For that, we use the well known analytical models for channel thermal noise and the induced gate noise [39]:

$$\overline{\left|I_{\mathrm{nD}}\right|^2} = 4kTB \cdot G_{\mathrm{nD}}, \tag{4.7a}$$

$$\overline{\left|I_{\mathrm{nG}}\right|^2} = 4kTB \cdot G_{\mathrm{nG}}, \tag{4.7b}$$

$$\overline{I_{\mathrm{nG}} I_{\mathrm{nD}}^*} = jc_{\mathrm{g}} \cdot 4kTB \cdot \sqrt{G_{\mathrm{nG}} G_{\mathrm{nD}}}, \tag{4.7c}$$

$$\overline{I_{\mathrm{nG}}^* I_{\mathrm{nD}}} = -jc_{\mathrm{g}} \cdot 4kTB \cdot \sqrt{G_{\mathrm{nG}} G_{\mathrm{nD}}}, \tag{4.7d}$$

$$\tag{4.7e}$$

where G_{nD} is the drain thermal noise conductance given by:

$$G_{\mathrm{nD}} = \gamma_{\mathrm{nD}} G_{\mathrm{m_{eff}}}, \tag{4.8}$$

and G_{nG} is the gate thermal noise conductance defined as:

$$G_{\mathrm{nG}} = \delta_{\mathrm{nG}} \frac{\omega^2 C_{\mathrm{GG}} C_{\mathrm{m}}}{2 \cdot G_{\mathrm{m_{eff}}}}. \tag{4.9}$$

In Eq. (4.8), Eq. (4.9) and Eq. (4.7), γ_{nD}, δ_{nG} and c_{g} are the three RF noise model parameters that will be used in the analytical expressions, with γ_{nD} being the thermal noise excess

factor at the drain, δ_{nG} being the thermal noise parameter at the gate and c_g the correlation parameter [39]. The long-channel values of these three RF noise model parameters in saturation are demonstrated in Table 4.1.

Table 4.1 – Long-channel values of the three RF noise model parameters in saturation

Parameter	Weak-Inversion	Strong-Inversion
γ_{nD}	$\frac{n}{2}$	$\frac{2}{3}n$
δ_{nG}	1	$\frac{4}{3}$
c_g	0.6	0.4

Finally, by combining Eq. (4.3), Eq. (4.6), Eq. (4.7), Eq. (4.8) and Eq. (4.9) we are able to calculate the analytical expressions for $R_{n,int}$, $G_{i,int}$ and $Y_{c,int}$ parameters of the internal part as:

$$R_{n,int} = \frac{G_{m_{eff}}\gamma_{nD}}{\left|Y_{21,int}\right|^2}, \tag{4.10a}$$

$$G_{i,int} = \omega^2 C_{GG}\left(\frac{G_{m_{eff}}\left(C_{GG}\gamma_{nD} - c_g\sqrt{2\cdot C_{GG}C_m\gamma_{nD}\delta_{nG}}\right)}{\left|Y_{21,int}\right|^2} + \frac{C_m\delta_{nG}}{2\cdot G_{m_{eff}}}\right), \tag{4.10b}$$

$$Y_{c,int} = -\omega^2 C_{noi}c_g\frac{C_{GD}+C_m}{G_{m_{eff}}} + j\omega\left(C_{GG} - C_{noi}c_g\right), \tag{4.10c}$$

where:

$$C_{noi} = \sqrt{\frac{C_{GG}C_m\delta_{nG}}{2\cdot\gamma_{nD}}}. \tag{4.11}$$

▷ **Step 3: Calculation of $C_{Y_{int}}$**

Knowing the $R_{n,int}$, $G_{i,int}$ and $Y_{c,int}$ parameters of the internal part, we can now proceed to the calculation of the admittance correlation matrix $\mathbf{C_{Y_{int}}}$ of the internal part, which is given by [91]:

$$\begin{aligned}\mathbf{C_{Y_{int}}} &= \begin{bmatrix} C_{Y_{11,int}} & C_{Y_{12,int}} \\ C_{Y_{21,int}} & C_{Y_{22,int}} \end{bmatrix} = \\ &= \begin{bmatrix} G_{iu,int} + \left(\left|Y_{11,int} - Y_{c,int}\right|^2\right)R_{n,int} & Y^*_{21,int}\left(Y_{11,int} - Y_{c,int}\right)R_{n,int} \\ Y_{21,int}\left(Y_{11,int} - Y_{c,int}\right)^* R_{n,int} & \left|Y_{21,int}\right|^2 R_{n,int} \end{bmatrix},\end{aligned} \tag{4.12}$$

where $G_{iu,int}$ is the uncorrelated part of G_i defined as:

$$G_{iu,int} = G_{i,int} - \left|Y_{c,int}\right|^2 R_{n,int}. \tag{4.13}$$

Using Eq. (4.10) in Eq. (4.12), we can express analytically the elements of $C_{\mathrm{Y_{int}}}$ as:

$$\mathbf{C_{Y_{int}}} = \begin{bmatrix} \dfrac{\omega^2 C_{\mathrm{GG}} C_{\mathrm{m}} \delta_{\mathrm{nG}}}{2 \cdot G_{\mathrm{m_{eff}}}} & j\omega c_{\mathrm{g}} \sqrt{\dfrac{C_{\mathrm{GG}} C_{\mathrm{m}} \gamma_{\mathrm{nD}} \delta_{\mathrm{nG}}}{2}} \\ -j\omega c_{\mathrm{g}} \sqrt{\dfrac{C_{\mathrm{GG}} C_{\mathrm{m}} \gamma_{\mathrm{nD}} \delta_{\mathrm{nG}}}{2}} & G_{\mathrm{m_{eff}}} \gamma_{\mathrm{nD}} \end{bmatrix}. \tag{4.14}$$

▷ **Step 4: Calculation of $C_{\mathbf{Y_{ext}}}$**

In this step the admittance correlation matrix $C_{\mathbf{Y_{ext}}}$ of the external part is calculated. First, a Y-parameter analysis is carried out, to calculate the four-port admittance matrix of the external part given by:

$$\mathbf{Y_{ext}} = \begin{bmatrix} \mathbf{Y_{ee}} & \mathbf{Y_{ei}} \\ \mathbf{Y_{ie}} & \mathbf{Y_{ii}} \end{bmatrix}, \tag{4.15}$$

where $\mathbf{Y_{ee}}$, $\mathbf{Y_{ei}}$, $\mathbf{Y_{ie}}$, $\mathbf{Y_{ii}}$ are 2x2 matrices, ending up with:

$$\mathbf{Y_{ee}} = \begin{bmatrix} \dfrac{1}{R_{\mathrm{G}}} & 0 \\ 0 & \dfrac{1}{R_{\mathrm{B}}} \end{bmatrix}, \tag{4.16a}$$

$$\mathbf{Y_{ei}} = \begin{bmatrix} -\dfrac{1}{R_{\mathrm{G}}} & 0 \\ 0 & -\dfrac{1}{R_{\mathrm{B}}} \end{bmatrix}, \tag{4.16b}$$

$$\mathbf{Y_{ie}} = \begin{bmatrix} -\dfrac{1}{R_{\mathrm{G}}} & 0 \\ 0 & -\dfrac{1}{R_{\mathrm{B}}} \end{bmatrix}, \tag{4.16c}$$

$$\mathbf{Y_{ii}} = \begin{bmatrix} \dfrac{1}{R_{\mathrm{G}}} & 0 \\ 0 & \dfrac{1}{R_{\mathrm{B}}} \end{bmatrix}. \tag{4.16d}$$

Then, knowing $\mathbf{Y_{ext}}$ we can calculate $\mathbf{C_{Y_{ext}}}$ based on [93]:

$$\mathbf{C_{Y_{ext}}} = \frac{1}{2}\left(\mathbf{Y_{ext}} + \mathbf{Y_{ext}}^{\dagger}\right) = \begin{bmatrix} \mathbf{C_{Y_{ee}}} & \mathbf{C_{Y_{ei}}} \\ \mathbf{C_{Y_{ie}}} & \mathbf{C_{Y_{ii}}} \end{bmatrix}, \tag{4.17}$$

where $\mathbf{C_{Y_{ee}}}$, $\mathbf{C_{Y_{ei}}}$, $\mathbf{C_{Y_{ie}}}$, $\mathbf{C_{Y_{ii}}}$ are also 2x2 matrices, and † denotes the Hermitian (conjugate-

transpose) of the associated matrix. Finally, $\mathbf{C_{Y_{ext}}}$ is calculated as:

$$\mathbf{C_{Y_{ext}}} = \begin{bmatrix} \frac{1}{R_G} & 0 & -\frac{1}{R_G} & 0 \\ 0 & \frac{1}{R_B} & 0 & -\frac{1}{R_B} \\ -\frac{1}{R_G} & 0 & \frac{1}{R_G} & 0 \\ 0 & -\frac{1}{R_B} & 0 & \frac{1}{R_B} \end{bmatrix}. \tag{4.18}$$

▷ **Step 5: Calculation of C_Y**

At this step we have all the elements to calculate the admittance correlation matrix $\mathbf{C_Y}$ of the DUT:

$$\mathbf{C_Y} = \begin{bmatrix} CY_{11} & CY_{12} \\ CY_{21} & CY_{22} \end{bmatrix} \tag{4.19}$$

from [93]:

$$\mathbf{C_Y} = \mathbf{C_{Y_{ee}}} + \mathbf{D}\,\mathbf{C_{Y_{ie}}} + \mathbf{C_{Y_{ei}}}\,\mathbf{D}^\dagger + \mathbf{D}\big(\mathbf{C_{Y_{ii}}} + \mathbf{C_{Y_{int}}}\big)\mathbf{D}^\dagger, \tag{4.20}$$

where $\mathbf{D}$ is the transformation matrix of the internal part given by:

$$\mathbf{D} = -\mathbf{Y_{ei}}\big(\mathbf{Y_{ii}} + \mathbf{Y_{int}}\big)^{-1}. \tag{4.21}$$

Note that from this step and on, all the calculations result to complex formulas, which require a mathematical tool to handle. Nevertheless, we chose to demonstrate these formulas and not only the final simplified ones, so that they can serve as a potential reference.

After performing the calculations indicated by Eq. (4.19), we end up with:

$$CY_{11} = \frac{\begin{aligned} &\Big(C_{GG}C_m\Big(1+\big(C_{DB}+C_{GD}\big)^2{R_B}^2\omega^2\Big)\delta_{nG} \\ &+2\cdot G_{m_{eff}}\Big({C_{GG}}^2R_G + R_B\Big(C_{GD}\big(C_{GD}+G_{m_{eff}}(2\cdot C_{GG}+C_{GD}G_{m_{eff}}R_B)R_G\big) \\ &+\Big(C_{DB}C_{GG}-C_{GD}\big(C_{GD}-C_{GG}+C_m\big)\Big)^2R_BR_G\omega^2\Big) \\ &+C_{GD}R_B\big(C_{GD}G_{m_{eff}}R_B\gamma_{nD}+c_g\sqrt{2\cdot C_{GG}C_m\delta_{nG}\gamma_{nD}}\big)\Big)\Big)\omega^2 \end{aligned}}{D}, \tag{4.22a}$$

$$CY_{12} = \frac{\left[2 \cdot G_{m_{eff}}\left(C_{GG}R_G\left(jG_{m_{eff}} - \left(C_{GD} + C_m + \left(C_{DB} + C_{GD}\right)G_{m_{eff}}R_B\right)\omega - j\left(C_{DB} + C_{GD}\right)C_mR_B\omega^2\right) + C_{GD}R_B\left(-\left(C_{DB} + C_{GD}\right)\omega\right) + R_G\left(jG_{m_{eff}} + C_m\omega\right)\left(G_{m_{eff}} + j\left(C_{GD} + C_m\right)\omega\right)\right)\right) + C_{GG}C_mR_G\omega\left(-jG_{m_{eff}} + \left(C_{GD} + C_m\right)\omega\right)\left(-j + \left(C_{DB} + C_{GD}\right)R_B\omega\right)\delta_{nG} + G_{m_{eff}}\left(2 \cdot C_{GD}G_{m_{eff}}R_B\left(j + C_{GG}R_G\omega\right)\gamma_{nD} + c_g\sqrt{2 \cdot C_{GG}C_m\delta_{nG}\gamma_{nD}}\left(j + \omega\left(C_{GG}R_G - R_B\left(C_{DB} + C_{GD} + C_{GD}G_{m_{eff}}R_G\right) - j\left(-\left(C_{DB}C_{GG}\right) + C_{GD}\left(C_{GD} - C_{GG} + C_m\right)\right)R_BR_G\omega\right)\right)\right)\right]\omega}{D}, \quad (4.22b)$$

$$CY_{21} = \frac{\left[2 \cdot G_{m_{eff}}\left(C_{GG}R_G\left(-jG_{m_{eff}} - \left(C_{GD} + C_m + \left(C_{DB} + C_{GD}\right)G_{m_{eff}}R_B\right)\omega + j\left(C_{DB} + C_{GD}\right)C_mR_B\omega^2\right) + C_{GD}R_B\left(-\left(C_{DB} + C_{GD}\right)\omega + R_G\left(-jG_{m_{eff}} + C_m\omega\right)\left(G_{m_{eff}} - j\left(C_{GD} + C_m\right)\omega\right)\right)\right) + C_{GG}C_mR_G\omega\left(jG_{m_{eff}} + \left(C_{GD} + C_m\right)\omega\right)\left(j + \left(C_{DB} + C_{GD}\right)R_B\omega\right)\delta_{nG} + G_{m_{eff}}\left(2 \cdot C_{GD}G_{m_{eff}}R_B\left(-j + C_{GG}R_G\omega\right)\gamma_{nD} + c_g\sqrt{2 \cdot C_{GG}C_m\delta_{nG}\gamma_{nD}}\left(-j + \omega\left(C_{GG}R_G - R_B\left(C_{DB} + C_{GD} + C_{GD}G_{m_{eff}}R_G\right) + j\left(-\left(C_{DB}C_{GG}\right) + C_{GD}\left(C_{GD} - C_{GG} + C_m\right)\right)R_BR_G\omega\right)\right)\right)\right]\omega}{D}, \quad (4.22c)$$

$$CY_{22} = \frac{\begin{aligned} &C_{GG}C_mR_G{}^2\omega^2 \cdot \left(G_{m_{eff}}{}^2 + \left(C_{GD} + C_m\right)^2\omega^2\right)\delta_{nG} \\ &+2 \cdot G_{m_{eff}}\left(\left(C_{DB} + C_{GD}\right)^2R_B\omega^2 + R_G\left(G_{m_{eff}}{}^2\right.\right. \\ &\left.\left.+\left(C_{GD} + C_m\right)^2\omega^2 + C_{GD}G_{m_{eff}}R_B\left(2 \cdot \left(C_{DB} + C_{GD}\right) + C_{GD}G_{m_{eff}}R_G\right)\omega^2\right.\right. \\ &\left.\left.+\left(C_{DB}C_{GG} - C_{GD}\left(C_{GD} - C_{GG} + C_m\right)\right)^2R_BR_G\omega^4\right) \\ &+\left(G_{m_{eff}} + C_{GG}{}^2G_{m_{eff}}R_G{}^2\omega^2\right)\gamma_{nD} - c_g\sqrt{2 \cdot C_{GG}C_m\delta_{nG}\gamma_{nD}}R_G\left(C_{GD} + C_m + C_{GG}G_{m_{eff}}R_G\right)\omega^2\Big)\end{aligned}}{D}, \quad (4.22d)$$

where D is the common denominator calculated by:

$$\begin{aligned} D = {} & 2\cdot G_{\mathrm{m_{eff}}}\Big(1+\Big(\big(C_{\mathrm{DB}}+C_{\mathrm{GD}}\big)^2 {R_{\mathrm{B}}}^2+2C_{\mathrm{GD}}R_{\mathrm{B}}(C_{\mathrm{GD}}+C_{\mathrm{m}} \\ & +\big(C_{\mathrm{DB}}+C_{\mathrm{GD}}\big)G_{\mathrm{m_{eff}}}R_{\mathrm{B}})R_{\mathrm{G}}+\big(C_{\mathrm{GG}}+C_{\mathrm{GD}}G_{\mathrm{m_{eff}}}R_{\mathrm{B}}\big)^2 {R_{\mathrm{G}}}^2\Big)\omega^2 \\ & +\Big(C_{\mathrm{DB}}C_{\mathrm{GG}}-C_{\mathrm{GD}}\big(C_{\mathrm{GD}}-C_{\mathrm{GG}}+C_{\mathrm{m}}\big)\Big)^2 {R_{\mathrm{B}}}^2{R_{\mathrm{G}}}^2\omega^4\Big). \end{aligned} \tag{4.23}$$

▷ **Step 6: Calculation of C_{A}**

In this step the admittance correlation matrix $\mathbf{C_Y}$ of the device is converted to its chain representation $\mathbf{C_A}$:

$$\mathbf{C_A}=\begin{bmatrix} CA_{11} & CA_{12} \\ CA_{21} & CA_{22}\end{bmatrix}, \tag{4.24}$$

as shown in [91, 93], using:

$$\mathbf{C_A}=\mathbf{V}^{-1}\mathbf{C_Y}\,\mathbf{V}^{\dagger-1}, \tag{4.25}$$

with:

$$\mathbf{V}=\begin{bmatrix} -\mathbf{Y_{e,11}} & 1 \\ -\mathbf{Y_{e,21}} & 0\end{bmatrix}, \tag{4.26}$$

and

$$\mathbf{Y_e}=\mathbf{Y_{ee}}+\mathbf{D}\mathbf{C_{Y_{ie}}}. \tag{4.27}$$

The calculations result to the following analytical expressions for the components of the $\mathbf{C_A}$ matrix:

$$\begin{aligned} CA_{11} = \frac{1}{2\cdot G_{\mathrm{m_{eff}}}\big|Y_{21,\mathrm{int}}\big|^2} & \Bigg(2\cdot G_{\mathrm{m_{eff}}}\Bigg(\big|Y_{21,\mathrm{int}}\big|^2 R_{\mathrm{G}}+R_{\mathrm{B}}\Bigg(\Big(C_{\mathrm{DB}}+C_{\mathrm{GD}}+C_{\mathrm{GD}}G_{\mathrm{m_{eff}}}R_{\mathrm{G}}\Big)^2 \\ & +\Big(C_{\mathrm{DB}}C_{\mathrm{GG}}-C_{\mathrm{GD}}\big(C_{\mathrm{GD}}-C_{\mathrm{GG}}+C_{\mathrm{m}}\big)\Big)^2 {R_{\mathrm{G}}}^2\omega^2\Bigg)\omega^2+G_{\mathrm{m_{eff}}}\gamma_{\mathrm{nD}}\Bigg) \\ & +R_{\mathrm{G}}\Bigg(C_{\mathrm{GG}}C_{\mathrm{m}}\big|Y_{21,\mathrm{int}}\big|^2 R_{\mathrm{G}}\delta_{\mathrm{nG}}+2\cdot G_{\mathrm{m_{eff}}}\Big({C_{\mathrm{GG}}}^2\cdot G_{\mathrm{m_{eff}}}R_{\mathrm{G}}\gamma_{\mathrm{nD}} \\ & -c_{\mathrm{g}}\sqrt{2\cdot C_{\mathrm{GG}}C_{\mathrm{m}}\delta_{\mathrm{nG}}\gamma_{\mathrm{nD}}}\big(C_{\mathrm{GD}}+C_{\mathrm{m}}+C_{\mathrm{GG}}G_{\mathrm{m_{eff}}}R_{\mathrm{G}}\big)\Big)\Bigg)\omega^2 \end{aligned}, \tag{4.28a}$$

$$CA_{12} = \frac{\left(C_{GG}C_m\left|Y_{21,int}\right|^2 R_G\delta_{nG}\omega + G_{m_{eff}}\left(2R_B\left(C_{GD}G_{m_{eff}}(C_{DB} + C_{GD} + C_{GD}G_{m_{eff}}R_G\right) + j\left(C_{DB}+C_{GD}\right)\left(-C_{DB}C_{GG} + C_{GD}(C_{GD} - C_{GG} + C_m)\right)\omega + R_G\left(C_{DB}C_{GG} - C_{GD}(C_{GD} - C_{GG} + C_m)\right)^2\omega^2\right)\omega + 2C_{GG}G_{m_{eff}}\left(-j + C_{GG}R_G\omega\right)\gamma_{nD} - c_g\sqrt{2\cdot C_{GG}C_m\delta_{nG}\gamma_{nD}}\left(-jG_{m_{eff}} + \left(C_{GD} + C_m + 2C_{GG}G_{m_{eff}}R_G\right)\omega\right)\right)\omega}{2\cdot G_{m_{eff}}\left|Y_{21,int}\right|^2}, \tag{4.28b}$$

$$CA_{21} = \frac{\left(C_{GG}C_m\left|Y_{21,int}\right|^2 R_G\delta_{nG}\omega + G_{m_{eff}}\left(2R_B\left(C_{GD}G_{m_{eff}}(C_{DB} + C_{GD} + C_{GD}G_{m_{eff}}R_G\right) + j\left(C_{DB}+C_{GD}\right)\left(C_{DB}C_{GG} - C_{GD}(C_{GD} - C_{GG} + C_m)\right)\omega + R_G\left(C_{DB}C_{GG} - C_{GD}(C_{GD} - C_{GG} + C_m)\right)^2\omega^2\right)\omega + 2C_{GG}G_{m_{eff}}\left(j + C_{GG}R_G\omega\right)\gamma_{nD} - c_g\sqrt{2\cdot C_{GG}C_m\delta_{nG}\gamma_{nD}}\left(jG_{m_{eff}} + \left(C_{GD} + C_m + 2C_{GG}G_{m_{eff}}R_G\right)\omega\right)\right)\omega}{2\cdot G_{m_{eff}}\left|Y_{21,int}\right|^2}, \tag{4.28c}$$

$$CA_{22} = \frac{\left(C_{GG}C_m\left|Y_{21,int}\right|^2\delta_{nG} + 2G_{m_{eff}}\left(C_{GD}{}^2G_{m_{eff}}{}^2R_B + R_B\left(C_{DB}C_{GG} - C_{GD}\left(C_{GD} - C_{GG} + C_m\right)\right)^2\omega^2 + C_{GG}G_{m_{eff}}\left(C_{GG}\gamma_{nD} - c_g\sqrt{2\cdot C_{GG}C_m\delta_{nG}\gamma_{nD}}\right)\right)\omega^2}{2\cdot G_{m_{eff}}\left|Y_{21,int}\right|^2}. \tag{4.28d}$$

with $\left|Y_{21,int}\right|^2$ being the magnitude of the Y_{21} of the internal part given by:

$$\left|Y_{21,int}\right|^2 = G_{m_{eff}}{}^2 + \left(C_{GD} + C_m\right)^2\omega^2. \tag{4.29}$$

▷ **Step 7: Calculation of the four RF noise parameters of the DUT**

The final step in this procedure is to calculate the analytical expressions for the four RF noise parameters of the DUT from [94]:

$$R_n = C_{A,11}, \tag{4.30a}$$

$$F_{min} = 1 + 2\left(\Re\{C_{A,12}\} + \sqrt{C_{A,11}C_{A,22} - \Im\{C_{A,12}\}^2}\right), \tag{4.30b}$$

$$Y_{opt} = \sqrt{C_{A,11}C_{A,22} - \Im\{C_{A,12}\}^2} + j\frac{\Im\{C_{A,12}\}^2}{C_{A,11}}. \tag{4.30c}$$

Performing the calculations indicated by Eq. (4.30) and after simplification, we can derive

the analytical expression for the four RF noise parameters of the device as shown below:

$$R_{\mathrm{n}} = \frac{\gamma_{\mathrm{nD}}}{G_{\mathrm{m_{eff}}}} + R_{\mathrm{G}}, \tag{4.31a}$$

$$F_{\mathrm{min}} = 1 + \omega \frac{b_{\mathrm{N}}}{\left|Y_{21,\mathrm{int}}\right|^2} + \omega^2 \frac{2 \cdot R_{\mathrm{B}}\Big(G_{\mathrm{m_{eff}}} C_{\mathrm{GD}}\big(C_{\mathrm{DB}} + C_{\mathrm{GD}} + G_{\mathrm{m_{eff}}} R_{\mathrm{G}} C_{\mathrm{GD}}\big)\Big) + \omega^2 e_{\mathrm{N}} R_{\mathrm{G}}}{{G_{\mathrm{m_{eff}}}}^2}, \tag{4.31b}$$

$$Y_{\mathrm{opt}} = G_{\mathrm{opt}} + jB_{\mathrm{opt}} = \frac{\omega\left(b_{\mathrm{N}} + j\, c_{\mathrm{N}} G_{\mathrm{m_{eff}}}\right)}{2\left(d_{\mathrm{N}} + \omega^2 R_{\mathrm{B}}\big(C_{\mathrm{DB}} + C_{\mathrm{GD}}\big)^2\right)}, \tag{4.31c}$$

with:

$$a_{\mathrm{N}} = c_{\mathrm{g}}\sqrt{2 \cdot C_{\mathrm{GG}} C_{\mathrm{m}} \delta_{\mathrm{nG}} \gamma_{\mathrm{nD}}}, \tag{4.32a}$$

$$b_{\mathrm{N}} = \sqrt{\frac{2 \cdot d_{\mathrm{N}}\Big[2\, G_{\mathrm{m_{eff}}}\Big(\omega^2 e_{\mathrm{N}} R_{\mathrm{B}} + C_{\mathrm{GG}} G_{\mathrm{m_{eff}}}\big(C_{\mathrm{GG}}\gamma_{\mathrm{nD}} - a_{\mathrm{N}}\big)\Big) + C_{\mathrm{GG}} C_{\mathrm{m}}\left|Y_{21,\mathrm{int}}\right|^2 \delta_{\mathrm{nG}}\Big] - c_{\mathrm{N}}^2 {G_{\mathrm{m_{eff}}}}^3}{G_{\mathrm{m_{eff}}}}}, \tag{4.32b}$$

$$c_{\mathrm{N}} = a_{\mathrm{N}} - 2 \cdot \gamma_{\mathrm{nD}} C_{\mathrm{GG}}, \tag{4.32c}$$

$$d_{\mathrm{N}} = G_{\mathrm{m_{eff}}}\gamma_{\mathrm{nD}} + \left|Y_{21,\mathrm{int}}\right|^2 R_{\mathrm{G}}, \tag{4.32d}$$

$$e_{\mathrm{N}} = \Big(C_{\mathrm{DB}} C_{\mathrm{GG}} - C_{\mathrm{GD}}\big(C_{\mathrm{GD}} - C_{\mathrm{GG}} + C_{\mathrm{m}}\big)\Big)^2. \tag{4.32e}$$

From the simplified analytical expressions in Eq. (4.31), we notice that R_{n} depends mainly on the channel thermal noise through the parameter γ_{nD}, whereas both F_{min} and Y_{opt} are strongly dependent on both the channel thermal noise and the induced gate noise [43]. We should mention here that if R_{B} is small, just a few Ohms, then it has almost no impact on the total RF noise and can be neglected. On the other hand the contribution of R_{G} is important for RF noise modeling and characterization and thus, special attention needs to be paid in the extraction of R_{G}, as it is crucial to separate the noise contribution of R_{G} from the intrinsic channel [95]. Furthermore, especially for low-noise applications, the gate resistance of devices should be minimized [48, 95–97].

4.3 Expressions for the Direct Extraction of the RF Noise Model Parameters

The simplified analytical expressions of the four RF noise parameters given by Eq. (4.31) can be used for the direct extraction of the RF noise model parameters, specifically γ_{nD}, δ_{nG} and

c_g. In order to extract directly these parameters from measurements, a five consecutive step procedure should be followed as described in Eq. (4.33). Obviously, for the extraction of the RF noise parameters, the RF parameters should have been extracted previously, using the procedure described in Sec. 3.4. First, γ_{nD} can be extracted from R_n. Then two intermediate steps are needed to extract a_N and b_N from B_{opt} and G_{opt}, respectively. Note that b_N can be extracted also from F_{min}, but it leads to a more complex analytical expression, so for simplicity G_{opt} is chosen for the extraction of b_N. Finally, solving a_N and b_N definitions given by Eq. (4.32a) and Eq. (4.32b), δ_{nG} and c_g parameters can be extracted, respectively. Similar, to the RF parameter extraction procedure, the RF noise parameter extraction procedure Eq. (4.33), should be performed for each operating point or device geometry. It should be noted here that in most cases RF noise measurements suffer a lot of scattering, which makes direct extraction of the RF noise model parameters even more challenging.

RF Noise Model Parameter Step-by-Step Extraction:

$$\text{Step 1:}\quad \gamma_{nD} = G_{m_{eff}}\left(R_n - R_G\right) \tag{4.33a}$$

$$\text{Step 2:}\quad a_N = \frac{2\cdot\left[B_{opt}\left(\left|Y_{21,int}\right|^2 R_G + \omega^2 R_B\left(C_{DB}+C_{GD}\right)^2\right) + G_{m_{eff}}\gamma_{nD}\left(B_{opt}+\omega C_{GG}\right)\right]}{\omega G_{m_{eff}}} \tag{4.33b}$$

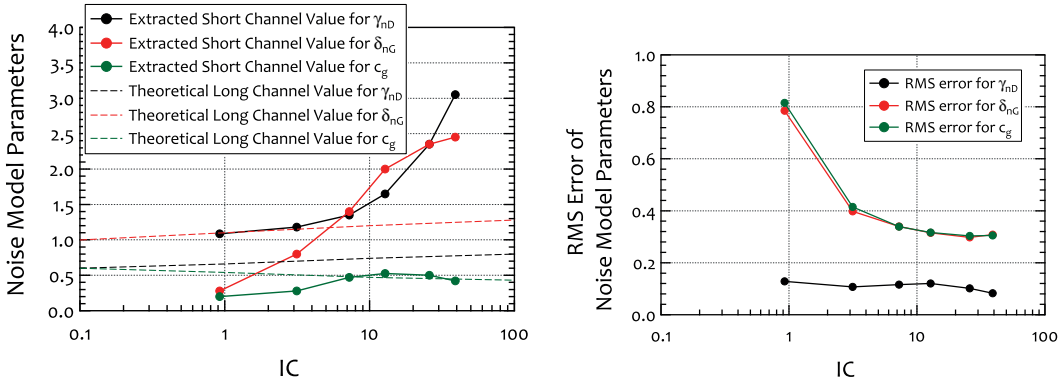
$$\text{Step 3:}\quad b_N = \frac{2\cdot G_{opt}\left(G_{m_{eff}}\gamma_{nD} + \left|Y_{21,int}\right|^2 R_G + \omega^2 R_B\left(C_{DB}+C_{GD}\right)^2\right)}{\omega} \tag{4.33c}$$

$$\text{Step 4:}\quad \delta_{nG} = \frac{G_{m_{eff}}\left[b_N^2 + c_N^2 {G_{m_{eff}}}^2 - 4\cdot d_N\left(\omega^2 e_N R_B + G_{m_{eff}} C_{GG}\left(C_{GG}\gamma_{nD} - a_N\right)\right)\right]}{2\cdot d_N C_{GG} C_m \left|Y_{21,int}\right|^2} \tag{4.33d}$$

$$\text{Step 5:}\quad c_g = \frac{a_N}{\sqrt{2\,C_{GG}C_m\gamma_{nD}\delta_{nG}}} \tag{4.33e}$$

4.3.1 Validation of the Analytical RF Noise Model and the Parameter Extraction Procedure

For the validation of the analytical noise expressions, the de-embedded RF noise measurements up to 18 GHz, of the same 40 nm RF DUT as in the RF analysis (M = 6, N_f = 10, W = 2 µm) were used. The measurements were carried out using a standard RF noise measurement set-up. Similar to the S-parameters, a procedure based on 'OPEN' and 'SHORT' dummy test structures [59] has been applied for the de-embedding of the RF noise measurements. In order to extract the four RF noise parameters the NF50 method is used [60]. Following the procedure in Eq. (4.33), the RF noise model parameters were extracted in saturation (V_D = 1.1V) and from moderate to strong-inversion (V_G = 0.4V...1.1V). In Fig. 4.4a, the extracted values of the RF noise parameters of the DUT vs. *IC* are shown, along with the theoretical long channel values [39]. From moderate- to strong-inversion γ_{nD} varies from



(a) Noise model parameters vs. *IC*

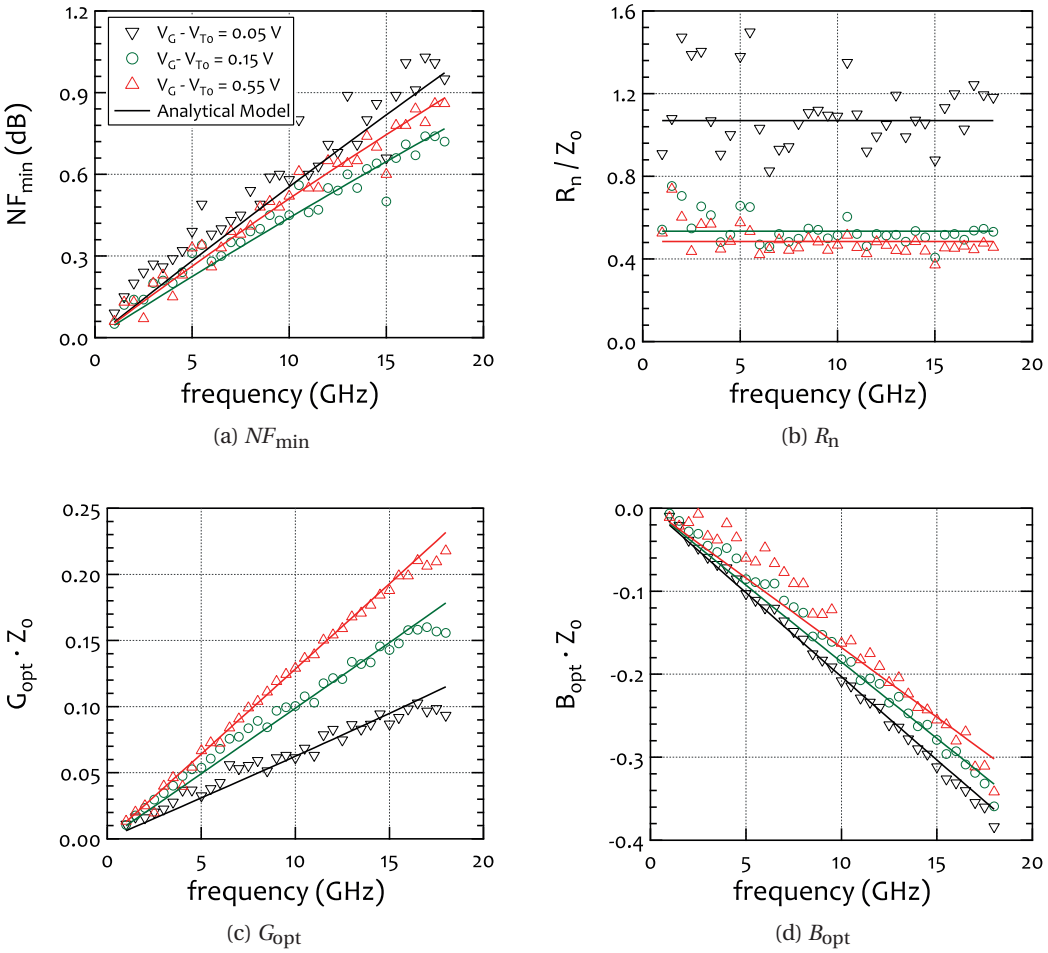
(b) RMS error of the extracted noise model parameters vs. *IC*, with respect to the analytical expressions

Figure 4.4 – (a) Extracted values of the RF noise model parameters of the small signal-equivalent circuit vs. *IC* of a minimum channel length RF nMOS DUT with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_D = 1.1$ V. The theoretical long channel values are also shown. (b) RMS error of the extracted RF noise model parameters vs. *IC*, with respect to the analytical expressions used for their extraction across frequency. The RMS error is expected to be large due to the great dispersion of the noise measurements, especially at lower levels of inversion.

~ 1 to 3, δ_{nG} from ~ 0.3 to 2.5 and c_g from ~ 0.2 to 0.5. The difference from the theoretical values for long channel devices can be attributed to short channel effects e.g., velocity saturation, channel length modulation, carrier heating etc. that directly impact the values of γ_{nD} and δ_{nG} [39, 43, 98, 99]. The extracted values for γ_{nD} and δ_{nG} are in agreement with the obtained results in [97, 98], corroborating the adopted approach. Additionally, the RMS error for each of the extracted RF noise model parameters vs. *IC*, with respect to the analytical expressions used for their extraction across frequency, is shown in Fig. 4.4b. As expected, the RMS error is quite large due to the great dispersion of the noise measurements and especially at lower levels of inversion, at which the RF noise is very difficult to measure.

Using the extracted values of the parameters of the small-signal equivalent circuit and the RF noise model parameters, we can compare the analytical model Eq. (4.31) against measurements. In Fig. 4.5, the four RF noise parameters vs. frequency, for different levels of inversion are displayed, while in Fig. 4.6, the four RF noise parameters vs. *IC* are shown as well. From the plots we see that the analytical model is capturing accurately the RF noisy behavior of the device from moderate to strong-inversion. In Fig. 4.6a we observe also that the minimum NF_{min} is achieved in moderate-inversion region. This adds one more argument in favor of biasing RF circuits in this region when targeting for low-power operation [54, 86–88]. Note that in moderate-inversion the f_T is already some tens of GHz for the technology under study.

For further validation of the consistency of the analytical model, the noise figure *NF* at two different bias points, in moderate- and strong-inversion, is demonstrated in Fig. 4.7a. The measured *NF* at a varying Z_S around 50 Ω (as shown in Fig. 4.7b) is compared to the analytical



(a) NF_{min} (b) R_{n}

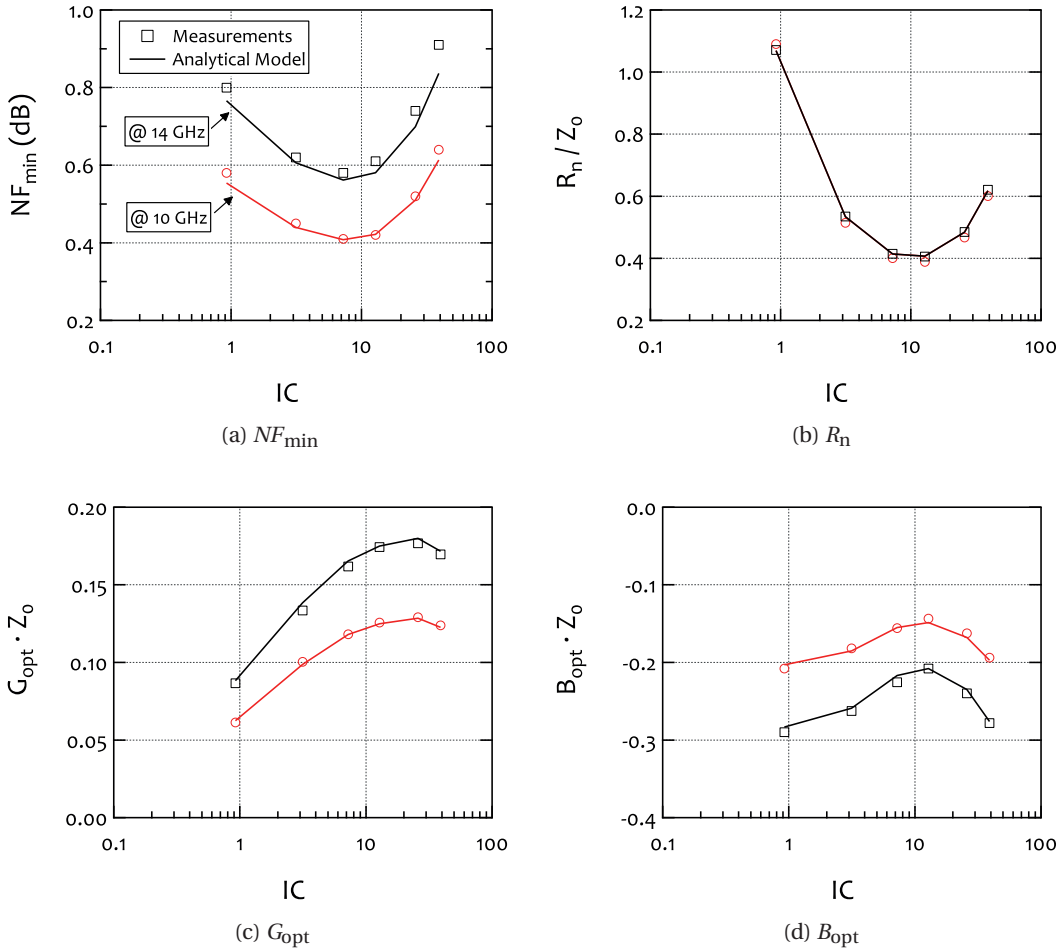
(c) G_{opt} (d) B_{opt}

Figure 4.5 – The four RF noise parameters vs. frequency of a minimum channel length RF nMOS DUT with $M = 6$, $N_{\mathrm{f}} = 10$, $W = 2$ µm and $L = 40$ nm, at $V_{\mathrm{G}} - V_{\mathrm{T0}} = [0.05, 0.15, 0.55]$ V and $V_{\mathrm{D}} = 1.1$ V with $Z_0 = 50\,\Omega$: (a) NF_{min}, (b) R_{n}, (c) G_{opt} and (d) B_{opt}. Note that at $V_{\mathrm{G}} - V_{\mathrm{T0}} = [0.05, 0.15]$ V the device is in moderate inversion ($0.1 < IC < 10$) and at $V_{\mathrm{G}} - V_{\mathrm{T0}} = 0.55$ V the device is in strong-inversion ($IC > 10$).

$NF = 10 \cdot \log(F)$, with F being calculated by the formula [49]

$$F = F_{\mathrm{min}} + \frac{R_{\mathrm{n}}}{G_{\mathrm{S}}} \cdot \left| Y_{\mathrm{S}} - Y_{\mathrm{opt}} \right|^2, \tag{4.34}$$

using the measured source admittance $Y_{\mathrm{S}} = 1/Z_{\mathrm{S}}$, the measured source conductance $G_{\mathrm{S}} = \mathrm{Real}(Y_{\mathrm{s}})$ and the analytical four RF noise parameters shown in Fig. 4.7. In addition, the $NF_{50} = 10 \cdot \log(F_{50})$, which is the noise figure at a constant source impedance $Z_{\mathrm{S}} = R_{\mathrm{S}} = 50\,\Omega$, is calculated for the analytical model. The analytical model shows a good agreement with the measured NF, especially accounting for the fact that an uncertainty remains at higher frequencies between the actual source impedance value during the noise measurement and the one obtained from a separate source impedance measurement. Note that the noise factor



(a) NF_{min} (b) R_{n}

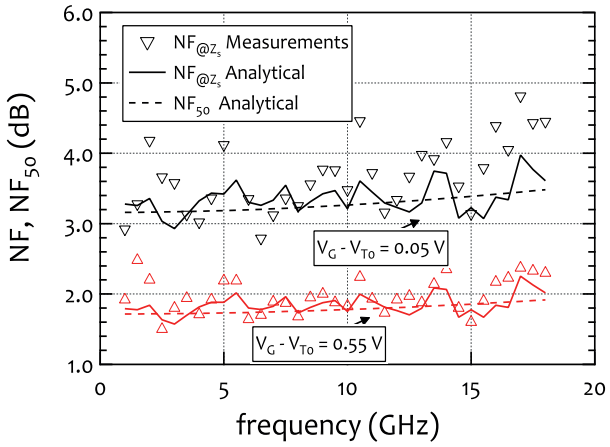
(c) G_{opt} (d) B_{opt}

Figure 4.6 – The four RF noise parameters vs. *IC*, of a minimum channel length RF nMOS DUT with $M = 6$, $N_{\text{f}} = 10$, $W = 2$ µm and $L = 40$ nm, at $V_{\text{D}} = 1.1$ V and f = [10 GHz, 14 GHz], with $Z_0 = 50\ \Omega$. It is clearly observed that the minimum values of NF_{min} and R_{n} are achieved in the onset between moderate and strong-inversion, whereas regarding G_{opt} and B_{opt} their minimum value is in moderate inversion.

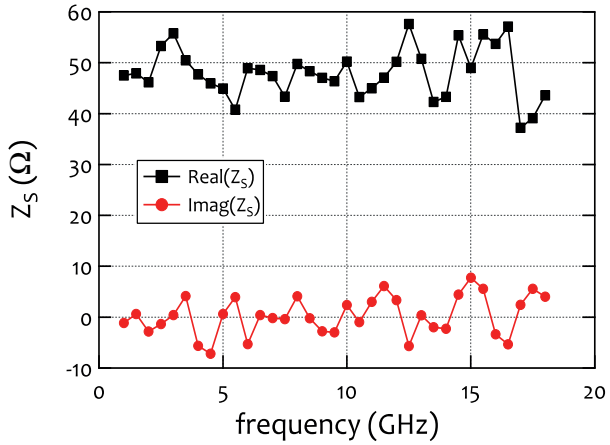
defined by Eq. (4.34) is frequency dependent, so it is sometimes called spot noise factor while the noise factor is obtained from the spot noise factor after integration over a certain bandwidth. If the bandwidth is narrow the spot noise factor and the noise factor are almost equal [39].

4.4 Conclusions

Accurate RF noise modeling is a prerequisite for RF IC design, especially for low-power applications. In this chapter we presented analytical expressions able to model the four RF noise parameters of nanoscale MOSFETs. For the analysis, a generic, and simple RF small-signal equivalent circuit of the transistor was used but a different approach, reverse to the noise



(a) *NF*



(b) Z_S

Figure 4.7 – *NF* and Z_S vs. frequency of a minimum channel length RF nMOS DUT with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_G - V_{T0} = [0.05, 0.55]$ V and $V_D = 1.1$ V: (a) The measured $NF_{@Z_s}$ is compared to the analytical $NF_{@Z_s}$, which is calculated using Eq. (4.34) and the measured Z_S shown in (b). The analytical NF_{50} is calculated using Eq. (4.34) but with a constant impedance $Z_S = R_S = 50\ \Omega$. (b) Measured Z_S (real and imaginary parts). Note that at $V_G - V_{T0} = 0.05$ V the device is in moderate inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = 0.55$ V the device is in strong-inversion ($IC > 10$).

de-embedding procedure, was adopted. We were able to model analytically the four RF noise parameters, accounting for the channel thermal noise and the induced gate noise. These expressions can be directly used to predict the RF noise parameters of the device, without the need to calculate first the individual noise contributions e.g., I_{nD}, I_{nG} etc.

For the extraction of the RF noise model parameters, a step-by-step procedure for the extraction of their values directly from measurements was shown for the first time [5]. The V_G dependence of all the model parameters was also presented, which can be useful for implementation in compact models, for which a global fit is desirable. The advantage of

the proposed parameter extraction methodology is that it remains consistent with existing standard compact MOSFET models.

Compact Modeling with BSIM6 Part II

5 The BSIM6 Compact Model

5.1 Introduction

Advanced Systems-on-Chip (SOCs) consist of millions of devices with the MOS transistor being the key element among them. Due to the increasing complexity of modern ICs, designers rely heavily on device models that serve as an accurate interface between circuit design and fabrication. No matter how much design effort is put, it is the ability of the compact model to describe the device's characteristics accurately that will guarantee the precise operation of circuits. This is especially true with the aggressive downscaling of advanced bulk CMOS technologies that demands MOSFET models able to describe correctly the behavior of devices accounting for all the physical phenomena. A reliable model should have the ability to handle all the different operating regions of the MOS transistor in the whole geometry range of one technology. It should also be robust and cause no convergence issues during circuit simulation. Targeting to meet the aforementioned needs, the Berkeley Short-Channel Insulated-Gate-Fet Model (BSIM) family has introduced its latest member, BSIM6 as the next generation compact model for the conventional bulk MOSFET. In this chapter, we will describe the main characteristics of the BSIM6 compact model.

5.2 A Short History

The BSIM family of compact models, developed by the UC Berkeley, are extensively used by semiconductor and IC design companies for more than 20 years. The third version of BSIM3, namely BSIM3v3, became the first industry standard of its kind on December 1996 [100] and BSIM4 was released in 2000. And although, both BSIM3 and its successor BSIM4 were adopted by most IC companies worldwide due to their accuracy and simulation speed, a subtle but important asymmetry around $V_{DS} = 0\ V$, forced the BSIM group to start the development of BSIM6 in late 2010 [101]. Recognizing that the source of the asymmetry issue was the threshold-voltage-based scheme used in the core of BSIM3 and BSIM4, the BSIM group searched for

different known approaches for the core of the BSIM6 model, choosing the charged-based one [39,102]. In less than 3 years and after undergoing an intense and thorough benchmarking, BSIM6 was standardized in 2013 by the Compact Modeling Coalition (CMC) [103].

5.3 BSIM6 Main Characteristics

In the core of BSIM6, the charge-based approach of the EKV model has been adopted. The main advantages of this approach are a) its physical nature, which ensures consistency between the actual behavior of MOSFET and the behavior predicted by the compact model, b) accuracy and continuity in all regions of operation, from weak to strong inversion and from linear operation to saturation; this results from the continuous current-charge and charge-voltage relations [39,104], and c) computational efficiency since the calculated charges can be used without any intermediate steps, to calculate all the other quantities e.g. current, capacitances, noise etc. Unlike its predecessors that were source-referenced, BSIM6 is a bulk-referenced model and thus consistent with the existing symmetry between source and drain of real devices [105].

The core charge density equation of BSIM6 [104], is given by:

$$2q_{\mathrm{i}} + ln(q_{\mathrm{i}}) + ln\left(\frac{4n}{\gamma}\left(\frac{n}{\gamma}q_{\mathrm{i}} + \sqrt{\psi_{\mathrm{p}} - 2q_{\mathrm{i}}}\right)\right) = \psi_{\mathrm{p}} - 2\phi_f - v_{\mathrm{ch}}, \tag{5.1}$$

where q_{i} is the normalized inversion charge density:

$$q_{\mathrm{i}} = -\frac{Q_{\mathrm{i}}}{2nC_{\mathrm{ox}}U_T}, \tag{5.2}$$

with Q_{i} being the inversion charge density, n the slope factor, C_{ox} the oxide capacitance per unit area, and $U_{\mathrm{T}} = kT/q$ the thermodynamic voltage and ψ_{p} is the normalized pinch-off surface potential:

$$\psi_{\mathrm{p}} = \frac{\Psi_{\mathrm{p}}}{U_{\mathrm{T}}}. \tag{5.3}$$

In earlier compact model approaches [106], [107] the second log term of Eq. (5.1) was neglected in the evaluation of charge density at source and drain. In BSIM6, though, Eq. (5.1) has been solved with respect to q_{i}, without using any approximations, ensuring better accuracy of the model for the entire bias range [36].

The normalized drain to source current, obtained using the well-known drift-diffusion model, is given by [39]:

$$i_{\mathrm{ds}} = \frac{I_{\mathrm{DS}}}{I_{\mathrm{spec}}} = \frac{\left(q_{\mathrm{s}}^2 + q_{\mathrm{s}}\right) - \left(q_{\mathrm{d}}^2 + q_{\mathrm{d}}\right)}{\frac{1}{2}\left(1 + \sqrt{1 + \left(\lambda_{\mathrm{c}}\left(q_{\mathrm{s}} - q_{\mathrm{d}}\right)\right)^2}\right)}, \tag{5.4}$$

where q_s and q_d are the normalized charge densities at source and drain ends, respectively. I_{spec} is the specific current defined as $I_{spec} = 2n\beta U_T^2$, where $\beta = \mu C_{ox}\frac{W}{L}$ with μ being the low-field mobility of the carriers in the channel. The denominator term in the above equation accounts for velocity saturation for short channel transistors, with:

$$\lambda_c = \frac{L_{sat}}{L_{eff}}, \tag{5.5}$$

where:

$$L_{sat} = \frac{2\mu_{eff}U_T}{v_{sat}}. \tag{5.6}$$

Note, that the drain charge density q_d is the effective charge density at drain, which is obtained using effective drain voltage accounting for V_D to V_{Dsat} transition [39].

Apart from the advantages that stem from the charge-sheet approximation adopted by BSIM6, the model inherits from BSIM4 the expressions, adequately modified, that describe the multiple physical phenomena that appear in a real, non-ideal, device e.g. short channel effects, quantum-mechanical-effects, non-uniform doping effects, gate leakage current, etc. [28]. Although some parts of the model used to describe these effects are not electrically equivalent to BSIM4, the parameter names are kept identical in order to maintain the BSIM4's user experience.

During BSIM6's development, emphasis was placed so that the model preserves DC and AC symmetry, thus, the equations regarding all the second-order physical effects were updated accordingly. BSIM6 has been tested using different quality tests for compact MOSFET models [108, 109], e.g. slope-ratio test, Gummel-symmetry test, tree-top test, AC symmetry test, harmonic balance simulation test, etc. The benchmarking demonstrated that BSIM6 satisfies all the quality tests. For example, it maintains continuity and preserves its symmetry with respect to source-drain interchange for higher order derivatives of currents and capacitances, it shows accurate slopes up to the 5^{th} harmonic in harmonic balance simulations, it provides a smooth transition from weak to strong inversion, etc. [105, 110, 111].

In BSIM6 several other improvements were made. For example the conventional junction capacitance model of BSIM4 was improved in order to ensure symmetry and continuity around $V_{BS} = 0\ V$ and $V_{BD} = 0\ V$, the nodal capacitances are derived using physical charge derivation and Ward-Dutton partitioning [112], self heating that provides additional accuracy for high-power applications is included, etc. [111]. Last but not least, BSIM6 is provided freely online. The model has been coded and released in Verilog-A facilitating its use [113], while is being implemented in major EDA simulators.

6 Geometrical Scaling in BSIM6

6.1 Introduction

Despite its multiple advantages, the downscaling of devices results also in increased short and narrow-channel effects. These effects, naturally become dominant to the corresponding geometries and they must, thus, be accurately modeled. Consequently, one of the most important challenges for compact MOSFET models is their scaling property, namely their ability to accurately represent the behavior of the real devices across the whole range of W and L and for all bias conditions, using a single model parameter set without any binning [105]. The idea of binning is based on dividing the whole device geometry plane into smaller, typically rectangular, areas called bins. The devices belonging to each bin, determined by L_{min}, L_{max}, W_{min} and W_{max} of the bin, share the same set of model parameters, called local model card. At the end, the different sets of model parameters are unified into a complete scalable model that forms a model card for which ever geometry available. Binning can be implemented for any model, however, it might result to geometrical discontinuities at the borders between the binning areas. Further, we must underline that the binning procedure can be time inefficient as it requires as many local model cards as the bins used. Thus, a global set of model parameters combined with a complete set of scaling equations is generally preferred over the binning, as a solution covering the full geometrical range of a certain process. Nevertheless, the simplicity of the binning, and its applicability to all models, despite their scaling qualifications, does give a certain credit to binning as an engineering solution to a highly complicated problem.

In this chapter, we will present which equations were developed for the geometrical scaling of the BSIM6 model and the reasoning behind this process.

6.2 Geometrical Scaling Approaches

Every physical phenomenon that impacts the behavior of the device, follows each own geometrical scaling profile, e.g. the velocity saturation effect has different scaling properties

compared to the DIBL effect. The scaling profile of each effect depends on multiple factors. One major aspect is inherently the nature and the physics behind the phenomenon itself. In addition, the scaling trend is also affected strongly by the fabrication process and the exact doping profiles and shapes that are achieved in the actual integrated circuit.

Even though a physical geometrical scaling suitable for each effect is desired, the rapid technological advancements and the limited time provided to the model to be developed does not leave enough space to the modeling engineers to meticulously follow this path. Also note that to the time needed for a model to be developed one should include both the actual development of the model and its propagation into and adoption by the designer community and the fabs. This leads to the contradictory situation where the state-of-art-technologies struggle for the development of advanced and complicated compact models, which demand so long time for their implementation that their development process outlives the technology that it is targeted for. Under the described constraints, an empirical approach was chosen for implementing the geometrical scaling of BSIM6, which was able to provide good enough results for the needs of the state-of-the-art technologies against which the model was evaluated.

A variety of different options for the geometrical scaling expressions of the model parameters are already implemented and have been evaluated for their capacity in other compact models [106, 114]. Assuming that P_{Lscaled} is the value of the model parameter after the length geometrical scaling, P is the value of the model parameter for a long/wide device and P_{L1}, P_{L2} are the length scaling model parameters, the main length scaling expressions are presented below. Note that similar expressions are valid for the width scaling as well, where all the corresponding symbols have been replaced by W instead of L. The effective channel length of the device is denoted as L_{eff} and it is always divided with 10^{-6} so that it is expressed in μm, for convenience.

▷ **Option A:**

$$P_{\text{Lscaled}} = P + \frac{P_{\text{L1}} \cdot 10^{-6}}{L_{\text{eff}}} \tag{6.1}$$

▷ **Option B:**

$$P_{\text{Lscaled}} = P + P_{\text{L1}} \cdot \exp\left(\frac{-L_{\text{eff}}}{P_{\text{L2}} \cdot 10^{-6}}\right) \tag{6.2}$$

▷ **Option C:**

$$P_{\text{Lscaled}} = P + P_{\text{L1}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}}\right)^{\text{P}_{\text{L2}}} \tag{6.3}$$

▷ **Option D:**

$$P_{\text{Lscaled}} = P \cdot \left(1 + \frac{P_{\text{L1}} \cdot 10^{-6}}{L_{\text{eff}}}\right) \tag{6.4}$$

▷ **Option E:**

$$P_{\text{Lscaled}} = P + P_{\text{L1}} \cdot \frac{10^{-6}}{L_{\text{eff}}} + P_{\text{L2}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}}\right)^2 \tag{6.5}$$

▷ **Option F:**

$$P_{\text{Lscaled}} = \frac{P}{1 + \frac{P_{\text{L1}} \cdot 10^{-6}}{L_{\text{eff}}}} \tag{6.6}$$

The above options can be divided into first and second order models and differ in the number of the scaling parameters they use, one (only P_{L1}) or two (both P_{L1} and P_{L2}), and the flexibility they offer. Among them, Option A and D provide a similar scaling profile, while Option E can approach the behavior of Option C in many cases, although Option E is generally more abrupt than Option C. Especially Option E can provide a scaling profile where two separate regions across L follow a different trend. Options B and C on the other hand offer two degrees of freedom upon the length scaling

6.2.1 Comparison of Existing Geometrical Scaling Approaches

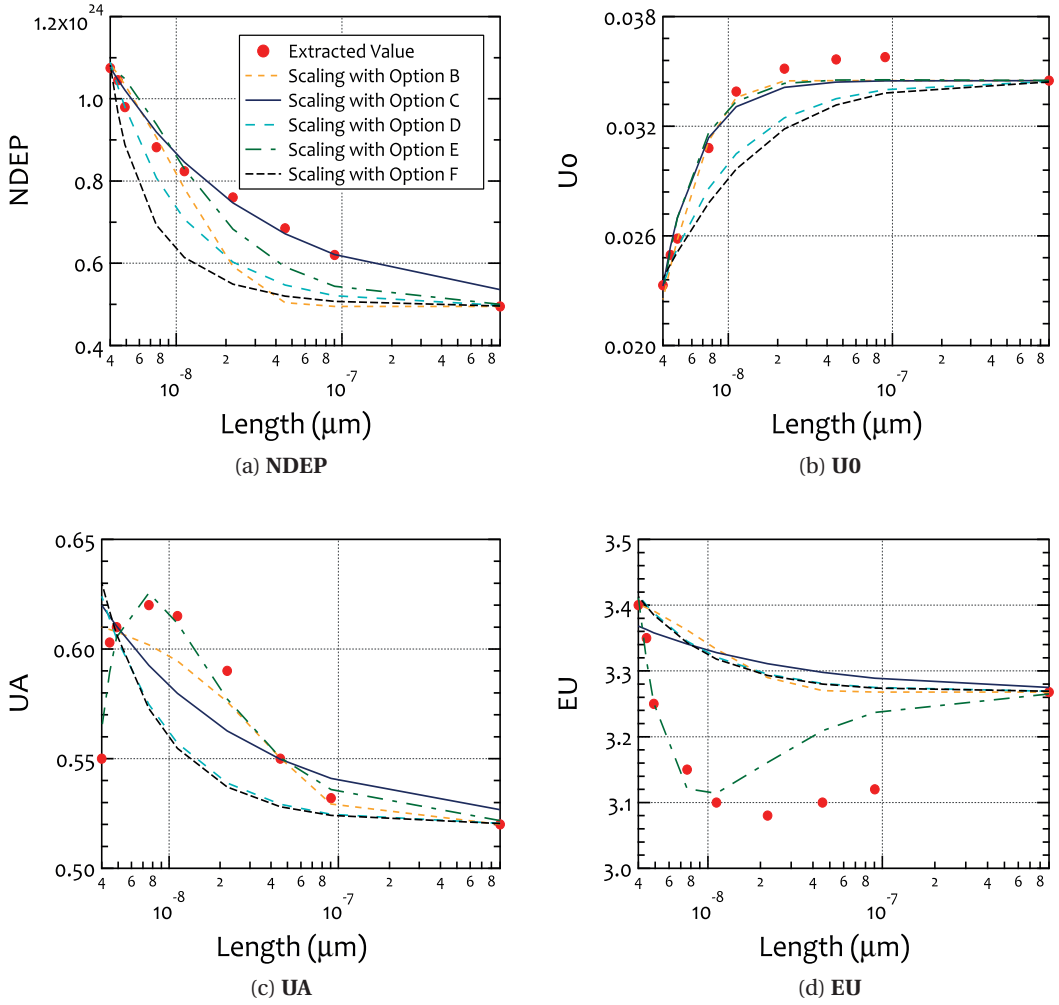
The identification of the most suitable geometrical scaling formulas is not a trivial task. First of all, the equations should demonstrate enough flexibility so that they can provide a good enough universal solution applicable to different CMOS processes from different fabs and technologies. Secondly, they should keep a useful balance over the trade-off between the flexibility, which is achieved by an increased number of parameters, and the simplicity, which requires the exactly opposite characteristic. Finally, the minimum number of the model parameters that need to be updated by the geometrical scaling should be identified and the scaling equations should be applied only to this limited subset. A crude approach where the complete set of the model parameters is affected by the scaling scheme would result into an overkill in terms of model complexity, while regarding the parameter extraction procedure it would increase dramatically the required time. Eventually, an excessive number of scalable parameters might lead to a undesired flexibility from the model, that could be able to provide similar results with different model cards.

Based on the critical points analyzed above and in order to select a suitable scaling approach, the first step was the identification of the smallest, yet sufficient, subset of parameters that should be scalable in order to provide accurate modeling for any device geometry. This is a

limited subset of the overall set of model parameters that includes the core model parameters, such as, the doping, the sub-threshold slope, the mobility, the parameters regarding the mobility reduction effect, the body factor, the velocity saturation, the series resistances, the parameters regarding the effects of CLM, DIBL, DITS and GIDL, etc. Then a dedicated local parameter extraction procedure was followed for each of the available geometries across the width/length plane, forming a group of model cards each one of which is targeted for each geometry. In our study thirty-two parameters were enough to form an adequate subset able to capture accurately the behavior of any device in all regions of CV and DC operation. In order to analyze the length scaling, devices with constant wide channel and different channel lengths, covering the whole range from a long device to the shortest ones, with a rather uniform distribution in logarithmic scale of the available length values in between, were used. This way, the influence of the narrow channel effects is avoided for this analysis as they would have no impact on the behavior of the group of the wide geometries. Similarly, for studying the width scaling, devices with constant long channel and different channel widths from narrow to wide were used. When the values of each one of the parameters were plotted across either the width or the length axis, we were able to see the scaling profile of each parameter. The last step, was to check whether the available scaling expressions were able to model correctly the behavior of the different scaling profiles of the various parameters.

In Fig. 6.1 and Fig. 6.2, we demonstrate a selection of the most characteristic length scaling profiles we encountered during the above procedure and how each scaling approach captures each profile. Note that option A is not demonstrated since it provides similar results to option D. Due to common length scaling profiles between the different parameters it was decided not to show all the 32 parameters profiles but to limit the demonstration to the most critical ones. Furthermore, the discussion here will be concentrated on the length scaling since the analysis for the width variations is more or less similar. In general, the width scaling is easier to be modeled compared to the length scaling for three reasons: (a) the minimum width of the devices does not typically scale down so agressively as the length ($W_{\text{min}} \approx 3 \sim 4\ L_{\text{min}}$), (b) the width axis of the device is perpendicular to the direction of the channel current and it, thus, influences less the electrical behavior of the device and (c) the actual doping profile and shape of the devices on the width axis is more uniform than on the length axis. The main reason of the narrow channel effects is as the edge conductance of the device which becomes prominent mostly for the narrow channel devices.

In Fig. 6.1 and Fig. 6.2, we see that when there is a certain, mostly monotonic trend throughout the L range, Option C proves to be the one that represents the scaling profiles with the best accuracy, e.g. Fig. 6.1a, Fig. 6.1b, Fig. 6.2a and Fig. 6.2b. On the other hand, when there is a strong change of the sign of the slope across the L, then the option that has the ability to approach the scaling profile closer is Option E, e.g. Fig. 6.1c, Fig. 6.1d and Fig. 6.2d,or Option B, e.g. Fig. 6.2c .



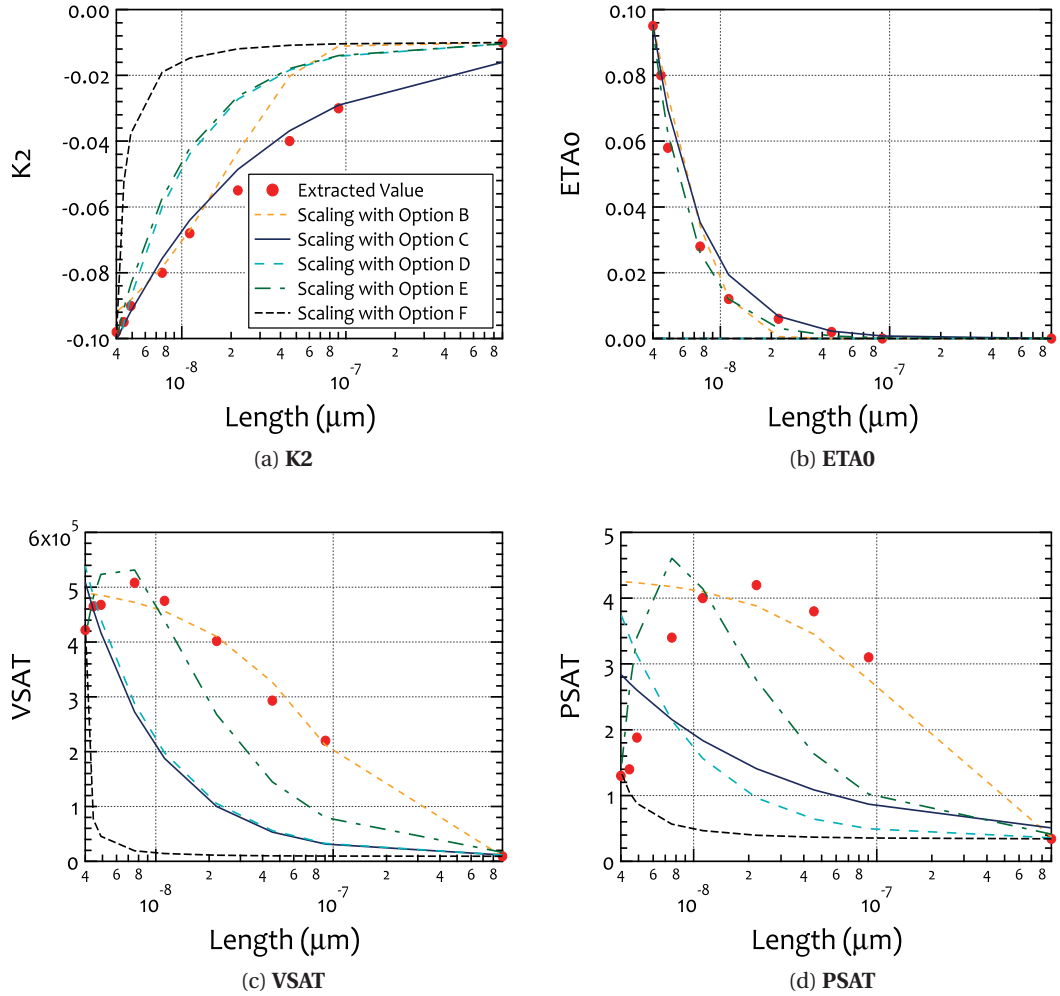
(a) **NDEP** (b) **U0**

(c) **UA** (d) **EU**

Figure 6.1 – Length scaling of the selected model parameters of nMOS DUTS of a 40 nm CMOS process - part A: (a) **NDEP** - Channel doping concentration, (b) **U0** - Low field mobility, (c) **UA** - Phonon/surface roughness scattering parameter and (d) **EU** - Phonon/surface roughness scattering parameter.

6.3 Selection of Geometrical Scaling Equations for BSIM6

In Fig. 6.1 and Fig. 6.2, the abilities of the different existing geometry scaling expressions on different scaling profiles were demonstrated. And although, in certain cases the available scaling options were able to capture accurately the profiles, there were other cases where none of the scaling formulas under evaluation was accurate enough, e.g. Fig. 6.1d , Fig. 6.2c and Fig. 6.2d, where even Option E that can approach the results better than the rest is not close enough to either the medium or the short devices. Furthermore, in many cases after extracting the scaling parameters, even though the scaling expression fits the short and medium channel length devices it does not fit accurately the long channel one, e.g. Fig. 6.1a, Fig. 6.2a. This is a highly undesirable characteristic since the scaling extraction might lead to loops where



(a) **K2** (b) **ETA0**

(c) **VSAT** (d) **PSAT**

Figure 6.2 – Length scaling of the selected model parameters of nMOS DUTS of a 40 nm CMOS process - part B: (a) **K2** - V_{TH} shift due to non-uniform vertical doping, (b) **ETA0** - DIBL coefficient, (c) **VSAT** - Saturation velocity and (d) **PSAT** - Velocity saturation exponent.

the wide and long channel extraction must be repeated. Finally, we would ideally want to be able to generalize the model and prime it with the flexibility to offer various scaling profiles with a single geometrical scaling expression. Subsequently, this would result into extending the model capabilities to cover more easily and accurately different CMOS technologies and processes. Taking into account the above considerations, we decided to restructure the scaling options that offer the most flexibility so that they meet our specified demands. So, Option C and Option E were changed accordingly to BSIM6_{L1} and BSIM6_{L2}, as shown below.

▷ **BSIM6$_{L1}$:**

$$P_{\text{Lscaled}} = P \cdot \left[1 + P_{\text{L1}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}} \right)^{P_{\text{L2}}} - P_{\text{L1}} \cdot \left(\frac{10^{-6}}{L_{\text{eff,long}}} \right)^{P_{\text{L2}}} \right] \tag{6.7}$$

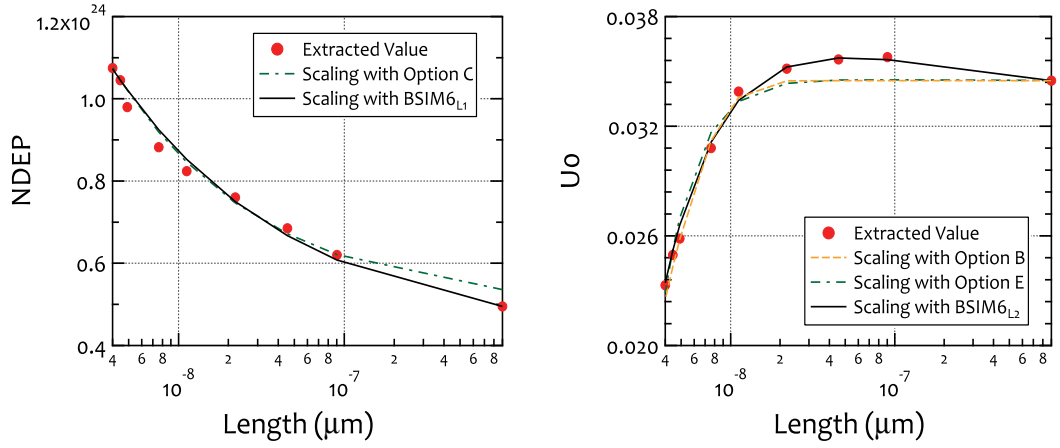
▷ **BSIM6$_{L2}$:**

$$P_{\text{Lscaled}} = P \cdot \left[1 + P_{\text{L1}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}} \right)^{P_{\text{L3}}} + P_{\text{L2}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}} \right)^{P_{\text{L4}}} - P_{\text{L1}} \cdot \left(\frac{10^{-6}}{L_{\text{eff,long}}} \right)^{P_{\text{L3}}} - P_{\text{L2}} \cdot \left(\frac{10^{-6}}{L_{\text{eff,long}}} \right)^{P_{\text{L4}}} \right] \tag{6.8}$$

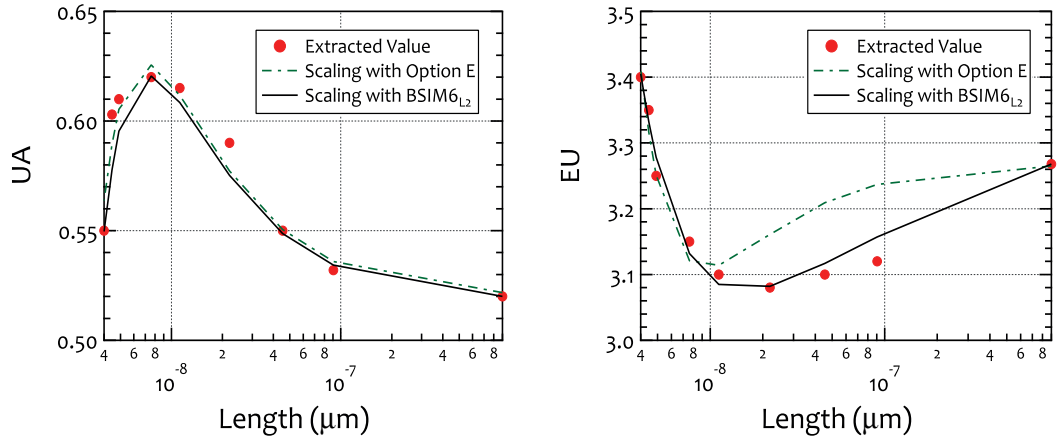
In Eq. (6.7) and Eq. (6.8), by subtracting the scaling term of the longest channel device, we ensure that the already extracted parameters of the long channel DUT will be kept unchanged after the scaling analysis, regardless of the values of the extracted scaling parameters. In this way the unnecessary refinement within loops that are in general required in such procedures is mitigated, since the scaling step does not influence at all the previous long and wide channel extraction. Additionally, if we compare Eq. (6.5) and Eq. (6.8), we see that in BSIM6$_{L2}$ the exponents of the terms are no longer constant but instead new parameters have been introduced. This of course doubles the number of the scaling parameters but at the same time it expands the applicability of the formula so that its usage is justified despite the additional computational cost and the increment in the size of the model card.

In Fig. 6.3 and Fig. 6.4, we compare the best of the already existing scaling options with the new scaling expressions (BSIM6$_{L1}$ and BSIM6$_{L2}$) based on the same scaling profiles as before. From the plots we see that BSIM6$_{L1}$ not only keeps the value of the long channel device constant but also improves the fitting of the overall scaling profile, e.g. Fig. 6.1a and Fig. 6.2a. Regarding the plots where two different trends are observed, we see that BSIM6$_{L2}$ can offer the same accuracy or even better than Option B or Option E, e.g. Fig. 6.1b, Fig. 6.1d, Fig. 6.2c and Fig. 6.2d.

As already mentioned above, the BSIM6$_{L2}$ expression might offer significantly higher flexibility but it also involves four length scaling parameters per core model parameter, which can be a major drawback if we consider (a) that the scaling equations are implemented for a few decades of the parameters and (b) that width scaling parameters and scaling parameters for short/narrow channel devices should be added as well. As a result, a compromise between accuracy and flexibility should be made and the modeling engineer should decide which of the two new equations will be used for which parameters. In an effort to keep the number of the model parameters and the complexity of the model as low as possible, maintaining though a good accuracy, BSIM6$_{L1}$ was chosen as the main scaling expression for BSIM6 model, while BSIM6$_{L2}$ was used only for the length scaling of the channel doping **NDEP**. It has to be noted here that **NDEP** is possibly the most critical parameter that also presents the higher variability with respect to the geometry in advanced CMOS technologies. By applying the same principles of the length scaling to the width scaling and to the scaling for combined short and narrow channel devices, also called small geometries, we end up with the equations



(a) **NDEP** (b) **U0**



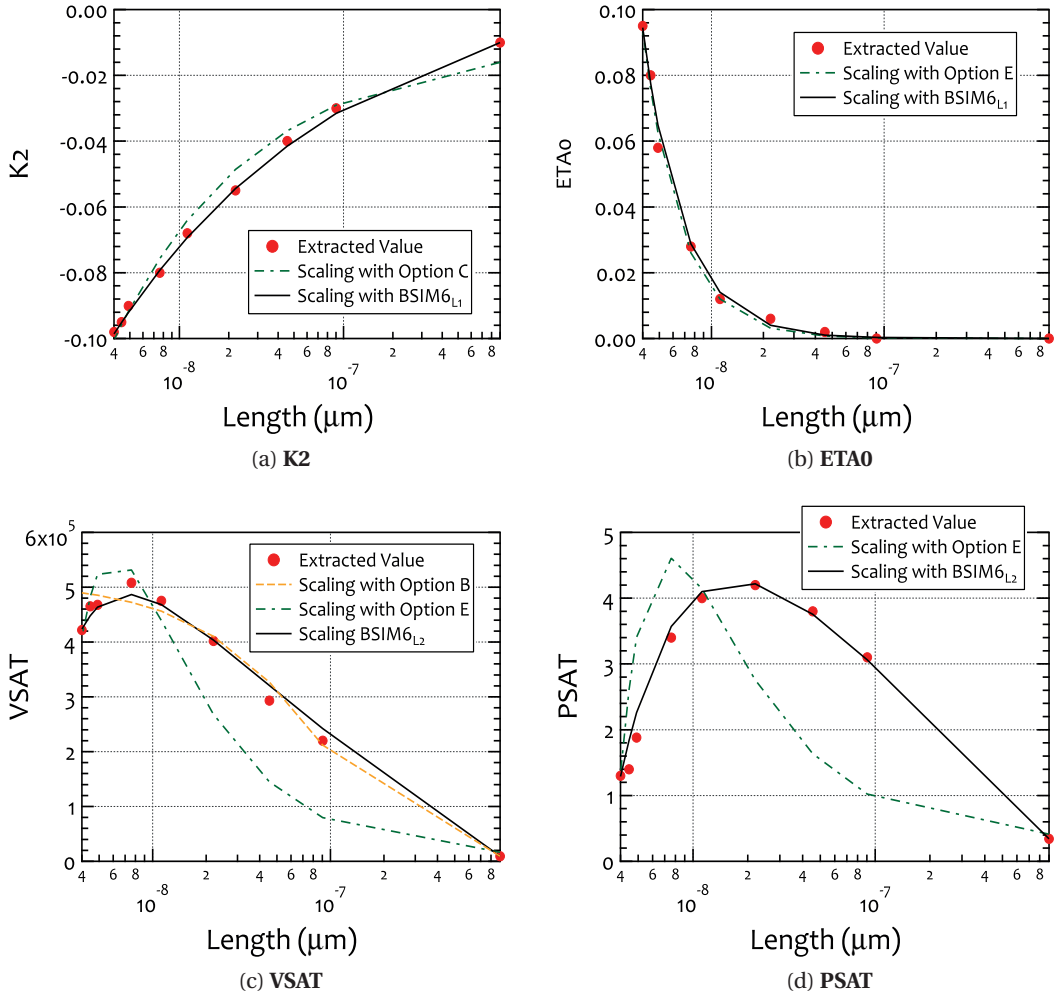
(c) **UA** (d) **EU**

Figure 6.3 – Comparison of the new BSIM6 length scaling expressions with the best of the already existing ones for the selected model parameters of nMOS DUTS of a 40 nm CMOS process - part A: (a) **NDEP** - Channel doping concentration, (b) **U0** - Low field mobility, (c) **UA** - Phonon/surface roughness scattering parameter and (d) **EU** - Phonon/surface roughness scattering parameter.

for the final scaled value of a parameter P_{scaled} as:

$$\begin{aligned} P_{\text{scaled}} &= P \cdot \left[1 + P_{\text{L}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}} \right)^{\text{P}_{\text{Lexp}}} - P_{\text{L}} \cdot \left(\frac{10^{-6}}{L_{\text{eff,long}}} \right)^{\text{P}_{\text{Lexp}}} \right. \\ &\left. + P_{\text{W}} \cdot \left(\frac{10^{-6}}{W_{\text{eff}}} \right)^{\text{P}_{\text{Wexp}}} - P_{\text{W}} \cdot \left(\frac{10^{-6}}{W_{\text{eff,wide}}} \right)^{\text{P}_{\text{Wexp}}} + P_{\text{WL}} \cdot \left(\frac{10^{-12}}{W_{\text{eff}} \cdot L_{\text{eff}}} \right)^{\text{P}_{\text{WLexp}}} \right] \end{aligned} \tag{6.9}$$

In Eq. (6.9), for the short/narrow scaling there is no need to subtract the scaling for the large device, because usually the scaling parameters for short/narrow channel devices have a negligible impact on the wide/long geometry. Especially for the doping, the final value for the scaled parameter, taking into account the formula BSIM6L2 which contains two different



(a) **K2** (b) **ETA0**

(c) **VSAT** (d) **PSAT**

Figure 6.4 – Comparison of the new BSIM6 length scaling expressions with the best of the already existing ones for the selected model parameters of nMOS DUTS of a 40 nm CMOS process - part B: (a) **K2** - V_{TH} shift due to non-uniform vertical doping, (b) **ETA0** - DIBL coefficient, (c) **VSAT** - Saturation velocity and (d) **PSAT** - Velocity saturation exponent.

critical lengths for the scaling of the parameter, is given by the expression:

$$\begin{aligned} NDEP_{\text{scaled}} &= NDEP \cdot \left[1 + NDEP_{\text{L1}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}} \right)^{\text{NDEP}_{\text{Lexp1}}} - NDEP_{\text{L1}} \cdot \left(\frac{10^{-6}}{L_{\text{eff,long}}} \right)^{\text{NDEP}_{\text{Lexp1}}} \right. \\ &+ NDEP_{\text{L2}} \cdot \left(\frac{10^{-6}}{L_{\text{eff}}} \right)^{\text{NDEP}_{\text{Lexp2}}} - NDEP_{\text{L2}} \cdot \left(\frac{10^{-6}}{L_{\text{eff,long}}} \right)^{\text{NDEP}_{\text{Lexp2}}} \\ &+ NDEP_{\text{W}} \cdot \left(\frac{10^{-6}}{W_{\text{eff}}} \right)^{\text{NDEP}_{\text{Wexp}}} - NDEP_{\text{W}} \cdot \left(\frac{10^{-6}}{W_{\text{eff,wide}}} \right)^{\text{NDEP}_{\text{Wexp}}} \\ &\left. + NDEP_{\text{WL}} \cdot \left(\frac{10^{-12}}{W_{\text{eff}} \cdot L_{\text{eff}}} \right)^{\text{NDEP}_{\text{WLexp}}} \right]. \end{aligned} \tag{6.10}$$

6.4 Evaluation of the BSIM6 Scalability

After implementing the new scaling equations in BSIM6, the validation of the model's ability to accurately represent the behavior of nanoscale bulk MOSFETs in the whole geometry plane of a technology follows. The first step before the evaluation of the model was the extraction of two global model cards (different for nMOS and pMOS devices), that could be used: a) across the geometry plane (W/L), b) for all bias conditions; from weak- to strong-inversion and from linear operation to saturation and c) for CV and IV operation. In order to extract the global model cards, the parameter extraction procedure described in details in Ch. 7 was used. The model is then compared against measurements of the state-of-the-art 40 nm bulk CMOS technology, over a wide range of biases and geometries, for nMOS and pMOS devices, and for static (IV) operation. In the plots that follow only the length scaling abilities of the model are shown. Due to the fact that the W of the devices does not scale as much as the L, the width scaling properties of a process can be modeled easier and thus it is more critical to demonstrate the length scaling abilities of the model. Further results are presented in Ch. 8.

All the different quantities that are displayed throughout this section are normalized according to the relations presented in Table 6.1, while a more in depth analysis regarding the normalization procedure can be found in [39]. The different threshold voltages are calculated using the constant current method proposed in [115]. According to this method, the current at which the threshold voltage is calculated is $I_{TH} = 0.4804 \cdot I_{spec}$ when $V_D = 0.05\ V$ (linear mode) and $I_{TH} = 0.608 \cdot I_{spec}$ when $V_D = 1.1\ V$ (saturation) at $T = 25$ °C, where I_{spec} denotes the specific current. For the validation of the model in IV operation measurements of DC DUTs were used.

Table 6.1 – Description of the normalization process

Quantity	Normalization Factor	Normalized Quantity
I_{DS}	$I_{spec} = 2n\beta U_T^2$	$i_d = I_{DS}/I_{spec}$
G_m	$G_{spec} = I_{spec}/U_T$	$g_m = G_m/G_{spec}$
G_{m2}	$G_{spec2} = I_{spec}/U_T^2$	$g_{m2} = G_{m2}/G_{spec2}$
G_{m3}	$G_{spec3} = I_{spec}/U_T^3$	$g_{m3} = G_{m3}/G_{spec3}$
G_{ds}	$G_{spec} = I_{spec}/U_T$	$g_{ds} = G_{ds}/G_{spec}$

where:
$\beta = \mu C_{ox} W/L$ the transfer parameter,
n the slope factor, C_{ox} the oxide capacitance per unit area,
μ the mobility of the carriers and
$U_T = kT/q$ the thermodynamic voltage

In Fig. 6.5, Fig. 6.6 and Fig. 6.7 the characteristics for the fundamental DC analyses of both nMOS and pMOS devices and for different channel lengths, varying from 40 nm to 10 μm, are

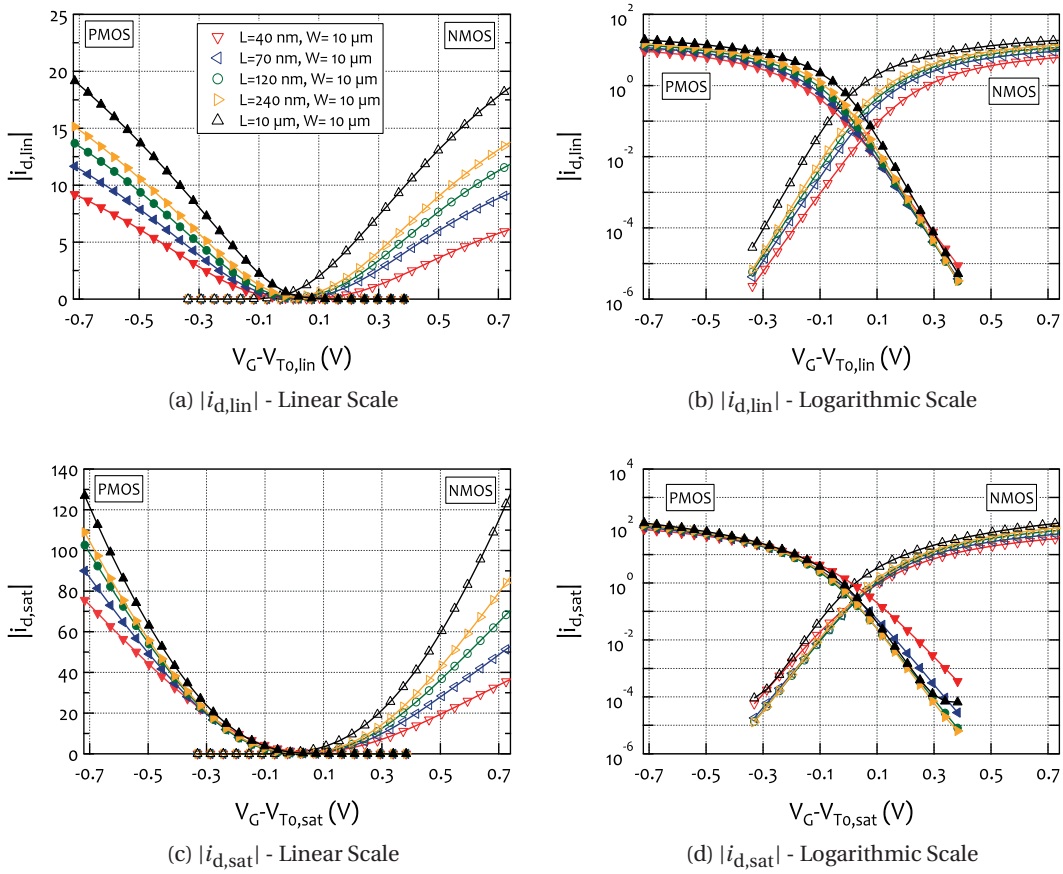
shown. Fig. 6.5 presents the normalized drain current $|i_d|$ vs. $V_G - V_{T0}$, for linear operation ($V_D = 0.05\ V$) and saturation ($V_D = 1.1\ V$), with V_{T0} being the threshold voltage of the long channel device in the equivalent region of operation and type of device when $V_{SB} = 0V$. The normalized transconductance namely, g_m is shown in Fig. 6.6a and Fig. 6.6b in linear mode and saturation. For further demonstrating the excellent scaling abilities of the model, the 2^{nd} and 3^{rd} derivative of the current, namely g_{m2} and g_{m3}, are shown in Fig. 6.6c/Fig. 6.6d and Fig. 6.6e/Fig. 6.6f respectively. Being higher order derivatives they are more sensitive to any inaccuracies of the model which would be visible immediately. In Fig. 6.7 the $|i_d|$ and the normalized output conductance g_{ds} vs. V_D for different V_G biases that cover from weak- to strong-inversion, is presented. The model is able to give an exact representation of the behavior of the nanoscale MOS transistor despite the fact that is highly influenced by complex higher order physical effects.

It should be noted that in the plots the normalized current reduces with the reduction of L. This is due to the short channel effects (SCEs), e.g. velocity saturation, RSCE etc., that strongly impact the behavior of nanoscale devices. Without the SCEs the characteristics of the normalized drain current for the devices with different channel lengths should overlap. However, the SCEs cause a reduction of the mobility with the reduction of L so, the normalized drain current of the shorter devices is smaller compared to that of the longer ones. This can be explained better using the following equation:

$$i_d = \frac{I_{DS}}{I_{spec}} = \frac{I_{DS}}{I_{spec\square} \cdot \frac{W}{L}} \propto \frac{\mu(L) \cdot \frac{W}{L}}{I_{spec\square} \cdot \frac{W}{L}} = \frac{\mu(L)}{I_{spec\square}}, \tag{6.11}$$

where the $I_{spec\square}$ is extracted from the long/wide channel device. In (6.11), it is shown that the normalized drain current i_d is proportional to the length dependent mobility $\mu(L)$. So, since SCEs cause a reduction of $\mu(L)$ with the reduction of L, i_d will also reduce for shorter channel devices.

Targeting to demonstrate further the abilities of the BSIM6 model, the length scaling plots of the sub-threshold slope SS ($SS = \partial V_{GS} / \partial log(I_D)$), the normalized threshold voltage υ_{tb} ($\upsilon_{tb} = V_{TB} / V_{T0,long}$), with $V_{T0,long}$ being the threshold voltage of the long channel device in the defined region of operation when $V_{SB} = 0\ V$ and the maximum normalized current $\max(i_d)$, are presented in Fig. 6.8. These characteristics describe the scaling properties of both nMOS and pMOS devices, in linear mode and saturation, for different V_{SB} biases. It can be observed that the model displays a very good scalability across L, despite the fact that a single model card (without binning) is used for all the simulations.



(a) $|i_{d,lin}|$ - Linear Scale

(b) $|i_{d,lin}|$ - Logarithmic Scale

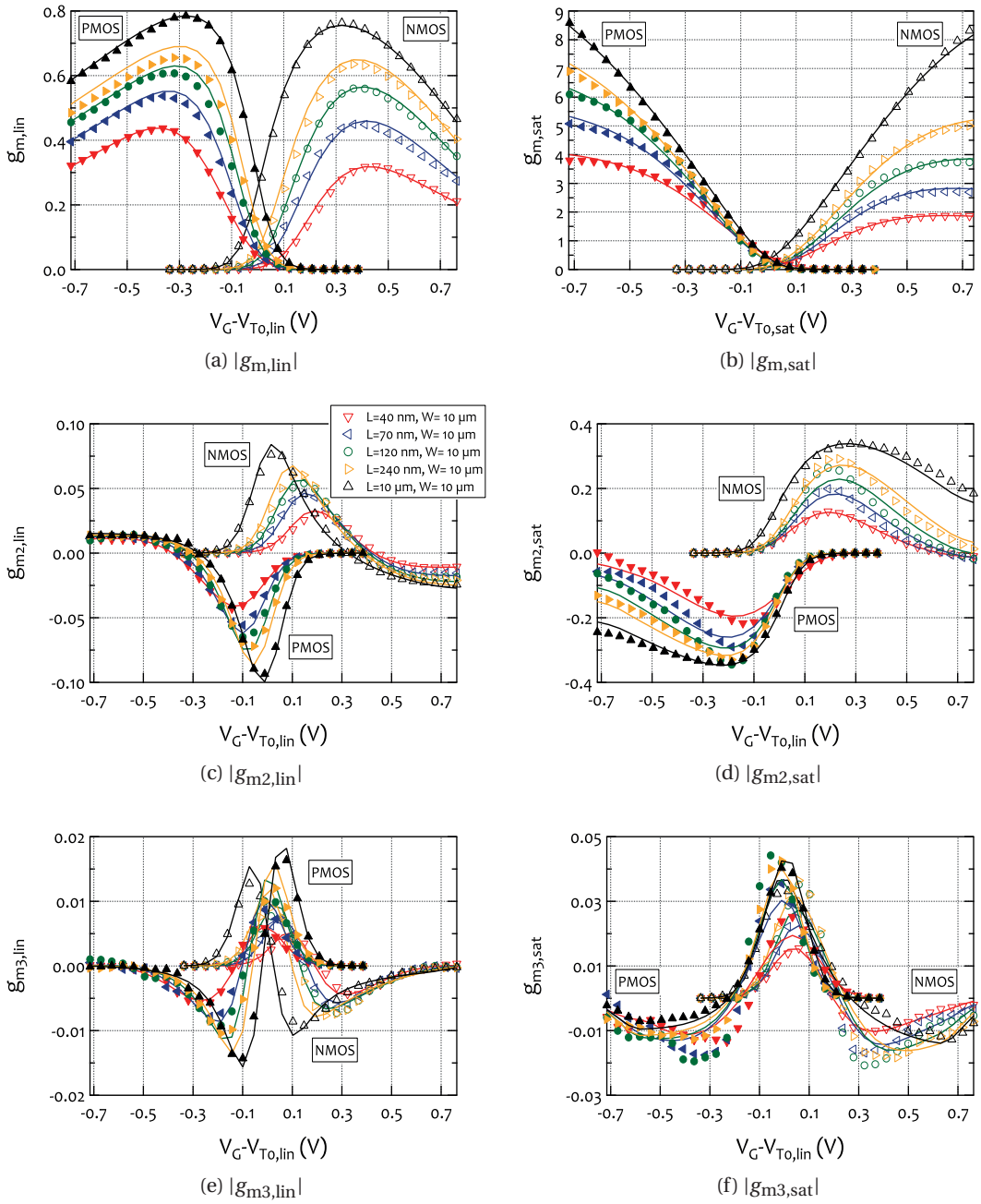
(c) $|i_{d,sat}|$ - Linear Scale

(d) $|i_{d,sat}|$ - Logarithmic Scale

Figure 6.5 – Normalized drain current $|i_d|$ vs. $V_G - V_{T0}$ in linear and logarithmic scale, for nMOS and pMOS DC DUTs, with $W = 10$ µm and $L = [40n, 70n, 120n, 240n, 10\mu]$ m, of a 40 nm CMOS process. (a) $i_{d,lin}$ vs. $V_G - V_{T0,lin}$ at $V_D = 0.05\ V$ (linear mode), in linear scale, (b) $i_{d,lin}$ vs. $V_G - V_{T0,lin}$ at $V_D = 0.05\ V$ (linear mode), in logarithmic scale, (c) $i_{d,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.1\ V$ (saturation), in linear scale, (d) $i_{d,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.1\ V$ (saturation), in logarithmic scale,

6.5 Conclusions

The outcome of the comparison of BSIM6 against measurements of an advanced CMOS process attests the excellent scaling abilities of the BSIM6 model and justifies the choice and the development of the new global scaling scheme. The model shows a smooth and continuous behavior, up to higher derivatives, in the prediction of the MOSFET performance. Furthermore, its accuracy is importantly boosted with respect to its previous version that included only binning equations, capturing, now, correctly all the scaling properties of one type of transistor using a single model card. This task has proved to be fairly challenging, especially in modern technologies where the device channel lengths extend over almost three decades. However, the systematic and careful way that this task was handled, allowed the optimization of the results, adding significant value to the compact model.

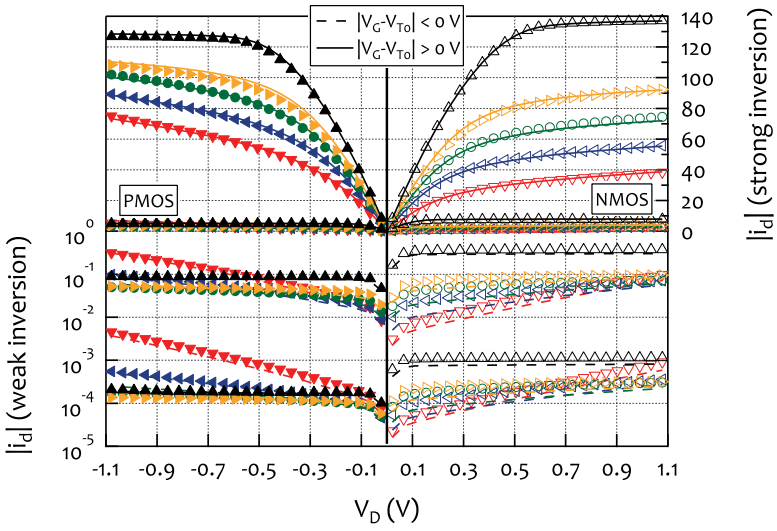


(a) $|g_{m,lin}|$ (b) $|g_{m,sat}|$

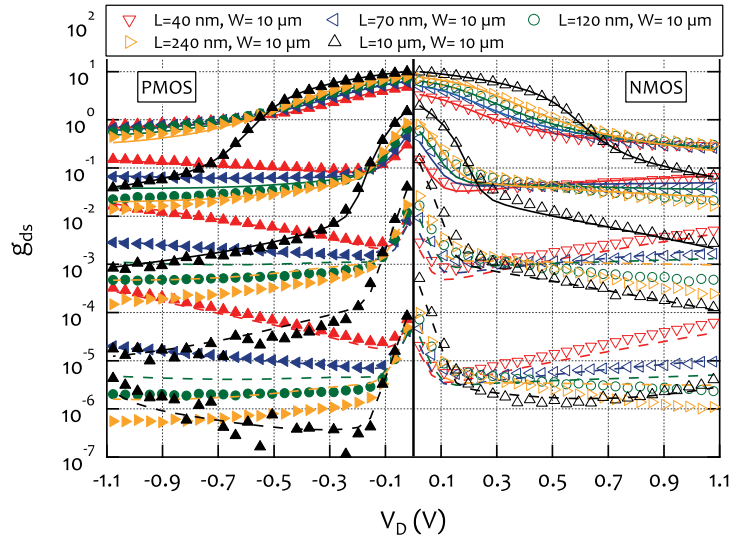
(c) $|g_{m2,lin}|$ (d) $|g_{m2,sat}|$

(e) $|g_{m3,lin}|$ (f) $|g_{m3,sat}|$

Figure 6.6 – Normalized drain current derivatives $|g_m|$, $|g_{m2}|$, $|g_{m3}|$ vs. $V_G - V_{T0}$, for nMOS and pMOS DC DUTs, with $W = 10$ µm and $L = [40n, 70n, 120n, 240n, 10\mu]$ m, of a 40 nm CMOS process. The V_D bias for linear mode is $V_D = 0.05$ V and for saturation $V_D = 1.1$ V. (a) $|g_{m,lin}|$, (b) $|g_{m,sat}|$, (c) $|g_{m2,lin}|$, (d) $|g_{m2,sat}|$, (e) $|g_{m3,lin}|$ and (f) $|g_{m3,lin}|$.

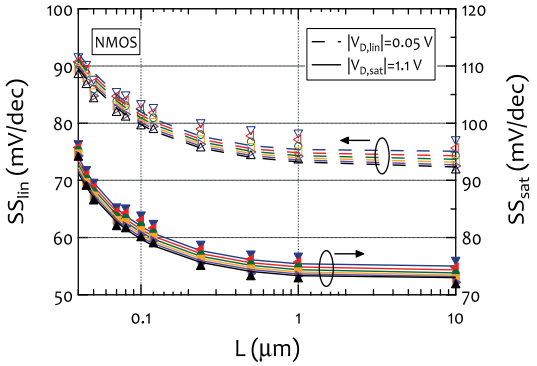


(a) i_d

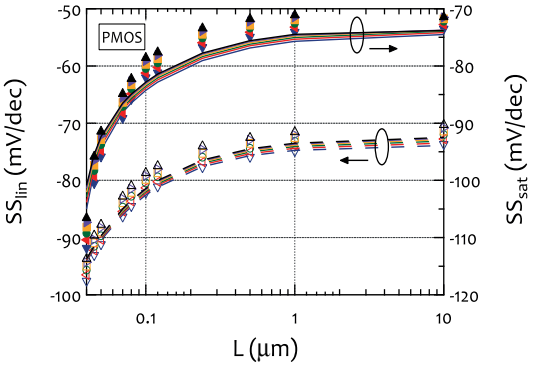


(b) g_{ds}

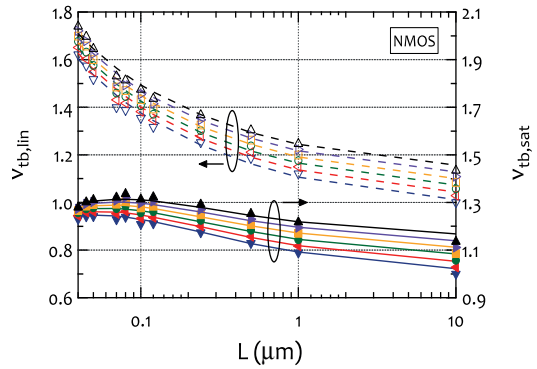
Figure 6.7 – Normalized drain current $|i_d|$ and its derivative g_{ds} vs. V_D for $V_G - V_{T0} = [-0.26, -0.07]$ V (weak inversion) and $V_G - V_{T0} = [0.12, 0.72]$ V (strong inversion) at $V_S = V_D = V_B = 0$ V, for nMOS and pMOS DC DUTs, with $W = 10$ µm and $L = [40n, 70n, 120n, 240n, 10\mu]$ m, of a 40 nm CMOS process. (a) $|i_d|$ and (b) g_{ds}.



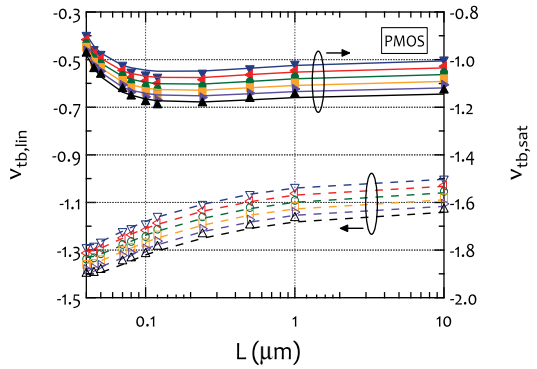
(a) SS - nMOS



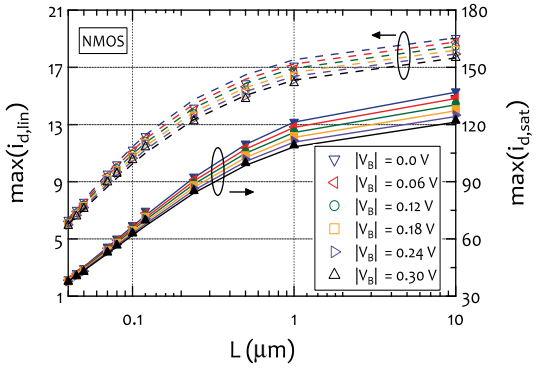
(b) SS - pMOS



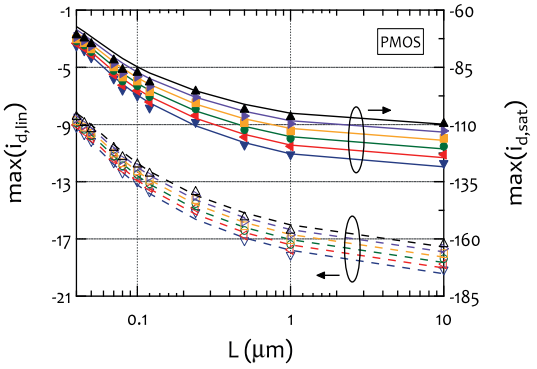
(c) v_{tb} - nMOS



(d) v_{tb} - pMOS



(e) max(i_d) - nMOS



(f) max(i_d) - pMOS

Figure 6.8 – Length Scaling of: (a), (b) SS, (c), (d) v_{tb}, (e), (f) max(i_d) for nMOS and pMOS DC DUTs of 40 nm CMOS process, at $V_D = 0.05\ V$ (linear mode) and $V_D = 1.1\ V$ (saturation) and for $|V_{SB}| = [0.0, 0.06, 0.12, 0.18, 0.24, 0.3]\ V$.

7 Parameter Extraction Methodology of BSIM6 for CMOS Technologies

7.1 Introduction

The development of the compact models of electron devices is based on the analysis of their physical behavior. The physics that lay inside the core of the device dictate the structure of the model and its formulation. However, the complexity of the real devices force the modeling engineers to proceed into a series of simplifications and approximations. These steps enable the development of a compact model which is appropriate for circuit level simulations [116]. On this direction, it is proved to be necessary to add certain empirical elements to physics based formulas in order to cover higher order effects where the physical approach is overcomplicated for the specifications of a compact model, resulting in a set of model parameters. These parameters are not strictly defined by the technology but their values are extracted against each different technology under the main criterion of modulating the behavior of the model towards the actual and measured performance of the fabricated devices of this technology. The set of values used in order to describe a certain technology is also refereed to as the "model card" of this technology.

The objective of this chapter is to provide the main guidelines for the extraction of the main model parameters of the BSIM6 MOSFET model in order to describe a certain technology. The procedure is structured in such a way that parameters linked to specific psychical phenomena are extracted from analyses where these effects are prominent. Although the parameter extraction is not always a straight-forward procedure, the aim is to minimize the effort invested and the number of the performed loops and iterations.

If all the steps of the described procedure are followed then a single model-card is obtained that can be globally used for all the devices of the technology regardless of their geometry. This means that the model can be used across the entire width/length plane of the technology, although it might not display equal accuracy for all DUTs. If a local fitting is needed, then only the parameters of Section 7.2.1 need to be extracted for each targeted DUT and a more accurate fitting is achieved. However, in that case, a higher level binning solution needs to be

found if the model card is to be used for the entire geometry range of the technology. A trade-off can be underlined at this point between a global, full-geometry study and a local extraction. The study of a wide range of corner geometries allows the more accurate investigation of the combination and the correlation mechanisms between the various phenomena which prevail in different geometries. On the other hand, the local extraction optimizes the flexibility that can be obtained by the model on a single DUT. However, if the local extractions are binned together to a full geometry structure, this comes at the expense of an effort at a higher level, together with possible discontinuities issues at the edges of the bins. Irrespectively of the choice between global and local fitting, different model cards should be extracted for nMOS and pMOS devices or for different technologies.

The guidelines reported here are based on the BSIM6 model and its model parameters and are, now, a part of the official BSIM6 technical manual [117]. Nevertheless, the applicability of the steps shown below has strong validity for any MOSFET compact model. The exact names of the parameters might differ between different models, yet the connection between the phenomena and effects with which the parameters are related and the analysis based on which they are extracted, as well as the order that is followed, is more connected with the nature and the characteristics of the MOSFET itself, rather than a specific compact model.

Note that in the following section the names of the parameters of the BSIM6 model are written in bold and capital letters so that it is easier for the reader to distinguish them.

7.2 Parameter Extraction Methodology

As the first step and before proceeding to the extraction of any parameter, it is important that the **TNOM** parameter is set to the value of the temperature at which the available or reference measurements were carried out. Further, it is recommended that, if certain process related parameters are available, their values are provided into the model card. The most common process parameters are shown in Table 7.1.

Parameter Name	Physical Description
EPSROX *	Relative Gate Dielectric Constant
EPSRSUB	Relative Dielectric Constant of the Channel
TOXE *	Electrical Gate Equivalent Oxide Thickness
TOXP or **DTOX**	Physical Gate Equivalent Oxide Thickness
NDEP *	Channel Doping Concentration
NGATE	Gate Doping Concentration
NSD	S/D Doping Concentration
XJ *	S/D Jucntion Depth
XW/**XL** *	Channel W/L Offset due to Mask/Etch Effect

Table 7.1 – Process parameters which are recommended to be provided before starting the parameter extraction procedure with BSIM6. Parameters that are followed by an asterisk (*) should be considered as the most important among them.

7.2.1 Extraction of Main Physical Effects & Geometry Independent Parameters

The first part of the model parameter extraction procedure is to extract the parameters that are related to the main physical phenomena, which define transistor's behavior and are geometry independent. To this direction, a wide and long channel device should be studied. At this point, **WWIDE** and **LLONG** parameters must be assigned to the values of the width and length of this DUT of large dimensions. This step ensures that once the behavior of the long/wide channel device is fitted, it will not be affected afterwards by the values of the scaling parameters that will be extracted in the following steps.

C_{GG} vs. V_G Analysis, $V_S = 0\ V$, $V_D = 0\ V$ & $V_B = 0\ V$

At this first step the process parameters and the parameters related to the Quantum Mechanical effect are extracted. Even if the values of these parameters have been already assigned from information directly related to the fabrication process itself, a fine tuning of them should be made in order to fit more accurately the electrical behavior of the model to the actual capacitive measurements of the device.

From the C_{GG} vs. V_G analysis the following process parameters can be extracted: **NDEP**, **TOXE**, **VFB** and **NGATE**. Each one of these parameters affects a different region of or in a different way the C_{GG} capacitance vs. V_G, so they should be extracted accordingly. More specifically:

- **NDEP** is affecting the C_{GG} in the depletion region. If possible, **NDEP**, which defines the doping level, is ideally extracted from the C_{GB} vs. V_G analysis (with the S and D terminals grounded).

- **TOXE** is affecting the deep accumulation and strong-inversion regions, where the maximum capacitive load is seen from the gate node towards the rest of the device.

- **VFB** is defining the flat-band voltage of the device and it can be extracted by studying the region from depletion till the onset of strong-inversion. If the threshold voltage is used by the model as a parameter, then the threshold voltage can be extracted instead or in parallel, depending of the model parameter set.

- **NGATE** is related to the poly-silicon depletion effect, so it affects the slope of C_{GG} in the strong-inversion region.

Furthermore, the value of C_{OX} is affected by the Quantum Mechanical effect. The parameters: **ADOS**, **BDOS**, **QM0** and **ETAQM** are also extracted from the C_{GG} vs. V_G analysis, when focusing at the slope of C_{GG} at the onset of the strong-inversion region.

I_D vs. V_G Analysis, $V_D = [V_{D,lin}, V_{D,sat}]$, $V_S = 0\ V$ & $V_B = 0\ V$

In this step, the V_G dependence of the drain current (I_D), is studied. Different parameters are extracted in two different regions of operation, namely the *linear mode* (i.e. $V_D \ll V_G - V_{TH}$) and the *saturation* (i.e. $V_D \gg V_G - V_{TH}$). It is very important that during the extraction in this step, both I_D and the transconductance (g_m) are studied at once.

Linear Mode

- Focusing in the *weak-inversion* region (I_D vs. V_G curve with y-axis in logarithmic scale), **NFACTOR**, which is related to the sub-threshold slope of the I_D, can be extracted. Furthermore, a fine tuning of the **NDEP** and **VFB** parameters may be performed, if needed. In case the values of **NDEP** and **VFB** obtained during the fitting of I_D vs. V_G characteristic differ importantly from those obtained during the fitting of the C_{GG} vs. V_G characteristic before (Section 7.2.1), then the parameters **NDEPCV** and **VFBCV** can be used for the dynamic operation (CV) and the **NDEP** and **VFB** for the static current operation (IV). In general, though, using different values for **NDEP** and **NDEPCV** for IV and CV operation is not recommended as it introduces an artificial discrepancy of the model behavior between the two regimes of operation.

- From the *strong-inversion* region, the low-field mobility **U0**, the parameter **ETAMOB** for the calculation of the vertical effective field, the parameters related to the effect of mobility reduction due to vertical field **UA** and **EU** and the parameters for the coulomb scattering effect **UD** and **UCS**, are extracted. Furthermore, the parameters for S/D series resistances are also extracted under the same bias conditions. If **RDSMOD = 0** (internal S/D series resistances), **RDSW** is extracted. Otherwise, **RSW** and **RDW** are extracted.

Saturation

- From the *weak-inversion* region (I_D vs. V_G curve with y-axis in logarithmic scale), the **CDSCD** parameter, which is linked to the dependence of the sub-threshold slope on drain bias, is extracted.

- Focusing in the *strong-inversion* region, the parameters that are connected to the velocity saturation effect, namely **VSAT**, **PSAT**, **PTWG** and **PSATX**, can be extracted.

Finally, from the *accumulation* to the *depletion* region, in both *linear mode* and *saturation*, the parameters related to the GIDL effect are extracted. First, the selector **GIDLMOD** should be set to 1, in order to activate GIDL/GILS currents, and then the parameters **AGIDL**, **BGIDL**, **CGIDL** and **EGIDL** are extracted. In principle, and depending on the fabrication process and the layout of the device, GIDL and GILS currents should be equal, so it is sufficient to extract

the **AGIDL**, **BGIDL**, **CGIDL** and **EGIDL** parameters. But in case GIDL and GISL currents differ, then parameters **AGISL**, **BGISL**, **CGISL** and **EGISL** can also be used and should be extracted separately from corresponding source current measurements.

I_G vs. V_G Analysis, various V_D, V_S = 0 *V* & V_B = 0 *V*

From the I_G vs. V_G analysis the parameters related to the gate current can be extracted. First, the gate oxide tunneling components should be activated by setting to 1 the selectors **IGCMOD** and **IGBMOD**. Different parameters are extracted in different regions of operation. More specifically the following points can be drawn:

Accumulation to Weak-inversion Region

- The **AIGBACC**, **BIGBACC**, **CIGBACC** and **NIGBACC**, which are linked to the gate-to-substrate current component determined by electron tunneling from conduction band (ECB).

- The **AIGS**, **BIGS** and **CIGS**, which are linked to the tunneling current between the gate and the source diffusion region and the **AIGD**, **BIGD** and **CIGD**, which are linked to the tunneling current between the gate and the drain diffusion region.

- The **DLCIG** and **DLCIGD**, which are linked to the S/D overlap length for the I_{GS} and the I_{GD}, respectively.

Weak- to Strong-inversion Region

- **AIGBINV**, **BIGBINV**, **CIGBINV**, **EIGBINV** and **NIGBINV**, which are linked to the gate-to-substrate current component determined by EVB.

- **AIGC**, **BIGC**, **CIGC**, **NIGC** and **PIGCD**, which are linked to the gate-to-channel current. **PIGCD** is expressing the V_D dependence of gate-to-channel current.

I_D vs. V_D Analysis at various V_G, V_S = 0 *V* & V_B = 0 *V*

In this step, both the I_D vs. V_D and the output conductance (g_{ds}) vs. V_D characteristics are studied simultaneously. Different effects impact both characteristics and the parameters related to those effects are extracted. In detail the following parameters are extracted, based on the corresponding notes:

- **DELTA**, which is a smoothing factor for the transition between V_{DS} and $V_{DS,sat}$.

- **PDITS** and **PDITSD**, which are linked to the DITS effect.

- **PCLM**, **PCLMG** and **FPROUT**, which are linked to the channel length modulation (CLM) effect.

- **PDIBLC**,which is linked to the impact of the DIBL effect on R_{out}.

- **PVAG**, which is linked to the V_G dependence on Early voltage.

C_{GG} vs. V_G Analysis at $V_{DS} \neq 0\ V$ & $V_B = 0\ V$

The velocity saturation (VSAT) and the channel length modulation (CLM) effects not only affect the static behavior of the transistor but its dynamic performance as well. The extraction of **VSAT** and **PCLM** from the I_D vs. V_G and the I_D vs. V_D curves should be sufficient in order to capture these effects for CV operation consistently. To verify that, the C_{GG} vs. V_G characteristic for different $V_{DS} \neq 0\ V$ values, from linear mode to saturation must be studied. If different values for **VSAT** and **PCLM** are necessary for accurate fitting of the CV behavior at different V_D biases, then the **VSATCV** and **PCLMCV** can be used to fine tune the dynamic operation without affecting the static one.

I_D vs. V_G Analysis at $V_D = [V_{D,lin}, V_{D,sat}]$ & various V_{SB}

In this step almost the same procedure as in Section 7.2.1 is repeated in order to extract the parameters that are linked to the body effect. Similar to the Section 7.2.1, it is very important that during the extraction in this step both I_D and g_m are studied simultaneously.

Linear Mode

- Focusing in the *weak-inversion* region, **CDSCB**, which is linked to the V_{SB} dependence of the sub-threshold slope, is extracted. Also **K2**, which is linked to the V_{TH} shift due to vertical non-uniform doping, is extracted in the same region.

- In the *strong-inversion* region, **UC**, which is linked to the V_B (or V_S) dependence of mobility, is extracted. The parameter for V_{SB} dependence of S/D series resistances, **PRWB**, is also extracted under the same bias conditions.

Saturation

- In the *strong-inversion* region, the parameter that is connected to V_{SB} dependence of the velocity saturation effect, i.e. **PSATB**, is extracted.

In order to validate that the values of the parameters, which are linked to V_{SB} dependencies, are correctly extracted, it is useful to check I_D vs. V_D and g_{ds} vs. V_D characteristics at various V_G & $V_{SB} \neq 0\ V$ and, if needed, to fine tune the values of the parameters.

Fitting Verification

After all the extraction steps of this part have been performed, the fitting of the model should be checked for all the analysis carried out up to this point. In case the fitting is not accurate enough for any of the analyses, all the steps (from 7.2.1 to 7.2.1) must be repeated for fine tuning the parameters.

7.2.2 Extraction of Short Channel Effects & Length Scaling Parameters

Once the behavior of the wide/long channel device has been accurately modeled, the next step is the extraction of the parameters that are either related to short channel effects or express the different length dependencies. consequently at this phase, devices across the entire length range of the technology, from the shortest to the longest one, are studied simultaneously. In order to avoid the impact of narrow channel effects or of the width dependencies these devices should, ideally, have the same channel **width**. The extraction that is carried out follows the same flow as in Section 7.2.1, but now a set of devices with constant channel **width** but different channel lengths is used.

C_{GG} vs. V_G Analysis at $V_S = 0\ V$, $V_D = 0\ V$ & $V_B = 0\ V$

In this step, the parameters related to the overlap and the fringing capacitances as well as those that are linked to the length dependence of the doping concentration and the flat-band voltage are extracted. More specifically:

- **NDEPL1**, **NDEPLEXP1**, **NDEPL1** and **NDEPLEXP1**, which are the length scaling parameters for the doping concentration, are extracted from C_{GG} in the depletion region. If possible, it is recommended that those parameters are extracted from C_{GB} vs. V_G analysis, while the S and D terminals are grounded.

- The extraction of parameters related to the overlap and the fringing capacitances is carried out by studying the entire range of V_G bias of C_{GG} vs. V_G characteristic. These parameters are: **CGSO**, **CGDO**, **CGBO**, **CGSL**, **CGDL**, **CKAPPAS**, **CKAPPAD** and **CF**. If possible, it is recommended that **CGSO**, **CGDO**, **CGBO** and **CF** are extracted from C_{GD} vs. V_G at low V_B values and with the S and D terminals being connected together and grounded and with the B terminal biased, while **CGSL**, **CGDL**, **CKAPPAS** and **CKAPPAD** are extracted from C_{GD} vs. V_G at high V_B, again with the S and D terminals being connected together and

grounded and with the B terminal biased.

- **DLC**, which is the channel-length offset parameter for the CV model, is extracted in the strong-inversion region of C_{GG}.

- **VFBCVL** and **VFBCVLEXP**, which express the length dependence of the flat-band voltage at CV, are extracted from the depletion region and till the onset of strong-inversion. In order to be able to use **VFBCVL** and **VFBCVLEXP** parameters, **VFBCV** must be $\neq \mathbf{0}$.

I_D vs. V_G Analysis at $V_D = [V_{D,lin}, V_{D,sat}]$, $V_S = 0\ V$ & $V_B = 0\ V$

In this step, parameters related to short channel effects or to length dependencies of I_D vs. V_G, are extracted. Similar to the procedure described in Section 7.2.1, the parameters are divided into two groups. The first group includes the parameters which are extracted in *linear mode* (i.e. $V_D \ll V_G - V_{TH}$) and the latter group includes the parameters which are extracted in *saturation* (i.e. $V_D \gg V_G - V_{TH}$). It is very important that during the extraction both I_D and g_m of all the devices are studied in parallel.

Linear Mode

- Focusing in the *weak-inversion* region (I_D vs. V_G characteristic in the $V_G < V_{TH}$ region and preferably with the y-axis in logarithmic scale), **NFACTORL** and **NFACTORLEXP**, which are related to the length dependence of the sub-threshold slope of I_D vs. V_G, can be extracted. Furthermore, **LINT**, which is the channel length offset parameter, is used to fit both the sub-threshold slope and the V_{TH} itself. For fitting the V_{TH} of the devices also **DVTP0** and **UD** can prove to be useful. **UD** should be used only for fine tuning because it mainly affects the strong-inversion region above threshold voltage. It is recommended that the parameters **NDEPL1**, **NDEPLEXP1**, **NDEPL1** and **NDEPLEXP1** keep the already extracted values from the C_{GG} vs. V_G analysis (Section 7.2.2). But, if the fitting of the V_{TH} across the entire length range cannot be achieved without changing the values of **NDEPL1**, **NDEPLEXP1**, **NDEPL1** and **NDEPLEXP1**, then these parameters are used for static current operation (IV) and **NDEPCVL1**, **NDEPCVLEXP1**, **NDEPCVL1** and **NDEPCVLEXP1** parameters are used for dynamic operation (CV).

- In the *strong-inversion* region, the parameters related to the length dependence of: i) the mobility; **U0L** and **U0LEXP**, ii) the effect of mobility reduction due to vertical field; **UAL**, **UALEXP**, **EUL** and **EULEXP** and iii) the coulomb scattering effect; **UDL** and **UDLEXP**, are extracted. Furthermore, parameters for the length dependence of S/D series resistances, namely **RDSWL** and **RDSWLEXP** (when **RDSMOD = 0**) or **RSWL**, **RSWLEXP**, **RDWL** and **RDWLEXP** (when **RDSMOD = 1**), are also extracted under the same bias conditions.

Saturation

- In the *weak-inversion* region (I_D vs. V_G characteristic with the y-axis in logarithmic scale), **CDSCDL** and **CDSCDLEXP** paramerers, which are linked to the length dependence of the sub-threshold slope dependence on drain bias, are extracted. Moreover, parameters for the DIBL effect, which control V_{TH} when $V_{DS} \neq 0$, namely **ETA0** and **DSUB**, are also extracted in the same region.

- Focusing in the *strong-inversion* region, the length scaling parameters linked to the velocity saturation effect, i.e **VSATL**, **VSATLEXP**, **PSATL**, **PSATLEXP**, **PTWGL** and **PTWGLEXP**, can be extracted.

Finally, from the *accumulation* to *depletion* region, in both *linear mode* and *saturation*, the parameters **AGIDLL**/**AGISLL**, which are related to the length dependence of the GIDL effect (GIDL/GISL currents), are extracted.

I_G vs. V_G Analysis at various V_D, V_S = 0 *V* & V_B = 0 *V*

From the I_G vs. V_G analysis, the parameters related to the length dependence of gate current are extracted. These parameters are: **AIGCL**, **AIGSL**, **AIGDL** and **PIGCDL**.

I_D vs. V_D Analysis at various V_G, V_S = 0 *V* & V_B = 0 *V*

In this step, both I_D vs. V_D and g_{ds} vs. V_D characteristics should be studied simultaneously. Similar to the procedure described in the Section 7.2.2 the parameters that are extracted are:

- **DELTAL** and **DELTALEXP**, which are related to the length dependence of the velocity saturation effect and the smoothing factor for the calculation of V_{DS} as it approaches $V_{DS,sat}$.

- **PDITSL**, which is linked to the length dependence of the DITS effect.

- **PCLML**, **PCLMLEXP**, **FPROUTL** and **FPROUTLEXP**, which are linked to the length dependence of the CLM effect.

- **PDIBLCL** and **PDIBLCLEXP**, which are linked to the length dependence of the impact of the DIBL effect on R_{out}.

It is very important to be mentioned here, that if the slope of the g_{ds} vs. V_D curve at low levels of inversion is steeper than the measurements, then **ETA0** should be decreased and, at the

same time, **DVTP1** can be used in order to achieve an accurate fit for the V_{TH} in *saturation*.

C_{GG} vs. V_G Analysis at $V_{DS} \neq 0\ V$ & $V_B = 0\ V$

The extraction of the length scaling parameters of **VSAT** and **PCLM** from the I_D vs. V_G and the I_D vs. V_D characteristics (Steps 7.2.2 and 7.2.2) should be sufficient in order to capture VSAT and CLM effects for the CV behavior as well. To verify that, C_{GG} vs. V_G characteristic of all wide devices with various gate lengths, for different $V_{DS} \neq 0\ V$, from linear mode to saturation, must be studied. If different values for **VSATL**, **VSATLEXP**, **PCLML** and **PCLMLEXP** are necessary for accurate fitting of the CV behavior across the length axis, then **VSATCVL**, **VSATCVLEXP**, **PCLMCVL** and **PCLMCVLEXP** can be used.

I_D vs. V_G Analysis at $V_D = [V_{D,lin}, V_{D,sat}]$ & various V_{SB})

In this step almost the same procedure as in Section 7.2.1 will be repeated in order to extract the length scaling parameters that are linked to the body effect. Similar to the Section 7.2.1, it is also very important that during the extraction in this step both I_D and g_m of all devices are studied simultaneously.

Linear Mode

- Focusing in the *weak-inversion* region, **K2L** and **K2LEXP**, which are linked to the length dependence V_{TH} shift due to vertical non-uniform doping, are extracted.

- In the *strong-inversion* region, **UCL** and **UCLEXP**, which are linked to the length dependence of mobility reduction on the V_{SB} bias, are extracted. The parameters for the length dependence of S/D series resistances on the V_{SB} bias, namely **PRWBL** and **PRWBLEXP**, are also extracted under the same bias conditions.

Saturation

- In the *weak-inversion* region (I_D vs. V_G characteristic when y-axis is in logarithmic scale), the parameters related to the length dependence of the DIBL effect dependence on V_{SB}, namely **ETAB** and **ETABEXP**, are extracted.

In order to validate that the values of the length scaling parameters, which are linked to V_{SB} dependencies, are correctly extracted, it is useful to check I_D vs. V_D and g_{ds} vs. V_D characteristics under various V_G & $V_{SB} \neq 0\ V$ conditions and, if needed, to fine tune the values of the parameters.

Fitting Verification

When all the steps for the extraction of the short channel effects and length scaling parameters have been performed, the fitting of the model should be checked for all the analysis carried out in the Section 7.2.2. In case the fitting is not accurate enough, all the steps (from 7.2.2 to 7.2.2) must be repeated for fine tuning the parameters.

7.2.3 Extraction of Narrow Channel Effects & Width Scaling Parameters

The next step in the parameter extraction procedure is the extraction of the parameters that are either related to narrow channel effects or express the different width dependencies. So at this part, devices across the entire width range of the technology, from the narrowest to the widest one, are studied simultaneously. In order to avoid the impact of short channel effects or of the length dependencies these devices should ideally have the same **long** channel. The extraction that is carried out follows the same flow as in Section 7.2.2, but now a set of devices with constant **long** channel but different channel widths is used.

C_{GG} vs. V_G Analysis at $V_S = 0\ V$, $V_D = 0\ V$ & $V_B = 0\ V$

In this step, parameters related to the width dependencies of the CV behavior of the device, e.g. width dependence of the doping concentration and flat-band voltage, are extracted. More specifically:

- **NDEPW** and **NDEPWEXP**, which are the width scaling parameters for the doping concentration, are extracted from C_{GG} in the depletion region. If possible, it is recommended that those parameters are extracted from C_{GB} vs. V_G analysis and with the S and D terminals being grounded.

- **DWC**, which is the channel-width offset parameter for the CV model, is extracted in the strong-inversion region of C_{GG}.

- **VFBCVW** and **VFBCVWEXP**, which express the width dependence of flat-band voltage at CV, are extracted along the entire V_G bias range of the C_{GG} characteristic. In order to be able to use **VFBCVW** and **VFBCVWEXP** parameters, **VFBCV** must be $\neq 0$.

I_D vs. V_G Analysis at $V_D = [V_{D,lin}, V_{D,sat}]$, $V_S = 0\ V$ & $V_B = 0\ V$

In this step, parameters related to the width dependencies of I_D vs. V_G, are extracted. Similar to the procedure described in Section 7.2.1, the parameters are divided in two groups, those which are extracted in the *linear mode* (i.e. $V_D \ll V_G - V_{TH}$) and those which are extracted in

saturation (i.e. $V_D \gg V_G - V_{TH}$). It is very important that during the extraction both I_D and g_m of all the devices are studied in parallel.

Linear Mode

- Focusing in the *weak-inversion* region (I_D vs. V_G characteristic when y-axis is in logarithmic scale), **NFACTORW** and **NFACTORWEXP**, which are related to the width dependence of the sub-threshold slope of I_D vs. V_G, can be extracted. Furthermore, **WINT**, which is the channel width offset parameter, is used to fit both the sub-threshold slope and the V_{TH} across W. It is recommended that the parameters **NDEPW** and **NDEPWEXP** keep the values extracted from the C_{GG} vs. V_G analysis (Section 7.2.3). However, in the case that a good description of the V_{TH} across the entire width range cannot be achieved without changing the values of **NDEPW** and **NDEPWEXP**, then these parameters are used for static operation (IV) and the **NDEPCVW** and **NDEPCVWEXP** parameters are used for dynamic operation (CV).

- In the *strong-inversion* region, the parameters related to the width dependence of the mobility reduction due to vertical field effect, namely **UAW**, **UAWEXP**, **EUW** and **EUWEXP**, are extracted.

Saturation

- Focusing in the *strong-inversion* region, the width scaling parameters linked to the velocity saturation effect, i.e. **VSATW** and **VSATWEXP**, can be extracted.

Finally, from the *accumulation* to *depletion* region, in both *linear mode* and *saturation*, the parameters **AGIDLW**/**AGISLW**, which are related to the width dependence of GIDL effect (GIDL/GISL currents), are extracted.

In order to validate that the values of the width scaling parameters are correctly extracted, it is useful to check I_D vs. V_D and g_{ds} vs. V_D characteristics under various V_G & $V_{SB} \neq 0$ V bias conditions and, if needed, to fine tune the values of the parameters, depending on the priorities and the modelling focus of the process.

I_G vs. V_G Analysis at various V_D, V_S = 0 V & V_B = 0 V

From the I_G vs. V_G analysis, the parameters related to the width dependence of gate current are extracted. These parameters are: **AIGCW**, **AIGSW** and **AIGDW**.

C_{GG} vs. V_G Analysis at $V_{DS} \neq 0\ V$ & $V_B = 0\ V$

The extraction of the width scaling parameters of **VSATW** and **VSATWEXP** from I_D vs. V_G and I_D vs. V_D characteristics (Step 7.2.3) should be sufficient in order to capture VSAT for CV operation. To verify that, the C_{GG} vs. V_G characteristic of all devices, for different $V_{DS} \neq 0\ V$, from linear mode to saturation, must be studied. If different values for **VSATW** and **VSATWEXP** are necessary for accurate fitting of the CV behavior across W, then **VSATCVW** and **VSATCVWEXP** can be used.

I_D vs. V_G Analysis at $V_D = [V_{D,lin}, V_{D,sat}]$ & various V_{SB}

In this step, from the *weak-inversion* region of *linear mode*, **K2W** and **K2WEXP**, which are linked to the width dependence V_{TH} shift due to vertical non-uniform doping, can be extracted. For validation purposes, it is useful to check: i) the I_D vs. V_G and g_m vs. V_G characteristics in *weak-* and *strong-inversion* and for both the *linear mode* and *saturation*, and ii) the I_D vs. V_D and g_{ds} vs. V_D characteristics at various V_G & $V_{SB} \neq 0\ V$ and, if needed, extract **K2W** and **K2WEXP** to fit both (i) and (ii) cases.

Fitting Verification

When all the extraction steps for the width scaling have been performed, the fitting of the model should be checked for all the analysis carried out in Section 7.2.3. In case the fitting is not accurate enough, all the steps (from 7.2.3 to 7.2.3) must be repeated for fine tuning the parameters.

7.2.4 Extraction of Parameters for Narrow/Short Channel Devices

The final part in the parameter extraction procedure from a geometrical point of view, is the extraction of the parameters for narrow/short channel devices. These devices have the minimum dimensions so it is of the highest difficulty to capture their behavior. Since the minimum channel device parameters can affect the already performed fitting across length and width, it is recommended that two different sets of devices are studied simultaneously, i.e. one set with constant **short** channels but different channel widths (from narrowest to widest) and one set with constant **narrow** channels but different channel lengths (from the shortest to the longest one). This way, the minimal dimensions device is reached by two different axis and the appropriate modeling level of all the geometrical region of the narrow and short geometries is assured.

C_{GG} vs. V_G Analysis at $V_S = 0\ V$, $V_D = 0\ V$ & $V_B = 0\ V$

In this step, the geometry dependent parameters for modeling the CV behavior of the narrow/short channel devices, are extracted. More specifically:

- **NDEPWL** and **NDEPWLEXP**, which are used to fit the doping concentration of small channel devices, are extracted from the C_{GG} in the depletion region. If possible, it is recommended that those parameters are extracted from the C_{GB} vs. V_G analysis with the S and D terminals being shorted to the ground.

- **LWLC** and **WWLC**, which are the coefficients of the length and the width dependencies for CV model, respectively, are extracted in the strong-inversion region of the C_{GG}.

- **VFBCVWL** and **VFBCVWLEXP**, which are used to fit the flat-band voltage at CV, are extracted from the depletion till the onset of strong-inversion region of the C_{GG} characteristic. In order to be able to use **VFBCVWL** and **VFBCVWLEXP** parameters, **VFBCV** has to be set $\neq \mathbf{0}$.

I_D vs. V_G Analysis at $V_D = [V_{D,lin}, V_{D,sat}]$, $V_S = 0\ V$ & $V_B = 0\ V$

In this step, the geometry dependent parameters for modeling I_D of the narrow and short channel devices, are extracted. Similar to the procedure described in Section 7.2.1, the parameters are divided in two groups, those which are extracted in the *linear mode* (i.e. $V_D \ll V_G - V_{TH}$) and those which are extracted in the *saturation* (i.e. $V_D \gg V_G - V_{TH}$). It is very important that, during the extraction, both I_D and g_m of all the devices are studied at the same time.

Linear Mode

- Focusing in *weak-inversion* region (I_D vs. V_G characteristic when y-axis is in logarithmic scale), **NFACTORWL** and **NFACTORWLEXP**, which are used to fit the sub-threshold slope of I_D vs. V_G for small channel devices, can be extracted. It is recommended that the parameters **NDEPWL** and **NDEPWLEXP** keep the values extracted from the C_{GG} vs. V_G analysis (Section 7.2.4). But, if the fitting of the V_{TH} for both sets of devices cannot be achieved without changing the values of **NDEPWL** and **NDEPWLEXP**, then these parameters are used for the static operation (IV) and the **NDEPCVWL** and **NDEPCVWLEXP** parameters are used for the dynamic operation (CV).

- In the *strong-inversion* region, the parameters which are used to model the effect of mobility reduction due to vertical field in small channel devices, namely **UAWL**, **UAWLEXP**, **EUWL** and **EUWLEXP**, are extracted.

Saturation

- Focusing in the *strong-inversion* region, the parameters which are used to model the velocity saturation effect in small channel devices, i.e. **VSATWL** and **VSATWLEXP**, can be extracted.

In order to validate that the values of the parameters, which model the behavior of narrow and short channel devices, are correctly extracted, it is useful to check the I_D vs. V_D and g_{ds} vs. V_D characteristics at various V_G & $V_{SB} \neq 0\ V$ and, if needed, to fine tune the values of the parameters.

C_{GG} vs. V_G Analysis at $V_{DS} \neq 0\ V$ & $V_B = 0\ V$

The extraction of the parameters, which are used to model to the velocity saturation effect in small channel devices, **VSATWL** and **VSATWEXP**, from the I_D vs. V_G and I_D vs. V_D characteristics (Step 7.2.4) should be sufficient in order to capture the VSAT for CV operation, as well. To verify that, the C_{GG} vs. V_G characteristic of all devices, for different $V_{DS} \neq 0\ V$, from linear mode to saturation, must be studied. If different values for **VSATWL** and **VSATWLEXP** are necessary for accurate fitting of the CV behavior of devices, then **VSATCVWL** and **VSATCVWLEXP** can be used.

I_D vs. V_G Analysis at $V_D = [V_{D,lin}, V_{D,sat}]$ & various V_{SB}

In this step, from the *weak-inversion* region of *linear mode*, the **K2WL** and **K2WLEXP** parameters, which are linked to the V_{TH} shift due to vertical non-uniform doping in small channel devices, can be extracted. For validation purposes, it is useful to check: i) the I_D vs. V_G and g_m vs. V_G characteristics in *weak-* and *strong-inversion* and for both *linear mode* and *saturation*, and ii) the I_D vs. V_D and g_{ds} vs. V_D characteristics at various V_G & $V_{SB} \neq 0\ V$ and, if needed, update **K2WL** and **K2WLEXP** to fit both (i) and (ii).

Fitting Verification

When all the steps for narrow and short channel devices have been performed, the fitting of the model should be checked for all the analysis carried out in Section 7.2.4. In case the fitting is not accurate enough, all the steps (from 7.2.4 to 7.2.4) must be repeated for the fine tuning the parameters.

7.2.5 Extraction of Temperature Dependence Parameters

Up to this point of the parameter extraction procedure, the temperature dependence of the parameters has been ignored since all the parameters were extracted at **TNOM**, and all the

reference data were considered at this temperature. In this part, the parameters that are related to the impact of the temperature on the behavior of devices are extracted, and for that, data across a certain temperature range of the technology are necessary. The behavior of devices is studied with the same geometrical sequence as the previous steps, while the temperature dependence parameters are extracted in the same regions of operation as the parameters of the corresponding physical effects.

Wide & Long Channel Devices

The first step, in the extraction of temperature (T) dependence parameters, is to study the behavior of a long and wide channel device **at different T** and for different analyses. It is recommended that the same device as the one in Section 7.2.1 is used. In more detail, the following guidelines are drawn:

I_D vs. V_G analysis at $V_D = V_{D,lin}$, $V_S = 0\ V$ & $V_B = 0\ V$

- From the *weak-inversion* region (I_D vs. V_G characteristic when y-axis is in logarithmic scale), the parameters **TBGASUB** and **TBGBSUB**, which control the temperature dependence of the energy band-gap (E_g), are extracted. Also, **TNFACTOR** is extracted in order to fit the sub-threshold slope of I_D in different T, while **KT1** and **KT1EXP** are extracted by fitting the V_{TH} across T.

- From the *strong-inversion* region, the mobility temperature exponent, **UTE** and the temperature coefficients: i) for mobility reduction due to vertical field effect, namely **UA1** and **UD1**, ii) for the coulomb scattering effect, **UCSTE** and iii) for the S/D series resistances, **PRT**, are extracted.

I_D vs. V_G analysis at $V_D = V_{D,sat}$, $V_S = 0\ V$ & $V_B = 0\ V$

- From the *strong-inversion* region, the parameters that are used to model the temperature dependence of the velocity saturation effect, i.e. **AT** and **PTWGT**, are extracted.

It is very important that in the above analysis both I_D and g_m of all the devices are studied at once. Furthermore, from the *accumulation* to *depletion* region, in both *linear mode* and *saturation* of the I_D vs. V_G analysis, the parameter **TGIDL**, which controls the temperature dependence of GIDL effect, is extracted.

I_D vs. V_D Analysis at various V_G, $V_S = 0\ V$ & $V_B = 0\ V$

From I_D vs. V_D analysis in different temperatures, **TDELTA**, which is related to the temperature dependence of the smoothing factor for the effective V_{DS} as it approaches $V_{DS,sat}$, is extracted.

I_D vs. V_G Analysis at $V_D = V_{D,lin}$ & various V_{SB}

- From the *weak-inversion* region (I_D vs. V_G characteristic when y-axis is in logarithmic scale) **KT2**, which is linked to the temperature dependence of V_{TH} shift due to vertical non-uniform doping with V_{SB} bias, is extracted.

- From the *strong-inversion* region, the temperature coefficient for the mobility reduction with V_{SB} bias, namely **UC1**, is extracted.

For validation purposes, it is useful to check: i) the I_D vs. V_G and g_m vs. V_G characteristics in the *weak-* and *strong-inversion* and for both *linear mode* and *saturation*, and ii) the I_D vs. V_D and g_{ds} vs. V_D characteristics at various V_G & $V_{SB} \neq 0\ V$ and, if needed, extract **KT2** and **UC1** to fit both (i) and (ii).

Length Scaling of Wide Channel Devices

The following step in the extraction of temperature dependence parameters, is to study the temperatures dependences across the length axes. For this reason, data **at different T** of a set of devices with constant **wide** channel but different channel lengths are used.

I_D vs. V_G analysis at $V_D = V_{D,lin}$, $V_S = 0\ V$ & $V_B = 0\ V$

- From the *weak-inversion* region (I_D vs. V_G characteristic when y-axis is in logarithmic scale), the parameter **KT1L** is extracted for fitting the V_{TH} across the length, at different temperatures.

- From the *strong-inversion* region, the length scaling parameters for: i) the mobility temperature exponent, **UTEL** and for the temperature coefficients or mobility reduction due to vertical field effect, namely **UA1L** and **UD1L**, are extracted.

I_D vs. V_G analysis at $V_D = V_{D,sat}$, $V_S = 0\ V$ & $V_B = 0\ V$

- From the *weak-inversion* region (I_D vs. V_G characteristic when y-axis is in logarithmic scale), the parameter **TETA0**, which is related to the temperature dependence of DIBL effect and thus is controlling the V_{TH} in saturation, is extracted.

- From *strong-inversion* region, the parameters that are used to model the temperature dependence of velocity saturation effect across the length, i.e. **ATL** and **PTWGTL**, are extracted.

It is very important that in the above analysis both I_D and g_m of all the devices are studied at once. For validating that the values of length scaling parameters for temperature dependence parameters are extracted correctly, it is useful to check also the I_D vs. V_D and g_{ds} vs. V_D characteristics and, if needed, to fine tune the values of the corresponding parameters by repeating Step 7.2.5.

7.3 Conclusions

The development of the compact models is based on their flexibility and their ability to adapt their predictive behavior on the actual measurements of a specific technology against which they are called to be used. The parameter extraction methodology can be regarded as a crude mathematical problem, on one side. However, this approach does not allow the user to exploit the maximum of the analytical and physics based structure of the model and it is characterized by an over-demanding computationally effort. On the other side, the parameter extraction procedure may be regarded as a physics-based problem, whose solution is based on identifying the regions and the bias conditions on which each phenomenon manifests it self more intensively. This approach minimizes the effort in computational terms, however, it might not exploit all the mathematical space for improvement due to various higher order effects that are correlated and depend on each other.

Practically, the best approach to be used is a path which lays in between the two extreme approaches mentioned above. The physics and the knowledge of the model will set a clear path on the process and the will define the main conditions under which each parameter and corresponding phenomenon is to be treated. However, at a second level some iterations are going to be needed in order to balance out all the correlations between the various effects and to minimize the distance between the model and the targeted data in the full range of the technology, covering various geometries and temperature conditions. This hybrid approach was described above and is illustrated in Fig. 7.1 and Fig. 7.2.

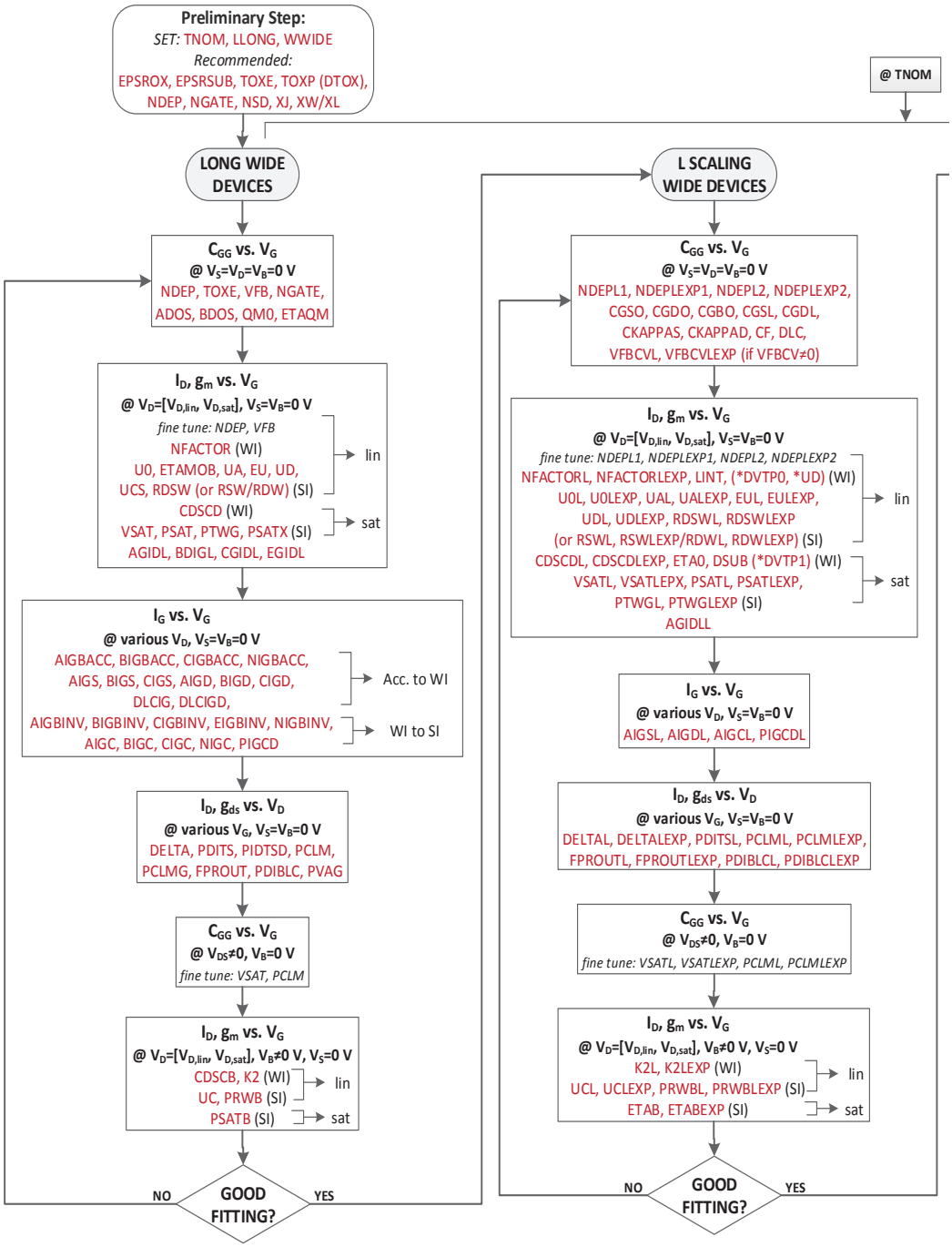


Figure 7.1 – The parameter extraction procedure and guidelines of BSIM6 expressed in the form of a flowchart - part A.

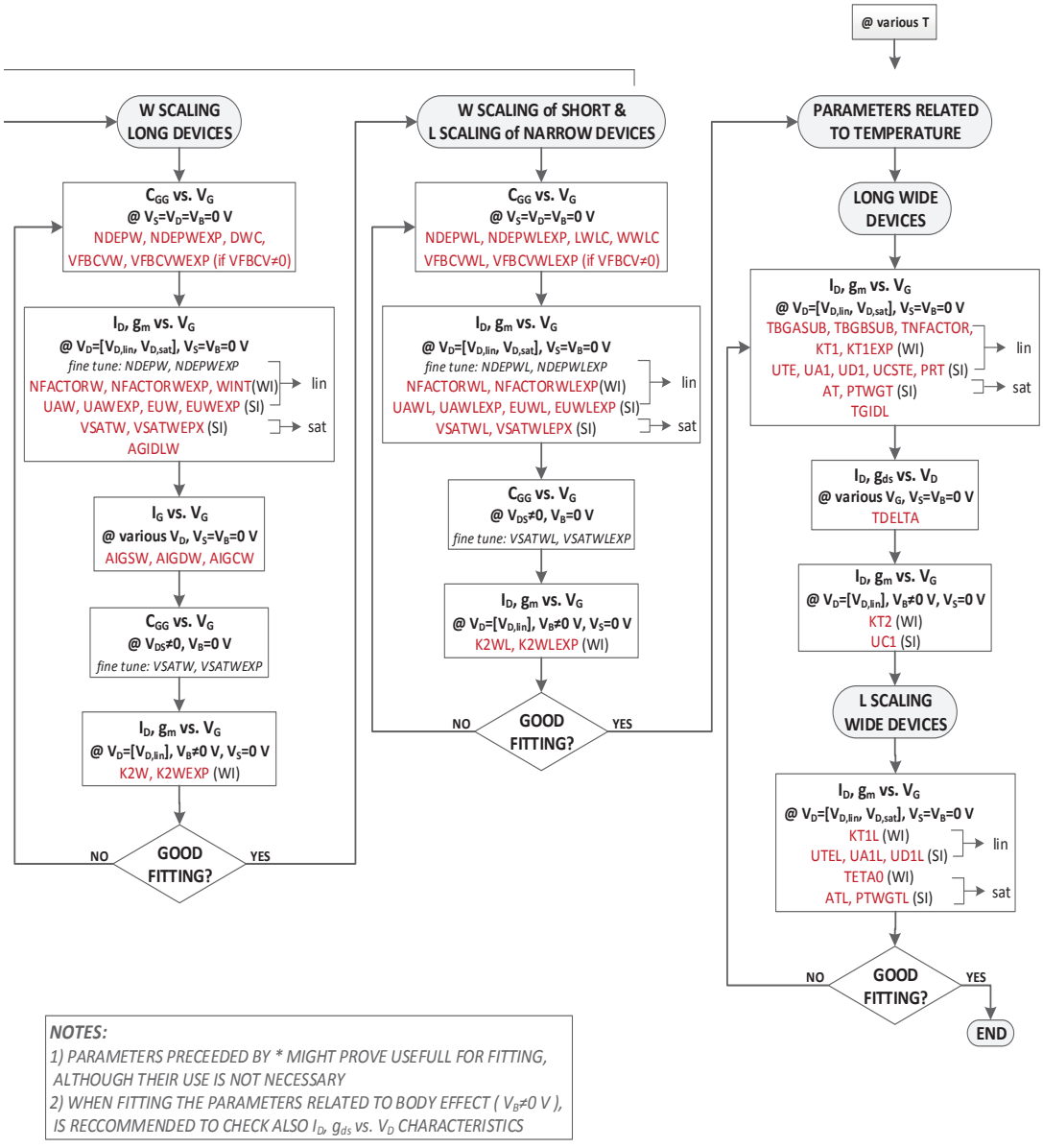


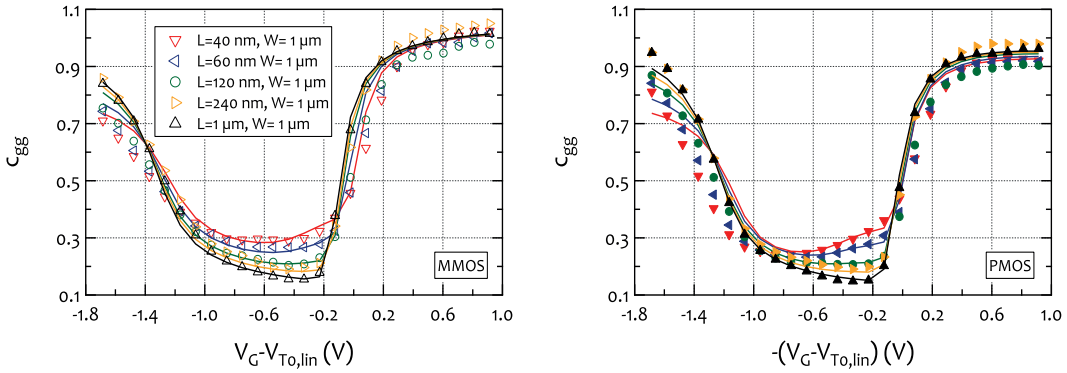
Figure 7.2 – The parameter extraction procedure and guidelines of BSIM6 expressed in the form of a flowchart -part B.

8 Evaluation of BSIM6 in Nanoscale CMOS Technologies

8.1 Introduction

State-of-the-art technologies need more complex models that subsequently demand more time to be developed. Taking also into account both the advancements of technology and the time required for a model to be adopted by the design community, the actual lifetime of a model is reduced significantly. Under the described conditions, it is crucial that the model is thoroughly benchmarked from its first steps to speedup its development and adoption. BSIM6 went through an extensive validation from its early development years. Thanks to this evaluation, the model proved its abilities to accurately represent the behavior of nanoscale bulk CMOS processes, facilitating its standardization.

In this chapter, results of this broad evaluation procedure are demonstrated. The model is validated against measurements of the two commercial, state-of-the-art 40 nm and 28 nm bulk CMOS technologies, over a wide range of biases and geometries, for nMOS and pMOS devices, and for different modes of operation, namely dynamic (CV), static (IV) and RF. In order for the model to be evaluated, different global model cards (different for nMOS and pMOS devices and different for the two processes) were extracted. A full extraction procedure is carried out as follows. The first step is to extract a global model card, that can be used: a) across the geometry plane (W/L), b) for all bias conditions; from weak- to strong-inversion and from linear operation to saturation, c) for the whole temperature (T) range of the technology and d) for CV and IV operation, as described in Ch. 7. Then this model card is extended for RF operation by: a) fine tuning some of the parameters so that the model can capture the changes in CV and IV behavior of the RF devices, which stem from the differences in the topology of the RF DUTs with respect to the DC ones, and b) including the contribution of the parasitic capacitances and resistances that surround the intrinsic channel of the devices and become dominant in the RF regime. Finally, the RF noise parameters are extracted, forming the final model card [111].



(a) nMOS (b) pMOS

Figure 8.1 – Normalized total gate capacitance c_{gg} vs. $|V_G - V_{T0,lin}|$ at $V_S = V_D = V_B = 0$ V of a 40 nm CMOS process. (a) nMOS and (b) pMOS DC DUTs, with $W = 1$ µm and $L = [40n, 60n, 120n, 240n, 1\mu]$ m.

8.2 Validation against a 40 nm CMOS Process

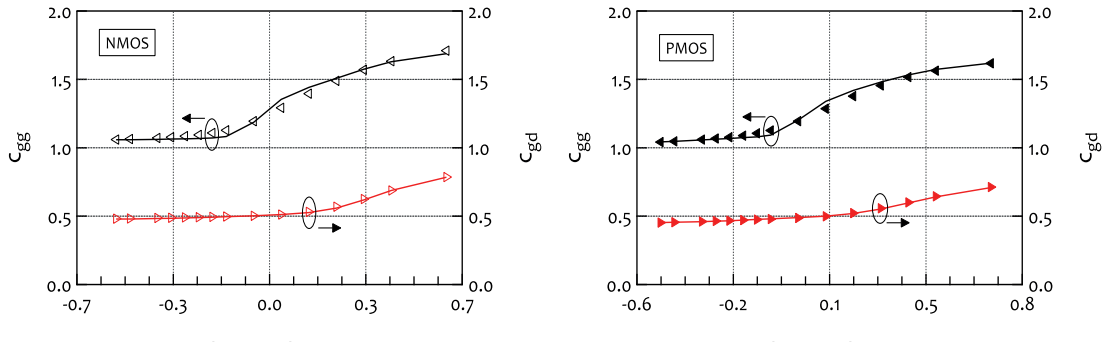
BSIM6 was evaluated extensively against measurements of an advanced 40 nm CMOS technology. At the time that BSIM6 was under development, the 40 nm node was the state-of-the-art commercial node. As a result, more results are demonstrated against this technology, compared to the 28 nm technology. In this section, a selection of the most representative results of the comparison between the model and measurements of the 40 nm process is presented. In each analysis the bias conditions and the DUTs that are used are described.

8.2.1 CV Operation

One of the most challenging, but also significant, properties of a compact model is its capability to provide a reliable description of the dynamic behavior of devices, especially of those that are scaled down to the nanometer range. In Fig. 8.1, the normalized total gate capacitance c_{gg} vs. $V_G - V_{T0,lin}$ for nMOS and pMOS DC DUTs, extending from the accumulation region to strong-inversion and for devices having different channel lengths, is presented. The C_{GG} capacitance is normalized to $C_{ox} \cdot W_{tot} \cdot L$ according to Eq. (3.18). $V_{T0,lin}$ represents the threshold voltage of the long channel device (different for nMOS and pMOS devices) when there is no body-effect, i.e. $V_{SB} = 0$ V. The results verify the correct behavior of the model even for the shortest channel devices with $L_{min} = 40$ nm [111].

Capacitances can be extracted not only through AC measurements but also from RF measurements. Using the Y-parameters, C_{GG} and C_{GD} can be obtained from Eq. (3.11c) and Eq. (3.11d), respectively. In [72, 86] BSIM6 was evaluated under very low-bias conditions, specifically in linear mode and from weak- to strong-inversion. It should be noted here that linear operation is a region that is usually unexplored during the RF characterization of nanoscale devices. Fig. 8.2, displays the normalized form of the capacitances C_{gg} and C_{gd} vs. $|V_G - V_{T0}|$,

from depletion region to strong-inversion, where V_{T0} is the threshold voltage of each DUT at $|V_D| = 0.3$ V. Even though $|V_D| = 0.3$ V is quite low, the impact of velocity saturation (VS) and channel length modulation (CLM) effects on capacitances should not be neglected. BSIM6 is able to capture the CV behavior of the device correctly, which is strongly affected by the overlap/fringing capacitances and the VS/CLM effects.



(a) nMOS (b) pMOS

Figure 8.2 – Normalized capacitances c_{gg} and c_{gd} vs. $|V_G - V_{T0}|$, at $|V_D| = 0.3$ V: (a) nMOS and (b) pMOS minimum channel length RF DUTs of a 40 nm CMOS, with $M = 6$, $N_f = 10$, $W = 2\,\mu$m and $L = 40$ nm.

8.2.2 DC Operation

In this section different plots of the comparison of the model against measurements, when the transistor is in DC operation, are demonstrated. All the results are also normalized according to the relations presented in Table 6.1.

Corner DUTs

In Sec. 6.4 different results displaying the length scaling capabilities of the model were demonstrated. The evaluation of the scaling features of a model is meaningful only after the model's ability to describe accurately the overall behavior of at least one device, with all the involved physical phenomena, is shown. For consistency, the results of the model simulations for the fundamental DC analyses compared to measurements concerning the four corner nMOS DC DUTs of the technology are presented [105].

In Fig. 8.3a, Fig. 8.3b, Fig. 8.3c and Fig. 8.3d, the normalized drain current i_d vs. $V_G - V_{T0}$ for linear operation and saturation, in both logarithmic and linear scale is demonstrated, while in Fig. 8.3e and Fig. 8.3f, the normalized gate transconductance g_m vs. $V_G - V_{T0}$ for linear operation and saturation is shown. V_{T0} is the threshold voltage of the specified region. Fig. 8.4a, shows the normalized normalized transconductance efficiency g_{ms}/i_d vs. IC, which

in saturation is calculated after:

$$\frac{g_{\mathrm{ms}}}{i_{\mathrm{d}}} = \frac{n \cdot g_{\mathrm{m}}}{i_{\mathrm{d}}} = \frac{n \cdot \dfrac{G_{\mathrm{m}}}{G_{\mathrm{spec}}}}{\dfrac{I_{\mathrm{D}}}{I_{\mathrm{spec}}}} = \frac{nU_{\mathrm{T}} \cdot G_{\mathrm{m}}}{I_{\mathrm{D}}}. \tag{8.1}$$

Finally, Fig. 8.4b and Fig. 8.4c demonstrate the i_{d} and g_{ds} vs. V_{D} for weak- and strong-inversion. The results verify the correct behavior of the model across all regions of operation for the extreme geometries of the studied technology.

W and L Scaling

A reliable model should be able to predict the drain current for all combinations of V_{G} and V_{D}, across W and L. In order to further demonstrate this property of the model, the normalized current in linear operation and in saturation for different levels of inversion is presented in Fig. 8.5. From the plots we observe that the model demonstrates a very good scalability not only across the L, which was verified in Sec. 6.4, but also across the W axis.

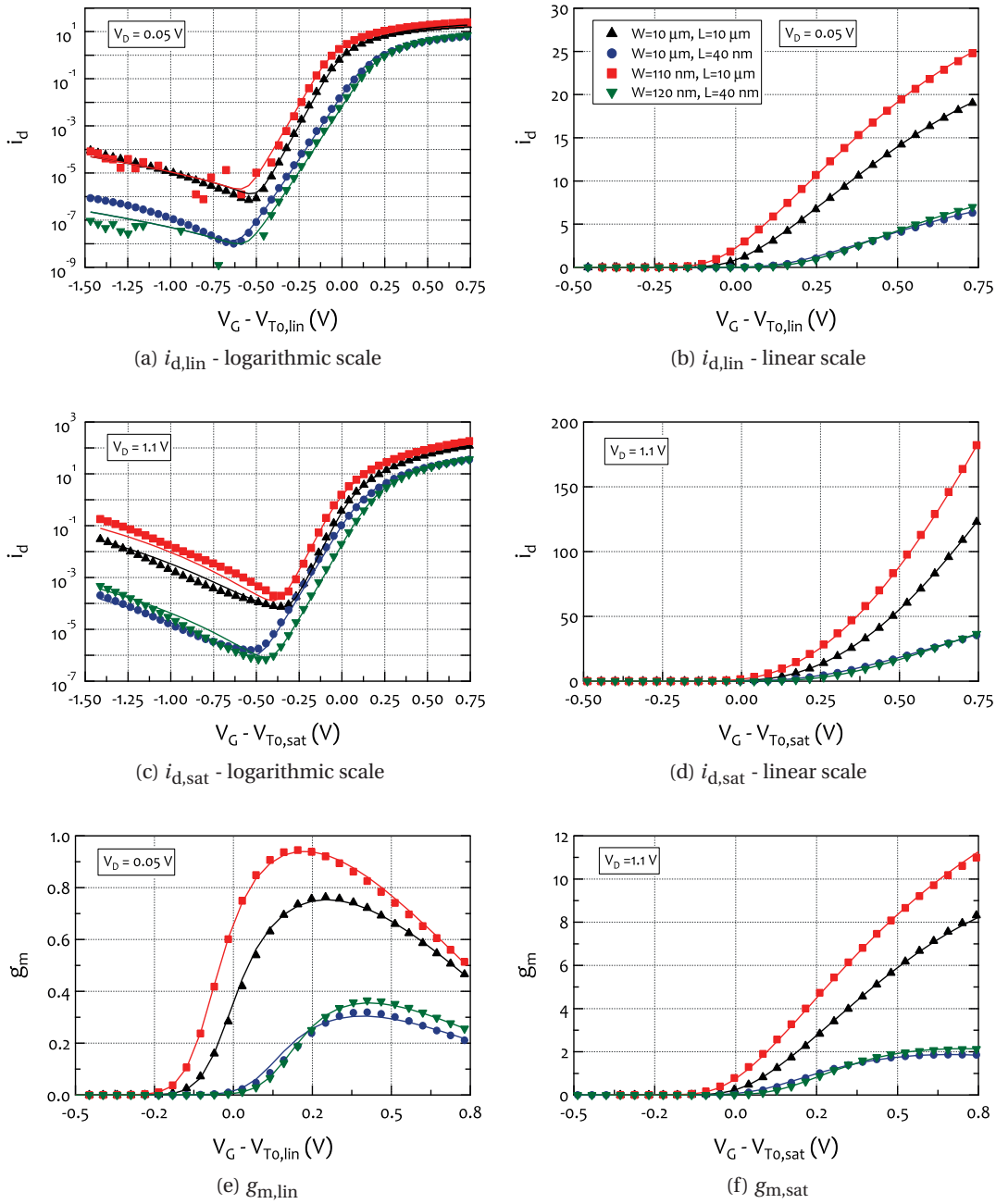
Temperature Scaling

The model was also evaluated for its ability to capture the static behavior across T. Fig. 8.6, shows the $|i_{\mathrm{ds}}|$ vs. $V_{\mathrm{G}} - V_{\mathrm{T0}}$ characteristics of the shortest channel DC devices, for 3 different temperatures including the extreme temperatures of the technology i.e., $T = -40\ ^{o}C$ and $T = 150\ ^{o}C$, in linear operation and saturation. The results show that the model is accurately capturing the impact of the temperature on the behavior of MOS transistor, while at the same time it is correctly predicting the temperature independent value of the drain current [111].

Study of the Model's Symmetry

Source and drain symmetry is a fundamental feature of an ideal MOSFET model. To check if this longitudinal symmetry is preserved by BSIM6 when the same parameter set extracted for this process is used, the Gummel symmetry test (GST) in weak- and strong-inversion was carried out [109]. The device that was selected is a short channel device where the symmetry of the model is more difficult to be preserved due to the prevailing short channel effects. The model was tested for the current $I_{\mathrm{X}} = I_{\mathrm{D}} - I_{\mathrm{S}}$ vs. $V_{\mathrm{X}} = V_{\mathrm{D}} - V_{\mathrm{S}}$, when $V_{\mathrm{D}} = -V_{\mathrm{S}}$, and its derivatives up to the 5^{th} degree (Fig. 8.7). The quantities are normalized to their maximum values as:

$$i_{\mathrm{X}} = \frac{I_{\mathrm{X}}}{\max(I_{\mathrm{X}})} \tag{8.2}$$



(a) $i_{d,lin}$ - logarithmic scale

(b) $i_{d,lin}$ - linear scale

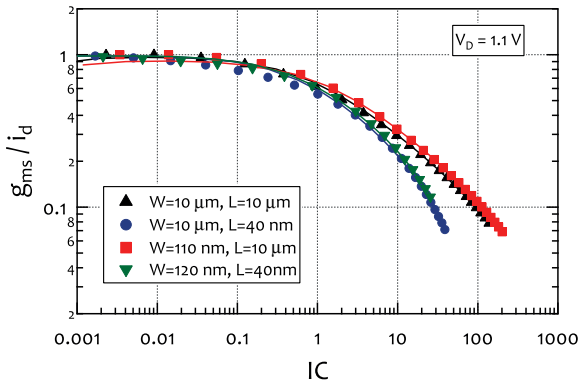
(c) $i_{d,sat}$ - logarithmic scale

(d) $i_{d,sat}$ - linear scale

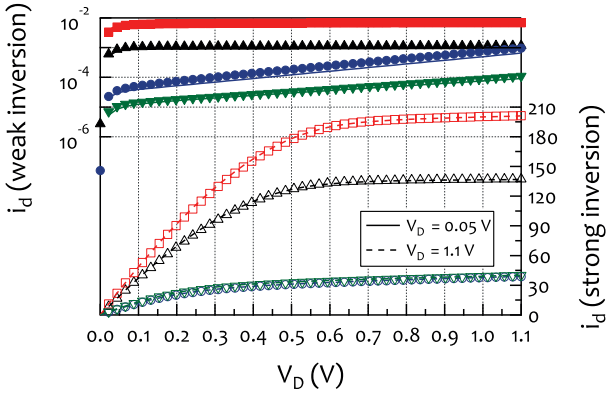
(e) $g_{m,lin}$

(f) $g_{m,sat}$

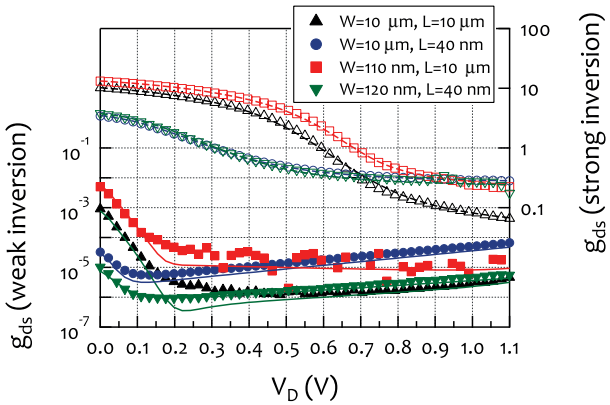
Figure 8.3 – Normalized drain current i_d and gate transconductance g_m vs. $V_G - V_{T0}$ for the corner nMOS DC DUTs of a 40 nm CMOS process: (a) $i_{d,lin}$ vs. $V_G - V_{T0,lin}$ at $V_D = 0.05$ V (linear mode), in logarithmic scale, (b) $i_{d,lin}$ vs. $V_G - V_{T0,lin}$ at $V_D = 0.05$ V (linear mode), in linear scale, (c) $i_{d,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.1$ V (saturation), in logarithmic scale, (d) $i_{d,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.1$ V (saturation), in linear scale, (e) g_m vs. $V_G - V_{T0,lin}$ at $V_D = 0.05$ V (linear mode) and (f) $g_{m,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.1$ V (saturation).



(a) g_{ms}/i_d

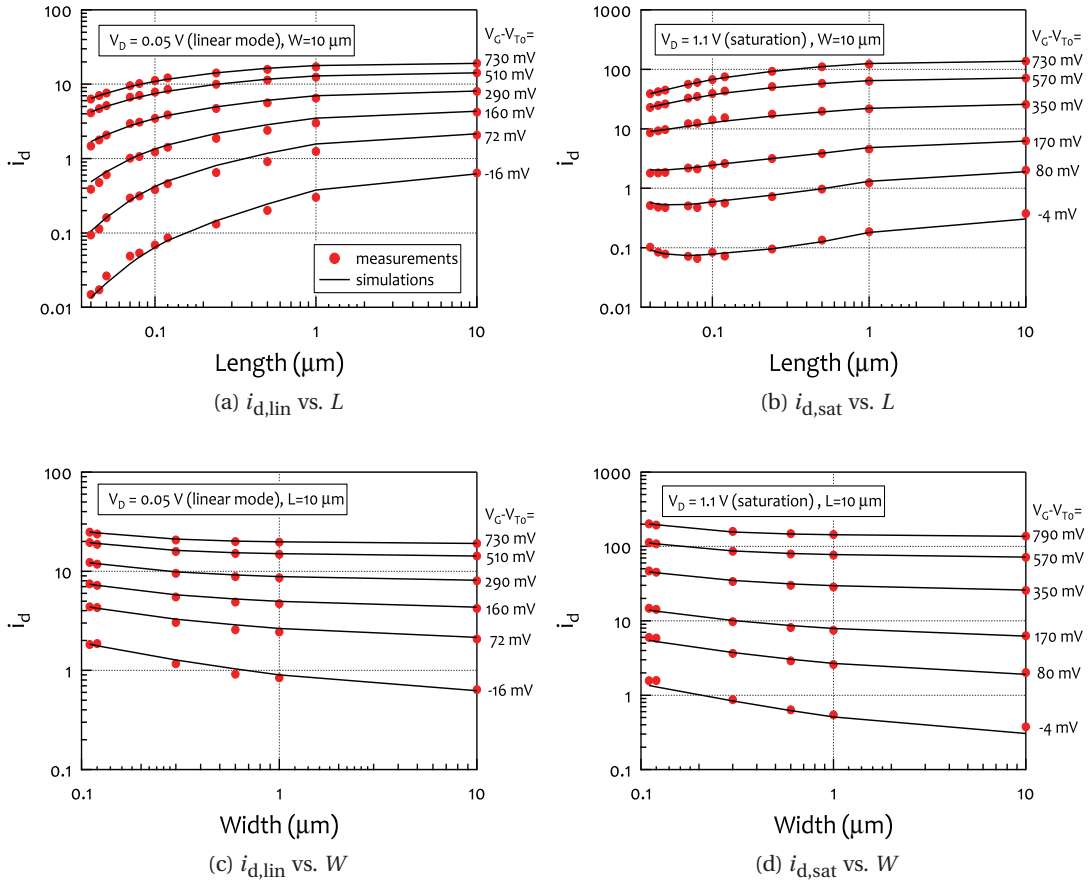


(b) i_d



(c) g_{ds}

Figure 8.4 – Transconductance efficiency g_{ms}/i_d vs. *IC* and normalized drain current i_d and output concuctance g_{ds} vs. V_D for the corner nMOS DC DUTs of a 40 nm CMOS process: (a) g_{ms}/i_d vs. i_d at $V_D = 1.1$ V (saturation), (b) i_d vs. V_D for $V_G - V_{T0} = -0.26$ V (weak-inversion) and $V_G - V_{T0} = 0.73$ V (strong-inversion) and (c) g_{ds} vs. V_D for $V_G - V_{T0} = -0.26$ V (weak-inversion) and $V_G - V_{T0} = 0.73$ V (strong-inversion).



(a) $i_{\mathrm{d,lin}}$ vs. L

(b) $i_{\mathrm{d,sat}}$ vs. L

(c) $i_{\mathrm{d,lin}}$ vs. W

(d) $i_{\mathrm{d,sat}}$ vs. W

Figure 8.5 – Length and width scaling of the normalized drain current i_{d} for various V_{G} for nMOS DC DUTs of a 40 nm CMOS process. (a) i_{d} vs. L, $V_{\mathrm{D}} = 0.05$ V (linear mode), (b) i_{d} vs. L, $V_{\mathrm{D}} = 1.1$ V (saturation) (c) i_{d} vs. W, $V_{\mathrm{D}} = 0.05$ V (linear mode) and (d) i_{d} vs. W, $V_{\mathrm{D}} = 1.1$ V (saturation).

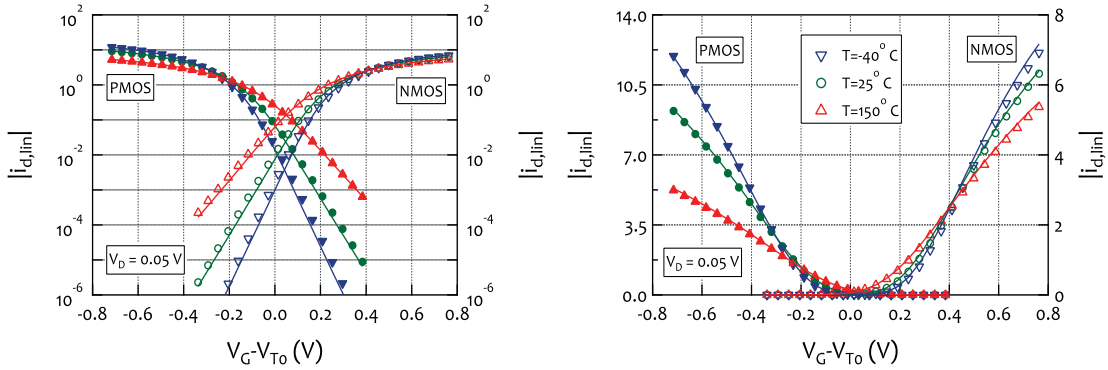
and

$$\frac{\partial^n i_{\mathrm{X}}}{\partial V_{\mathrm{X}}^n} = \frac{\dfrac{\partial^n I_{\mathrm{X}}}{\partial V_{\mathrm{X}}^n}}{\max\left(\dfrac{\partial^n I_{\mathrm{X}}}{\partial V_{\mathrm{X}}^n}\right)}. \tag{8.3}$$

The model exhibits a smooth behavior around $V_{\mathrm{DS}} = 0$.

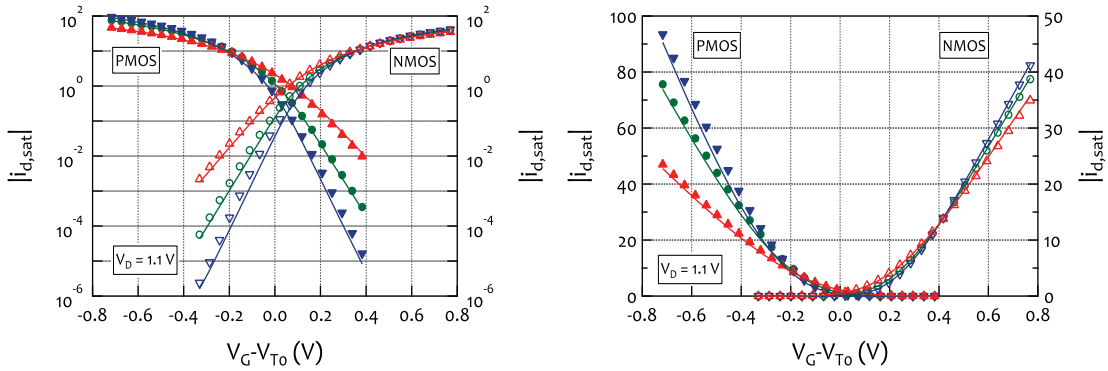
8.2.3 RF Operation

In this part the model is compared against RF measurements of the 40 nm technology. As described in Ch. 2, each RF DUT of this process consists of a number of multi-finger devices in parallel, isolated by a deep buried n-well layer (connected to the ground for nMOS devices). Since increased complexity can affect the simulation speed of a model, the surrounding



(a) $i_{\mathrm{d,lin}}$ - logarithmic scale

(b) $i_{\mathrm{d,lin}}$ - linear scale



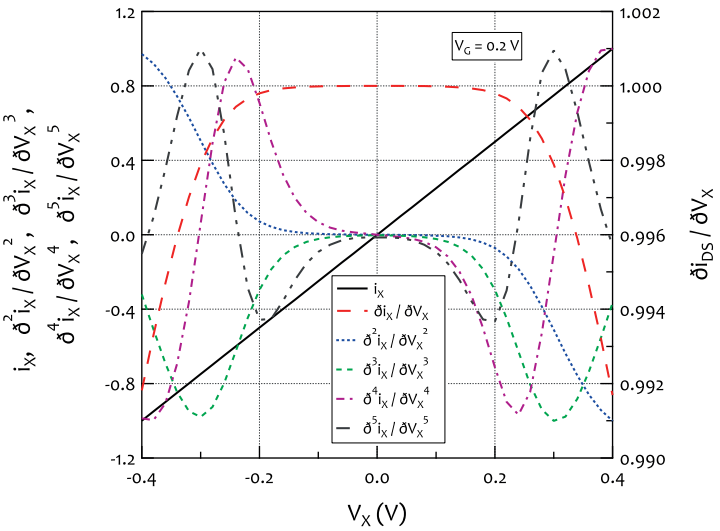
(c) $i_{\mathrm{d,sat}}$ - logarithmic scale

(d) $i_{\mathrm{d,sat}}$ - linear scale

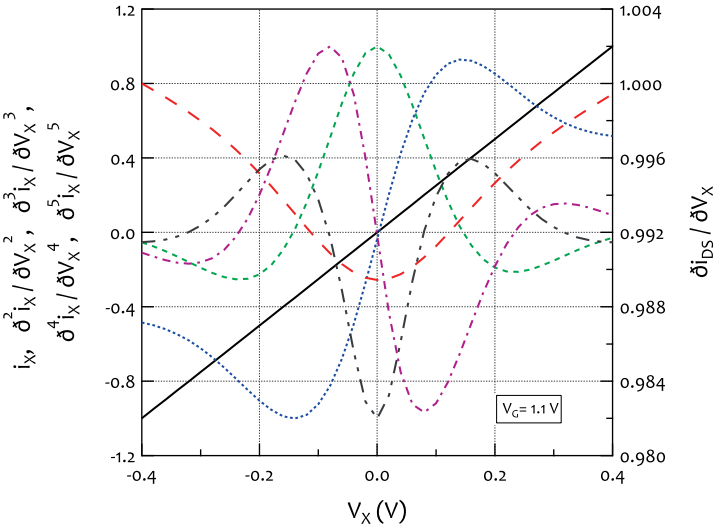
Figure 8.6 – Normalized drain current $|i_{\mathrm{d}}|$ vs. $V_{\mathrm{G}} - V_{\mathrm{T0}}$, at $T = [-40, 25, 250]\ ^{o}C$, for the shortest nMOS and pMOS DC DUTs of a 40 nm CMOS process, with $W = 10$ µm and $L = 40$ nm. (a) $|i_{\mathrm{d}}|$ at $V_{\mathrm{D}} = 0.05$ V (linear mode) in logarithmic scale, (b) $|i_{\mathrm{d}}|$ at $V_{\mathrm{D}} = 0.05$ V (linear mode) in linear scale, (c) $|i_{\mathrm{d}}|$ at $V_{\mathrm{D}} = 1.1$ V (saturation) in logarithmic scale and (d) $|i_{\mathrm{d}}|$ at $V_{\mathrm{D}} = 1.1$ V (saturation) in linear scale.

parasitic network should be kept as simple as possible. A 3-R network was found to be accurate enough to model the substrate network for the CMOS process under study, while it was very important that the parasitic contribution of the isolation layer was accounted for. Although the deep n-well layer that is spreading below the whole device is distributed in nature [118], a simplified model, proposed in [64], was used. A final overview of the RF MOSFET equivalent schematic, that was used to model RF devices of the specific 40 nm process, is presented in Fig. 8.8. Among all the extrinsic components only the resistance R_{iso} and the capacitance C_{iso} of the isolation layer are not provided by the BSIM6 compact model and had to be added externally [72, 86].

During the last years, the interest for ultra-low power RF applications which employ transistors at low-bias conditions has increased [119–121], while applications with operating frequencies above the F_{t} have already been proposed [122–124]. Thus, results from the validation of BSIM6 not only in saturation but also at those regions of very low-power operation, including regions close and above the transit frequency F_{t}, will be presented.



(a) weak-inversion



(b) strong-inversion

Figure 8.7 – Gummel symmetry test (GST) of BSIM6 model at (a) $V_G = 0.2V$ (weak-inversion) and (b) $V_G = 1.1V$ (strong-inversion), using the model card derived for nMOS DC DUTs of a 40 nm CMOS process. $i_X$vs. V_X and its partial derivatives up to 5^{th} degree are shown.

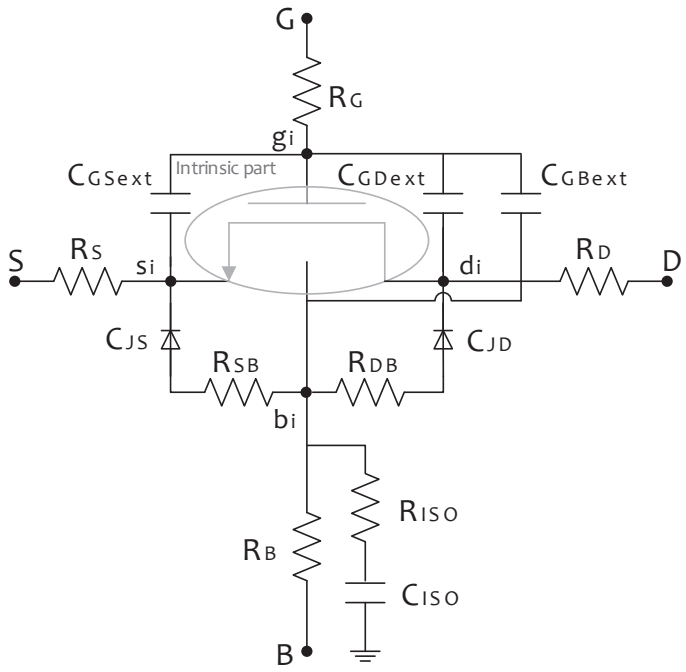
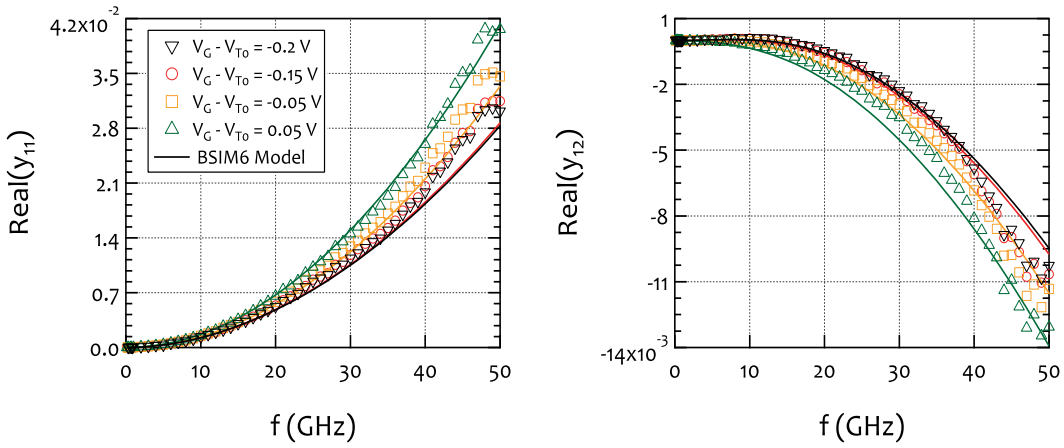


Figure 8.8 – RF MOSFET equivalent schematic including the parasitic contribution of the isolation layer.

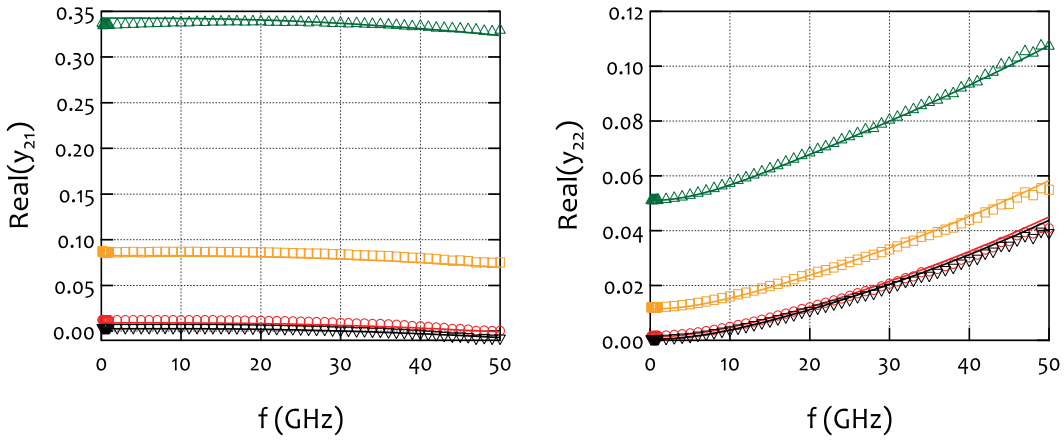
Y-parameters

The RF validation process begins with the study of the de-embedded Y-parameters. The BSIM6 model has been thoroughly validated in different bias conditions [5, 72, 86, 88, 111]. The Y-parameters are normalized to G_{spec} according to Eq. (3.20). In Fig. 8.9 and Fig. 8.10, the model is compared against measurements of the Y-parameters across frequency, in weak- and moderate-inversion at low V_{D} ($V_{\text{D}} = 0.3$ V), for the shortest nMOS RF device, showing an excellent consistency. In addition, in Fig. 8.11 the normalized Y-parameters vs. IC at 20 GHz are presented. Although it is challenging to achieve a good fitting of the Y-parameters over such a wide range of bias points, the model displays sufficient precision, verifying its RF abilities at very low-bias conditions and even for frequencies above the transit frequency of the device.

In Fig. 8.12 and Fig. 8.13, the model is compared against measurements of the Y-parameters across frequency, for different levels of inversion at $V_{\text{D}} = 1.1$ V, for the shortest nMOS RF device. The analytical model presented in Ch. 3 is also included for reference. We see that BSIM6 is representing correctly the Y-parameters, even at the lower RF frequencies of $\Re\{Y_{22}\}$, which is strongly affected by the isolation layer that expands below the DUT and is accounted for in the simulations with BSIM6, but not in the analytical model.



(a) $\Re\{y_{11}\}$ (b) $\Re\{y_{12}\}$

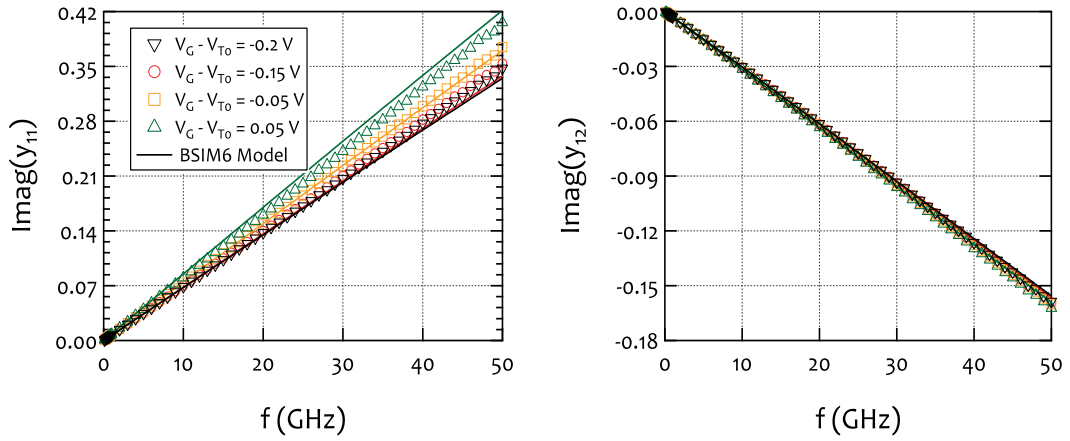


(c) $\Re\{y_{21}\}$ (d) $\Re\{y_{22}\}$

Figure 8.9 – Normalized real part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_G - V_{T0} = [-0.2, -0.15, -0.05, 0.05]$ V and $V_D = 0.3$ V: (a) Real(y_{11}), (b) Real(y_{12}), (c) Real(y_{21}) and (d) Real(y_{22}). Note that at $V_G - V_{T0} = [-0.2, -0.15]$ V the device is in weak-inversion ($IC < 0.1$), while at $V_G - V_{T0} = [-0.05, 0.05]$ V the device is in moderate inversion ($0.1 < IC < 10$).

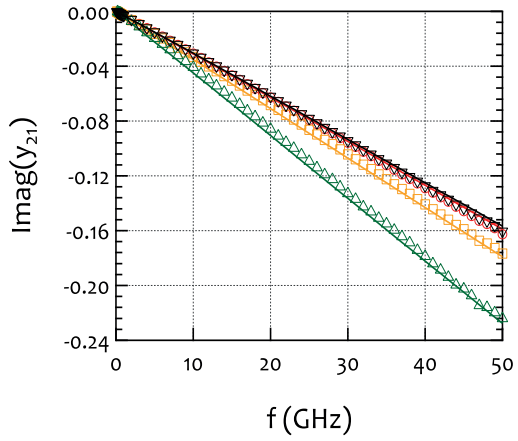
RF Figures-of-Merit

To further evaluate the BSIM6 model at RF, RF FoMs interesting from a designer's point of view are also studied, for two different V_D points, namely $V_D = 0.3$ V and $V_D = 1.1$ V. Especially for $V_D = 1.1$ V that the device operated in saturation the analytical model presented in Ch. 3 is also included for reference. In Fig. 8.14a and Fig. 8.14c, the magnitude of the small-signal current gain $|H_{21}|$ is displayed. We can observe that at low levels of inversion and beyond a specific frequency, $|H_{21}|$ becomes independent of the frequency. This occurs for frequencies above the transit frequency and as a result it does not have an impact on the behavior of F_t, which is the frequency at which $|H_{21}| = 1$. Therefore, even at low-bias conditions F_t can still be

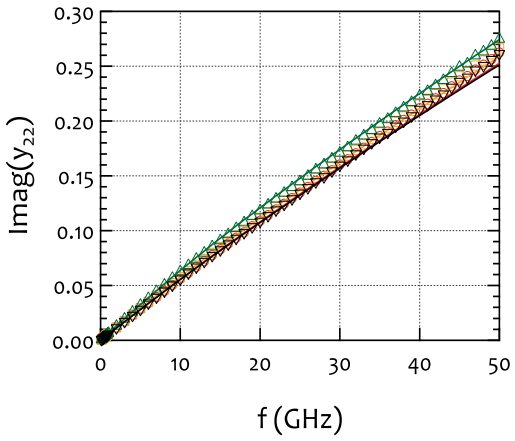


(a) $\Im\{y_{11}\}$

(b) $\Im\{y_{12}\}$



(c) $\Im\{y_{21}\}$



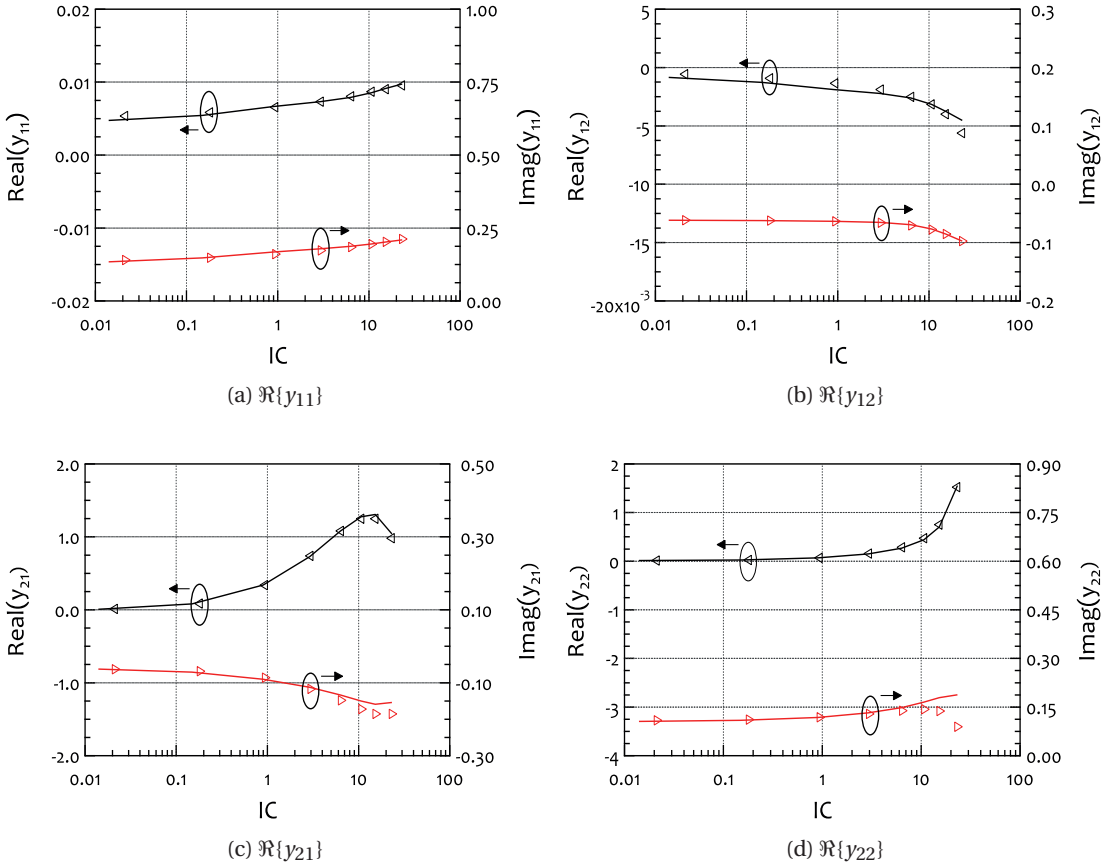
(d) $\Im\{y_{22}\}$

Figure 8.10 – Normalized imaginary part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_G - V_{T0} = [-0.2, -0.15, -0.05, 0.05]$ V and $V_D = 0.3$ V: (a) Imag(y_{11}), (b) Imag(y_{12}), (c) Imag(y_{21}) and (d) Imag(y_{22}). Note that at $V_G - V_{T0} = [-0.2, -0.15]$ V the device is in weak-inversion ($IC < 0.1$), while at $V_G - V_{T0} = [-0.05, 0.05]$ V the device is in moderate inversion ($0.1 < IC < 10$).

calculated using the same approximation as in strong-inversion:

$$F_t = f_{spot} \cdot |H_{21}(f_{spot})| \simeq \frac{1}{2\pi} \cdot \frac{G_m(f_{spot})}{C_{gg}(f_{spot})}, \tag{8.4}$$

where f_{spot} is the frequency at which F_t is calculated. However, especially for low levels of inversion, f_{spot} must be carefully chosen before the point at which $|H_{21}|$ levels off, i.e. in the -20 dB/dec part of the $|H_{21}|$ characteristic, so that Eq. (8.4) provides a correct estimation of the F_t. The transit frequency vs. IC is shown in Fig. 8.14b and Fig. 8.14d. It is worth mentioning that in moderate-inversion the F_t is already some tens of GHz.



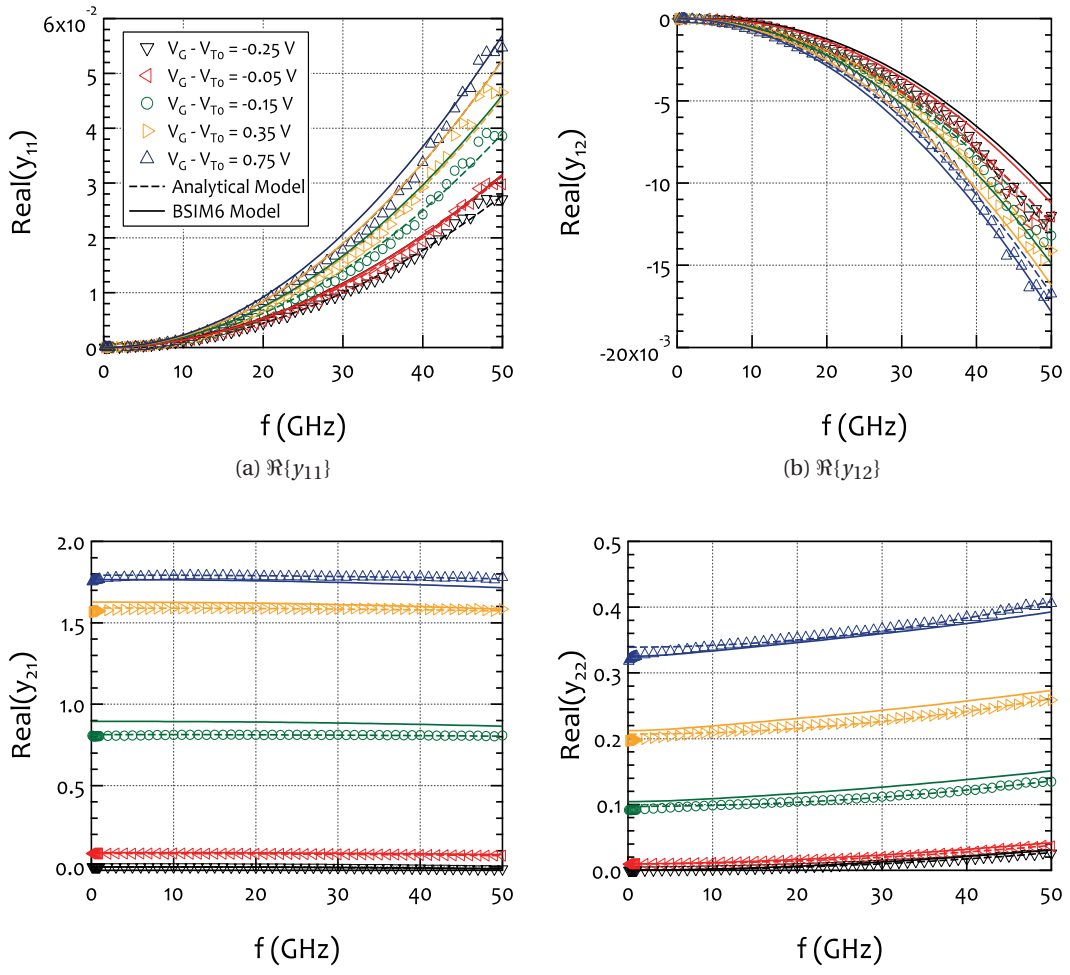
(a) $\Re\{y_{11}\}$ (b) $\Re\{y_{12}\}$

(c) $\Re\{y_{21}\}$ (d) $\Re\{y_{22}\}$

Figure 8.11 – Normalized Y-parameters vs. IC of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at 20 GHz and $V_D = 0.3$ V: (a) y_{11}, (b) y_{12}, (c) y_{21} and (d) y_{22}. The real part corresponds to the left axis whereas the imaginary part to the right.

Fig. 8.15a and Fig. 8.15c show Mason's Unilateral Gain U, which can be calculated as defined in Eq. (3.29). As already discussed in Ch. 3, the slope of U with respect to frequency changes as we move towards lower levels of inversion. The above observation points out that F_{max}, which is the frequency at which $U = 1$, can no longer be calculated as $F_{max} = f_{spot} \cdot \sqrt{U(f_{spot})}$, with f_{spot} being the frequency at which F_{max} is calculated. Instead, F_{max} must be calculated as the frequency at which U, extrapolated with the correct slope, becomes equal to 0 dB, for each operating point. This method was used to obtain F_{max} vs. IC in Fig. 8.15b and Fig. 8.15d.

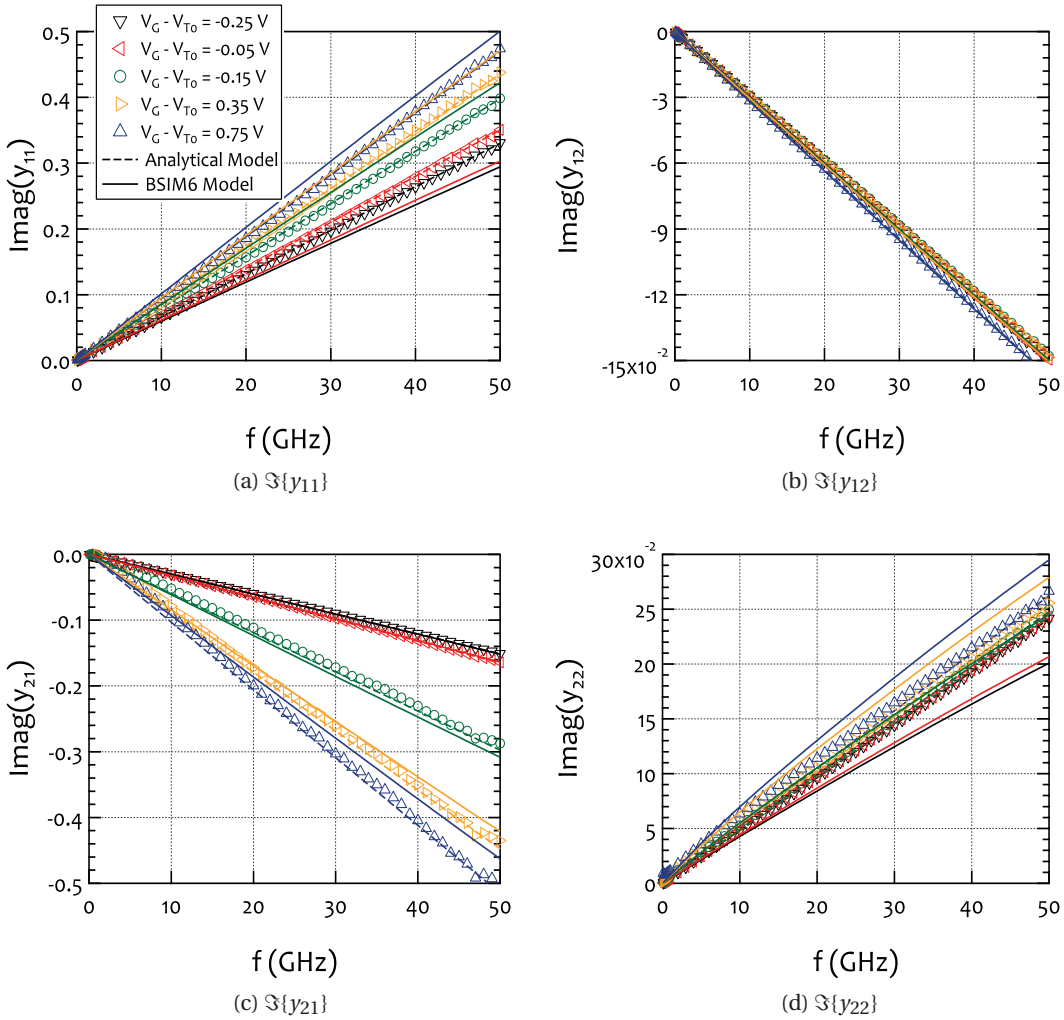
The comparison between the BSIM6 model and the measurements validates the abilities of the model to capture with precision the behavior of nanoscale devices, including at very-low bias conditions, provided that a correct parameter extraction is performed. The model is able to predict the RF operation, which is strongly affected by all the higher order physical phenomena, over more than four decades for RF of current density, across all different levels of inversion and specifically from weak- to strong-inversion.



(a) $\Re\{y_{11}\}$ (b) $\Re\{y_{12}\}$

(c) $\Re\{y_{21}\}$ (d) $\Re\{y_{22}\}$

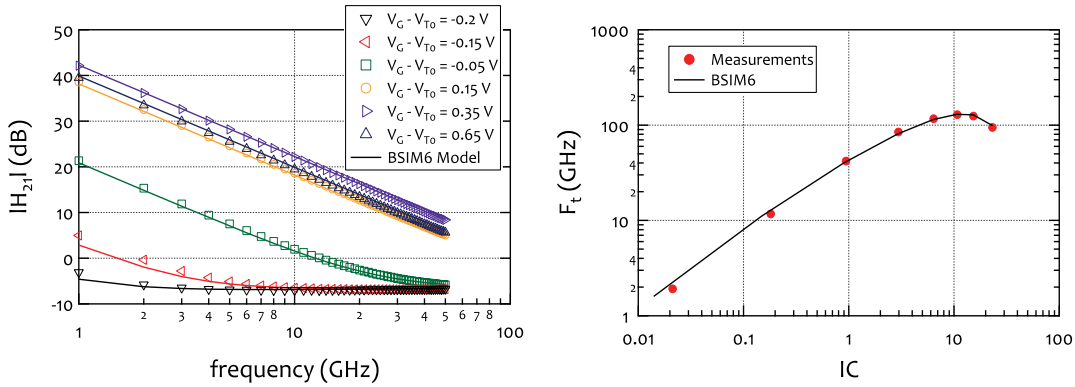
Figure 8.12 – Normalized real part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_G - V_{T0} = [-0.25, -0.05, 0.15, 0.35, 0.75]$ V and $V_D = 1.1$ V: (a) Real(y_{11}), (b) Real(y_{12}), (c) Real(y_{21}) and (d) Real(y_{22}). Note that at $V_G - V_{T0} = -0.25$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = [0.35, 0.75]$ V the device is in strong-inversion ($IC > 10$).



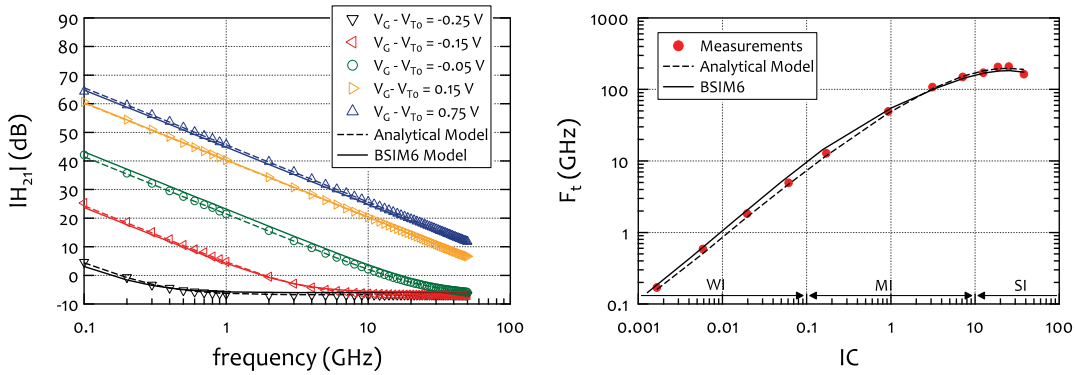
(a) $\Im\{y_{11}\}$ (b) $\Im\{y_{12}\}$

(c) $\Im\{y_{21}\}$ (d) $\Im\{y_{22}\}$

Figure 8.13 – Normalized imaginary part of the Y-parameters vs. frequency of a minimum channel length RF nMOS DUT of a 40 nm CMOS process, with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, at $V_G - V_{T0} = [-0.25, -0.05, 0.15, 0.35, 0.75]$ V and $V_D = 1.1$ V: (a) Imag(y_{11}), (b) Imag(y_{12}), (c) Imag(y_{21}) and (d) Imag(y_{22}). Note that at $V_G - V_{T0} = -0.25$ V the device is in weak-inversion ($IC < 0.1$), at $V_G - V_{T0} = [-0.05, 0.15]$ V the device is in moderate-inversion ($0.1 < IC < 10$) and at $V_G - V_{T0} = [0.35, 0.75]$ V the device is in strong-inversion ($IC > 10$).



(a) $|H_{21}|$ at $V_D = 0.3$ V (b) F_t at $V_D = 0.3$ V



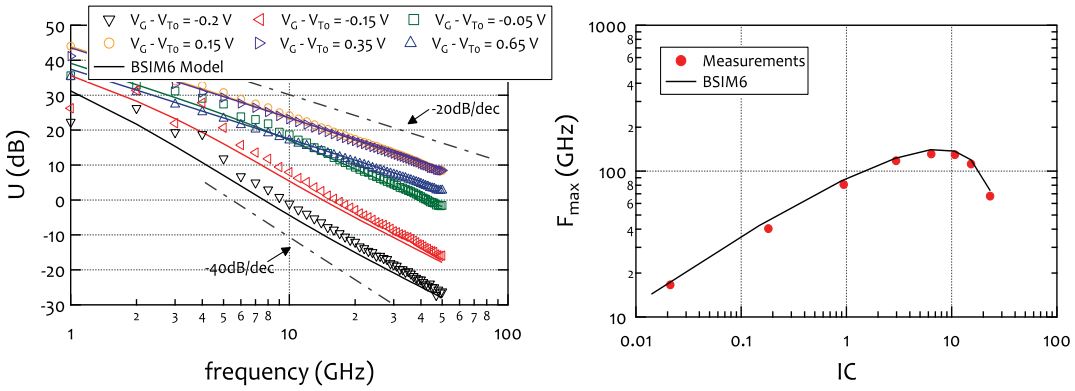
(c) $|H_{21}|$ at $V_D = 1.1$ V (d) F_t at $V_D = 1.1$ V

Figure 8.14 – $|H_{21}|$ vs. frequency and F_t vs. *IC* of a minimum channel length RF nMOS DUT of a 40 nm CMOS process. (a) $|H_{21}|$ vs. frequency at $V_G - V_{T0} = [-0.2...0.65]$ V and $V_D = 0.3$ V, (b) F_t vs. *IC* at $V_G - V_{T0} = [-0.2...0.65]$ V and $V_D = 0.3$ V, (c) $|H_{21}|$ vs. frequency at $V_G - V_{T0} = [-0.25...0.75]$ V and $V_D = 1.1$ V and (d) F_t vs. *IC* at $V_G - V_{T0} = [-0.25...0.75]$ V and $V_D = 1.1$ V.

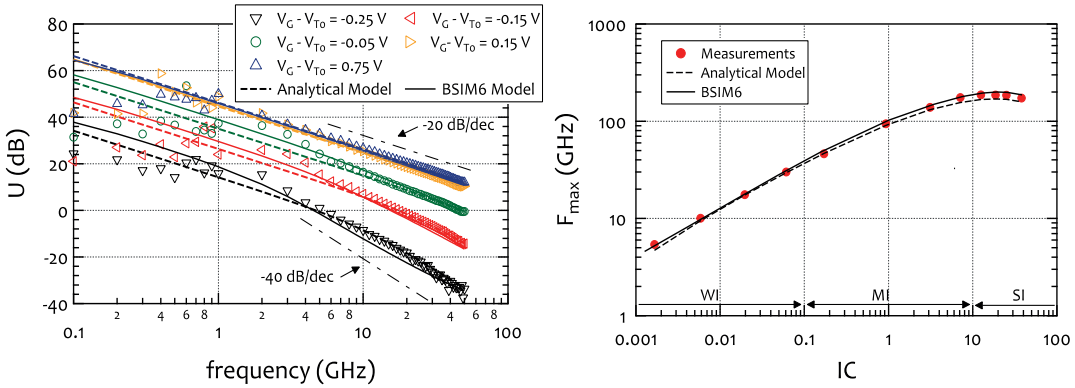
RF Noise

At RF and especially at low bias conditions the noise generated in the device plays an important role in the overall system characteristics. Therefore, accurate modeling of the RF noise is a requirement for low noise, RF IC design. BSIM6 includes all the different noise sources to accurately capture the noise behavior of the device, i.e. flicker noise, channel thermal noise (including induced gate noise), gate current shot noise and thermal noise of resistances.

In order to evaluate the accuracy of the BSIM6 model, the de-embedded RF noise parameters of the RF nMOS DUT, namely NF_{min}, R_n, G_{opt} and B_{opt}, were used. In Fig. 8.16 and Fig. 8.17 the four RF noise parameters of the device are compared with the BSIM6 model versus frequency, for low and high V_D bias, namely, $V_D = [0.3, 1.1]$ V. In saturation, i.e. $V_D = 1.1$ V, the analytical RF noise model presented in Ch. 4 is also included for reference. Since, it is very difficult to carry out RF noise measurements at very-low current densities the noise parameters are displayed only in moderate- and strong-inversion and not in weak-inversion. Besides, in



(a) U at $V_D = 0.3$ V (b) F_{max} at $V_D = 0.3$ V



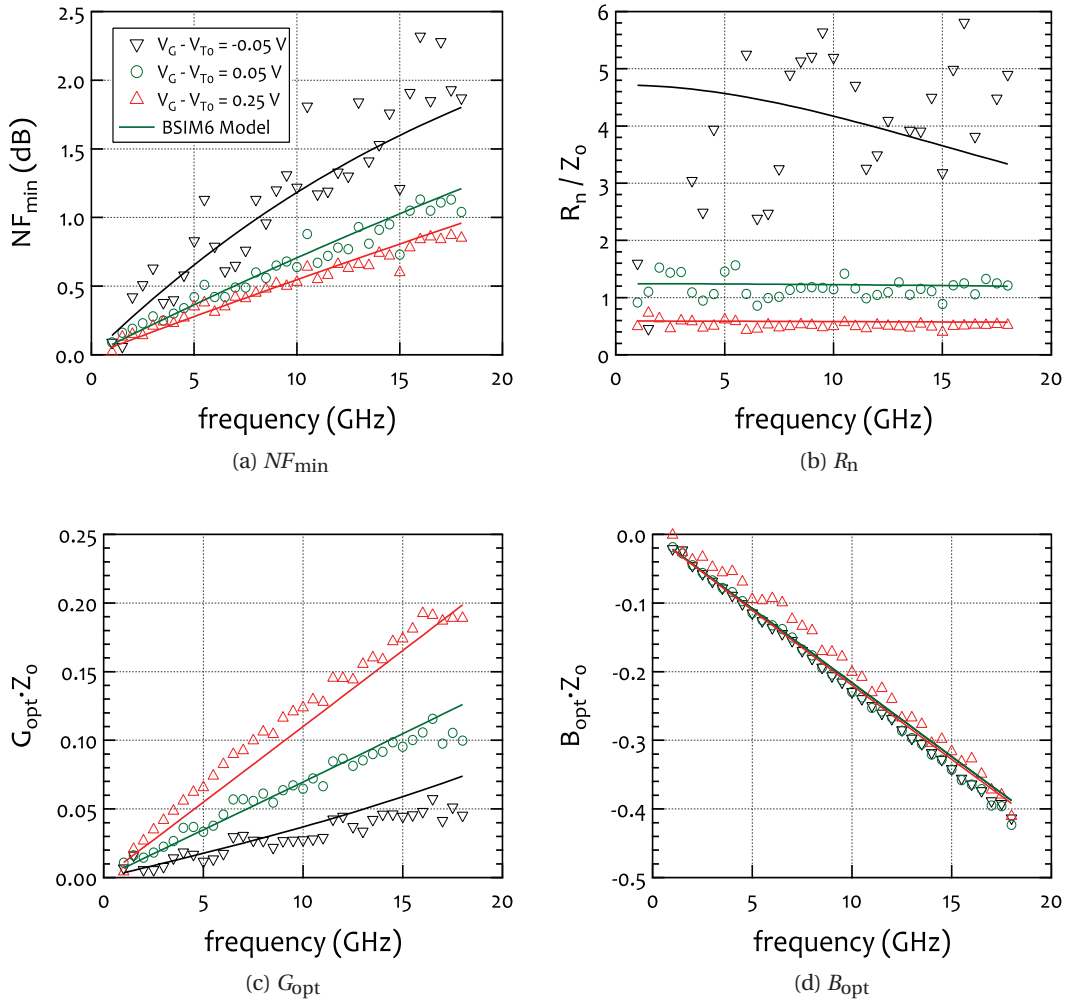
(c) U at $V_D = 1.1$ V (d) F_{max} at $V_D = 1.1$ V

Figure 8.15 – U vs. frequency and F_{max} vs. IC of a minimum channel length RF nMOS DUT of a 40 nm CMOS process. (a) U vs. frequency at $V_G - V_{T0} = [-0.2...0.65]$ V and $V_D = 0.3$ V, (b) F_{max} vs. IC at $V_G - V_{T0} = [-0.2...0.65]$ V and $V_D = 0.3$ V, (c) U vs. frequency at $V_G - V_{T0} = [-0.25...0.75]$ V and $V_D = 1.1$ V and (d) F_{max} vs. IC at $V_G - V_{T0} = [-0.25...0.75]$ V and $V_D = 1.1$ V.

Fig. 8.18, NF_{min} and R_n, are plotted with respect to IC, and although the noise measurements present a great variance across frequency, the model is still very close to the measurements. From Fig. 8.18a and Fig. 8.18c, we observe that the minimum NF_{min} is achieved in the higher levels of moderate inversion for both V_D biases. The RF noise characteristics verify the accuracy with which BSIM6 predicts the noisy behavior of the device from moderate- to strong-inversion region.

8.3 Validation against a 28 nm CMOS Process

Similar to the evaluation of the BSIM6 model against the 40 nm CMOS process, in this section the abilities of the model are verified even for the lowest standard bulk CMOS node, i.e. the 28 nm HK-MG bulk CMOS technology. For this part of the validation, we focus on the CV and DC operation of nMOS DC devices and we highlight particularly the model's length scaling



(a) NF_{min} (b) R_{n}

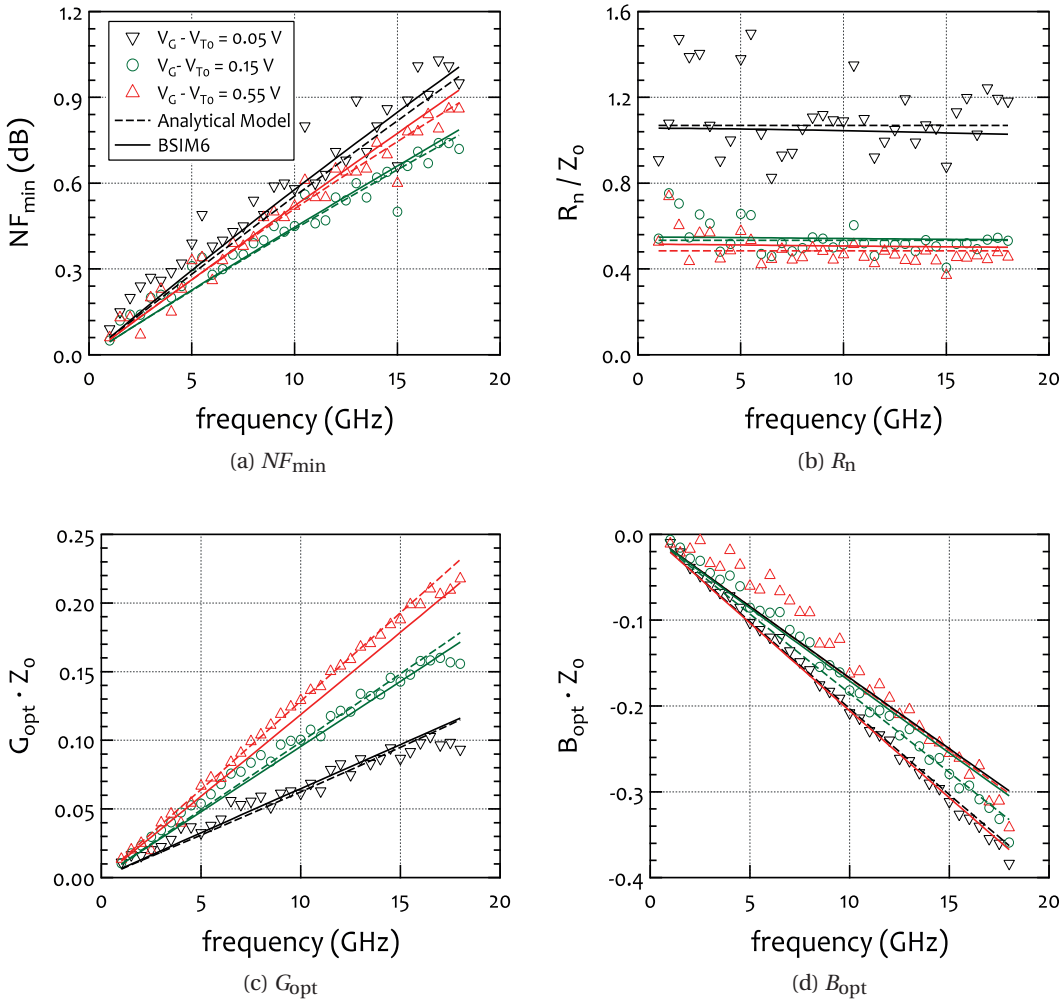
(c) G_{opt} (d) B_{opt}

Figure 8.16 – The four RF noise parameters vs. frequency of a minimum channel length RF nMOS DUT with $M = 6$, $N_{\text{f}} = 10$, $W = 2$ µm and $L = 40$ nm, at $V_{\text{D}} = 0.3$ V and $V_{\text{G}} - V_{\text{T0}} = [-0.05, 0.05, 0.25]$ V, with $Z_0 = 50\,\Omega$: (a) NF_{min}, (b) R_{n}, (c) G_{opt} and (d) B_{opt}.

abilities. Given the fact that a broad evaluation of the model has already been carried out for the 40 nm CMOS technology, for the 28 nm process only the most indicative results are demonstrated.

8.3.1 CV Operation

In Fig. 8.19, the normalized capacitances vs. $V - V_{\text{T0,lin}}$ of wide ($W = 10$ µm) nMOS DC devices with different channel lengths, from the long to short, are shown. More specifically, Fig. 8.19a shows the normalized total gate capacitance c_{gg} vs. $V_{\text{G}} - V_{\text{T0,lin}}$ when all the other device terminals, i.e. D, S, B, are connected to the ground, Fig. 8.19b shows the normalized gate to bulk capacitance c_{gb} vs. $V_{\text{GB}} - V_{\text{T0,lin}}$ when the D and S terminals are connected to the ground



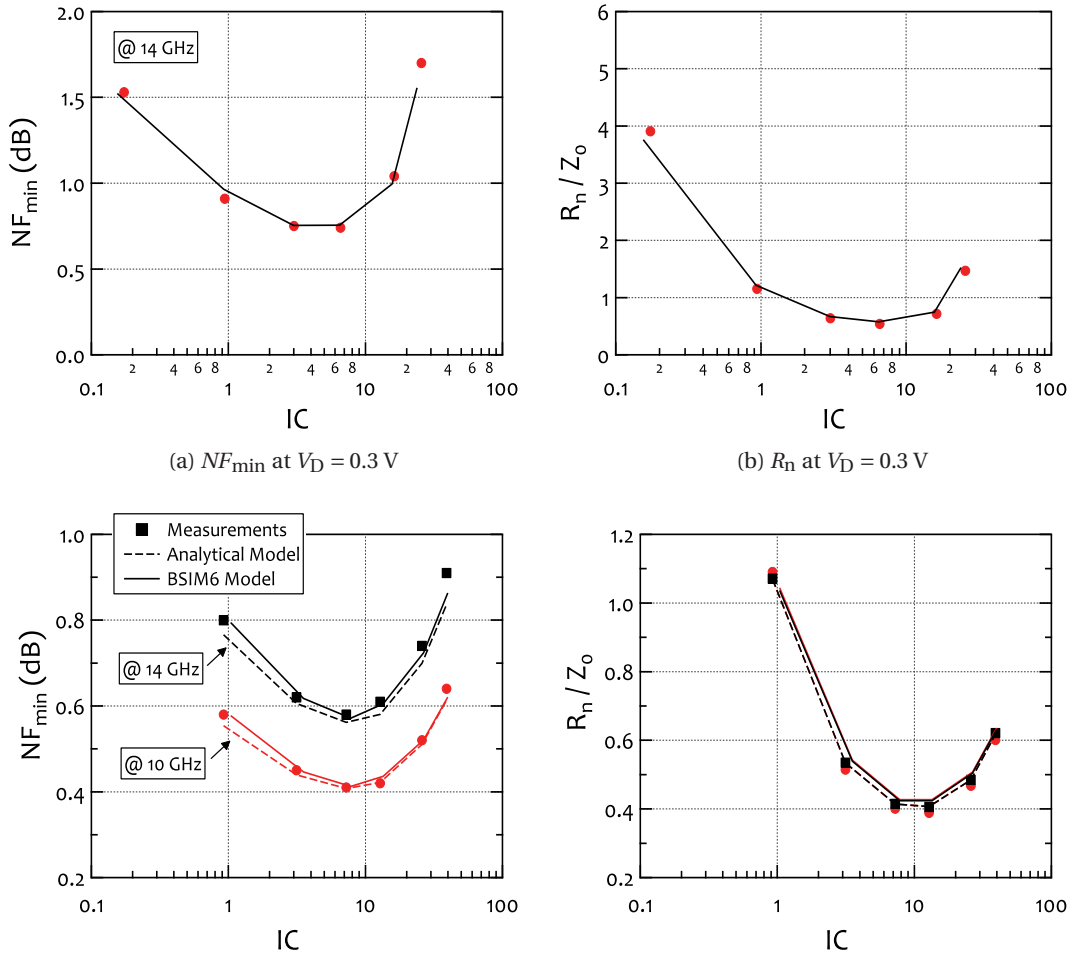
(a) NF_{min} (b) R_{n}

(c) G_{opt} (d) B_{opt}

Figure 8.17 – The four RF noise parameters vs. frequency of a minimum channel length RF nMOS DUT with $M = 6$, $N_{\text{f}} = 10$, $W = 2$ µm and $L = 40$ nm, at $V_{\text{D}} = 1.1$ V and $V_{\text{G}} - V_{\text{T0}} = [0.05, 0.15, 0.55]$ V, with $Z_0 = 50\,\Omega$: (a) NF_{min}, (b) R_{n}, (c) G_{opt} and (d) B_{opt}.

and Fig. 8.19c shows the normalized gate to bulk capacitance $c_{\text{gd,s}}$ vs. $V_{\text{GD,S}} - V_{\text{T0,lin}}$ when the voltage applied to the B terminal is $V_{\text{B}} = V_{\text{GD,S}}$. When $C_{\text{GD,S}}$ is measured, choosing to apply to the B terminal a bias that follows the $V_{\text{GD,S}}$ helps to be able to collect the AC signal at any $V_{\text{GD,S}}$. In this way, at any time, a zero voltage drop is applied to the diodes C_{JD} and C_{JS} and thus the limitation to stop the $V_{\text{GD,S}}$ biasing before the voltage drop at the diodes is equal to 0.7 V does not exist. The capacitances are normalized to $C_{\text{ox}} \cdot W_{\text{tot}} \cdot L$ according to Eq. (3.18), while $V_{\text{T0,lin}}$ represents the threshold voltage of the long channel device when there is no body-effect, i.e. $V_{\text{SB}} = 0$ V.

When observing closer the three CV plots, a few remarks concerning the impact of different short channel effects on the capacitance arise. In Fig. 8.19a and Fig. 8.19c, in depletion region



(a) $NF_{\min}$ at $V_D = 0.3$ V

(b) R_n at $V_D = 0.3$ V

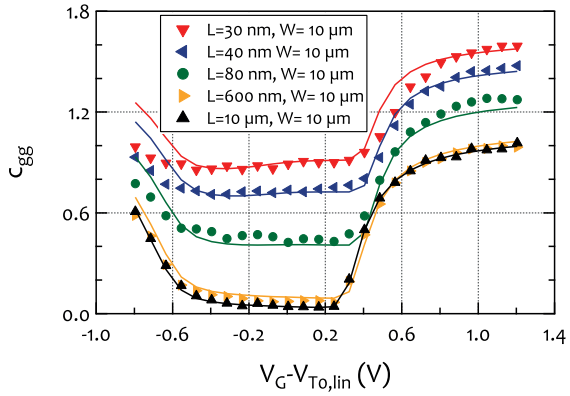
(c) $NF_{\min}$ at $V_D = 1.1$ V

(d) R_n at $V_D = 1.1$ V

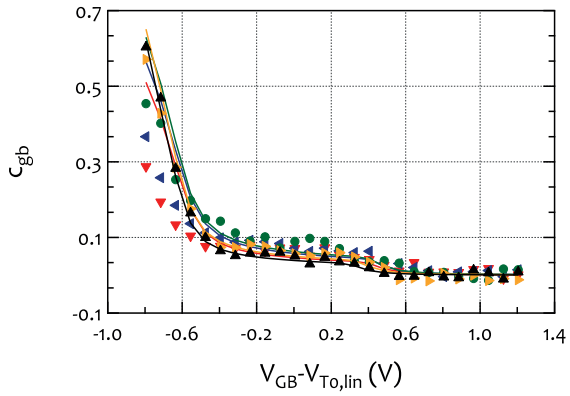
Figure 8.18 – $NF_{\min}$ and R_n vs. IC, of a minimum channel length RF nMOS DUT with $M = 6$, $N_f = 10$, $W = 2$ µm and $L = 40$ nm, with $Z_0 = 50\ \Omega$. (a) $NF_{\min}$ at $V_D = 0.3$ V and $f = 14$ GHz, (b) R_n at $V_D = 0.3$ V and $f = 14$ GHz, (c) $NF_{\min}$ at $V_D = 1.1$ V and $f =$ [10 GHz, 14 GHz] and (d) Rn at $V_D = 1.1$ V and $f =$ [10 GHz, 14 GHz]. It is clearly observed that the minimum values of $NF_{\min}$ and R_n are achieved in the onset between moderate and strong-inversion.

(where there is no inversion charge), we see an increase of the capacitances as we move from the long channel DUT to the shorter ones. This is the impact of the overlap and fringing capacitances. From the same plots, in the low part of the inversion region, we can see how the quantum mechanical effects impact the slope of the characteristics, e.g. the slope of c_{gg} between $V_G - V_{T0,lin} \approx 0.2$ to 0.5 V is steeper for the long channel device when compared to the the slope of c_{gg} between $V_G - V_{T0,lin} \approx 0.4$ to 0.6 V of the shortest one. Finally, from Fig. 8.19b in depletion region we can see the impact of the change in the value of the effective channel doping.

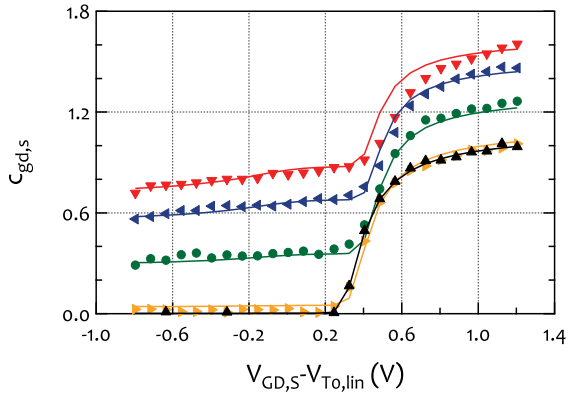
The comparison of the BSIM6 model against CV measurements of the 28 nm CMOS techno-



(a) c_{gg}



(b) c_{gb}



(c) c_{gd}

Figure 8.19 – Normalized capacitances for nMOS DC DUTs of a 28 nm CMOS process, with $W = 10$ µm and $L = [30\text{n}, 40\text{n}, 80\text{n}, 600\text{n}, 10\mu]$ m. (a) c_{gg} vs. $V_G - V_{T0,lin}$ at $V_S = V_D = V_B = 0$ V, (b) c_{gb} vs. $V_{GB} - V_{T0,lin}$ at $V_S = V_D = 0$ V and (c) c_{gd} vs. $V_{GD,S} - V_{T0,lin}$ at $V_B = V_{GD,S}$.

logy shows that the model is able to represent with sufficient accuracy the dynamic behavior of the transistors of this nanoscale node.

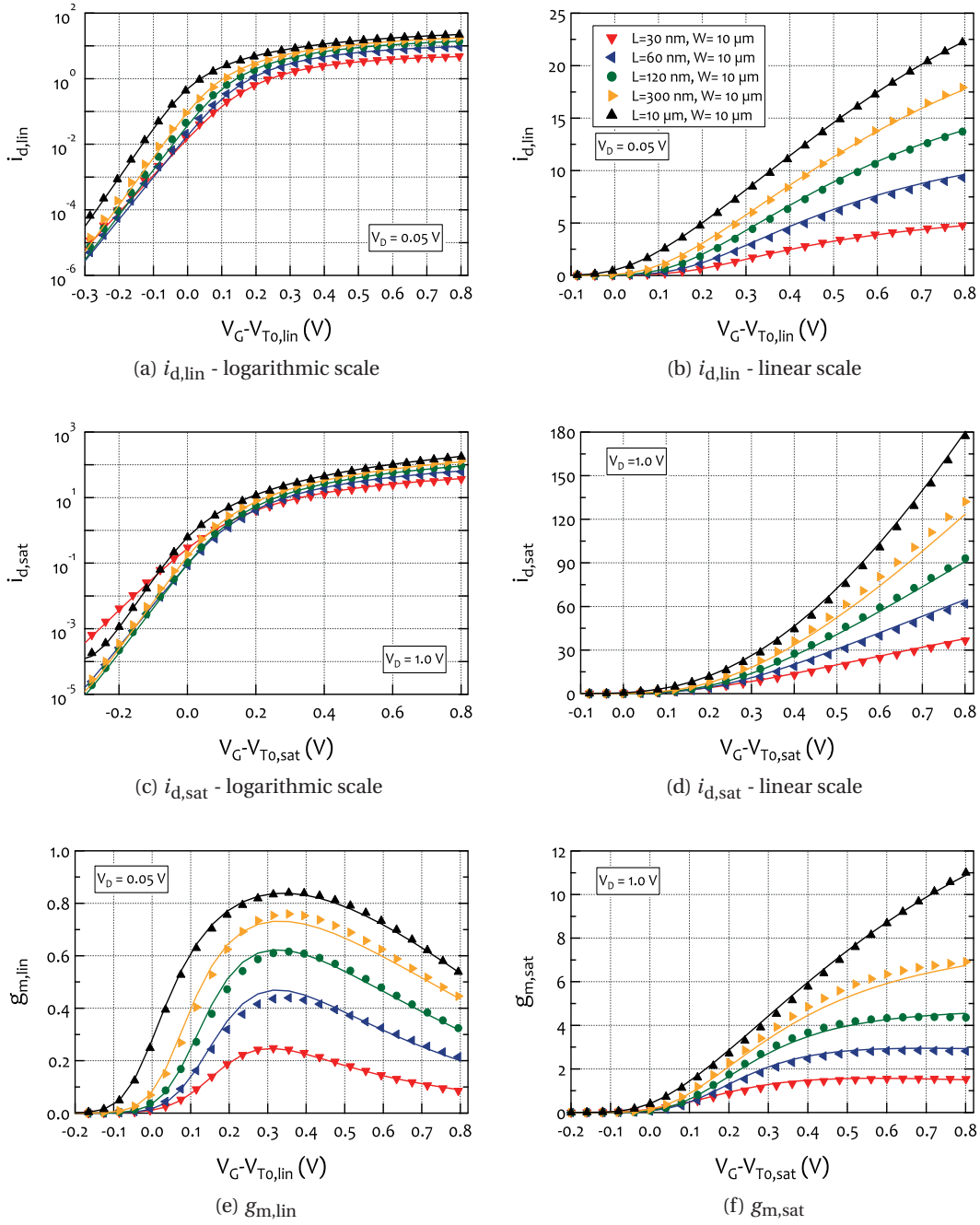
8.3.2 DC Operation

In the plots that are presented in this section, different IV characteristics are demonstrated. More specifically, Fig. 8.20 shows the i_d and g_m vs. $V_G - V_{T0}$, in linear mode and saturation, while in Fig. 8.21 the i_d and g_{ds} vs. V_D, in weak- and strong-inversion are presented. V_{T0} is the threshold voltage of the specified region, while all the results are also normalized according to the relations presented in Table 6.1. The different characteristics correspond to wide (W = 10 μm) nMOS DC devices with different channel lengths, from long to short. The model can capture the behavior of all the DUTs precisely, accounting for all the second order effects. To further demonstrate the abilities of the model, the sub-threshold slope SS, the normalized threshold voltage v_{tb} ($v_{tb} = V_{TB}/V_{T0,long}$, with $V_{T0,long}$ being the threshold voltage of the long channel device in the defined region of operation when $V_{SB} = 0$ V) and the normalized maximum current max(i_d) vs. L, in both linear mode and saturation, for different V_B biases, are presented in Fig. 8.22. The BSIM6 model is once again proving its excellent scalability and its suitability for advanced nanoscale CMOS technologies.

8.4 Conclusions

The comparison between the BSIM6 model and the measurements validates the abilities of the model to capture with precision the behavior of nanoscale devices, both nMOS and pMOS, including at very-low bias conditions, provided that a correct parameter extraction is performed. Despite the complexity of all the higher order effects that appear in nanoscale devices, the model is able to predict the static, dynamic and RF operation (including RF noise), over more than seven decades for DC and three decades for RF of current density, across all different levels of inversion and specifically from weak- to strong-inversion.

Concerning the RF operation, using a simple extrinsic parasitic network, the model captures the RF behavior accurately for a wide frequency range, even beyond F_t. In addition, the model represents precisely the limits of the technology, e.g. the slope change in the Unilateral gain, which can prove critical when designing close to these regions. The presented results demonstrate that BSIM6 is very well-suited for the design of ultra-low power analog/RF IC.



(a) $i_{d,lin}$ - logarithmic scale

(b) $i_{d,lin}$ - linear scale

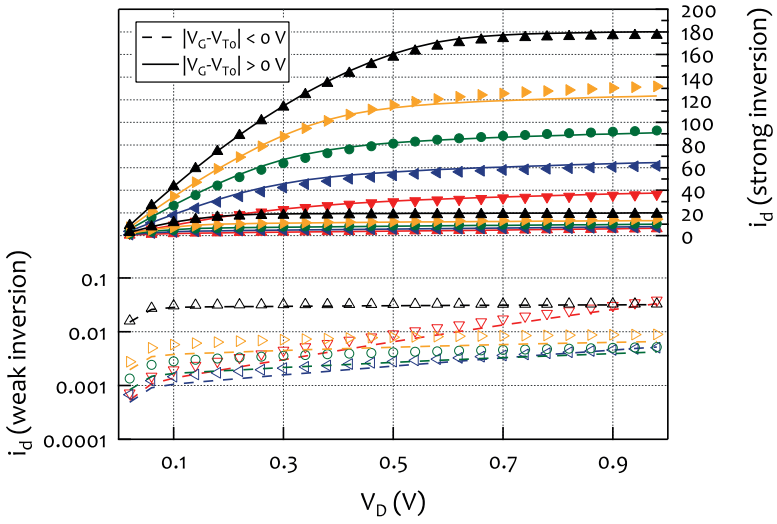
(c) $i_{d,sat}$ - logarithmic scale

(d) $i_{d,sat}$ - linear scale

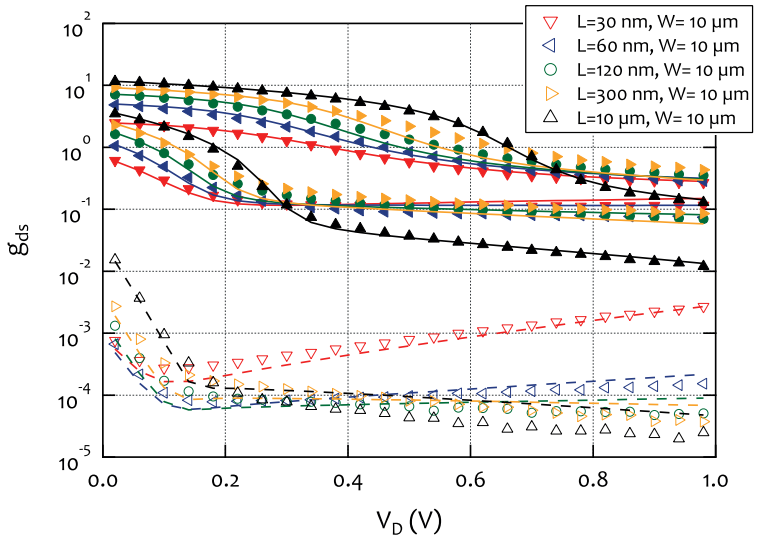
(e) $g_{m,lin}$

(f) $g_{m,sat}$

Figure 8.20 – Normalized drain current i_d and gate transconductance g_m vs. $V_G - V_{T0}$, for nMOS DC DUTs, with $W = 10$ µm and $L = [30n, 60n, 120n, 300n, 10\mu]$ m, of a 28 nm CMOS process. (a) $i_{d,lin}$ vs. $V_G - V_{T0,lin}$ at $V_D = 0.05$ V (linear mode), in logarithmic scale, (b) $i_{d,lin}$ vs. $V_G - V_{T0,lin}$ at $V_D = 0.05$ V (linear mode), in linear scale, (c) $i_{d,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.0$ V (saturation), in logarithmic scale, (d) $i_{d,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.0$ V (saturation), in linear scale, (e) g_m vs. $V_G - V_{T0,lin}$ at $V_D = 0.05$ V (linear mode) and (f) $g_{m,sat}$ vs. $V_G - V_{T0,sat}$ at $V_D = 1.0$ V (saturation).

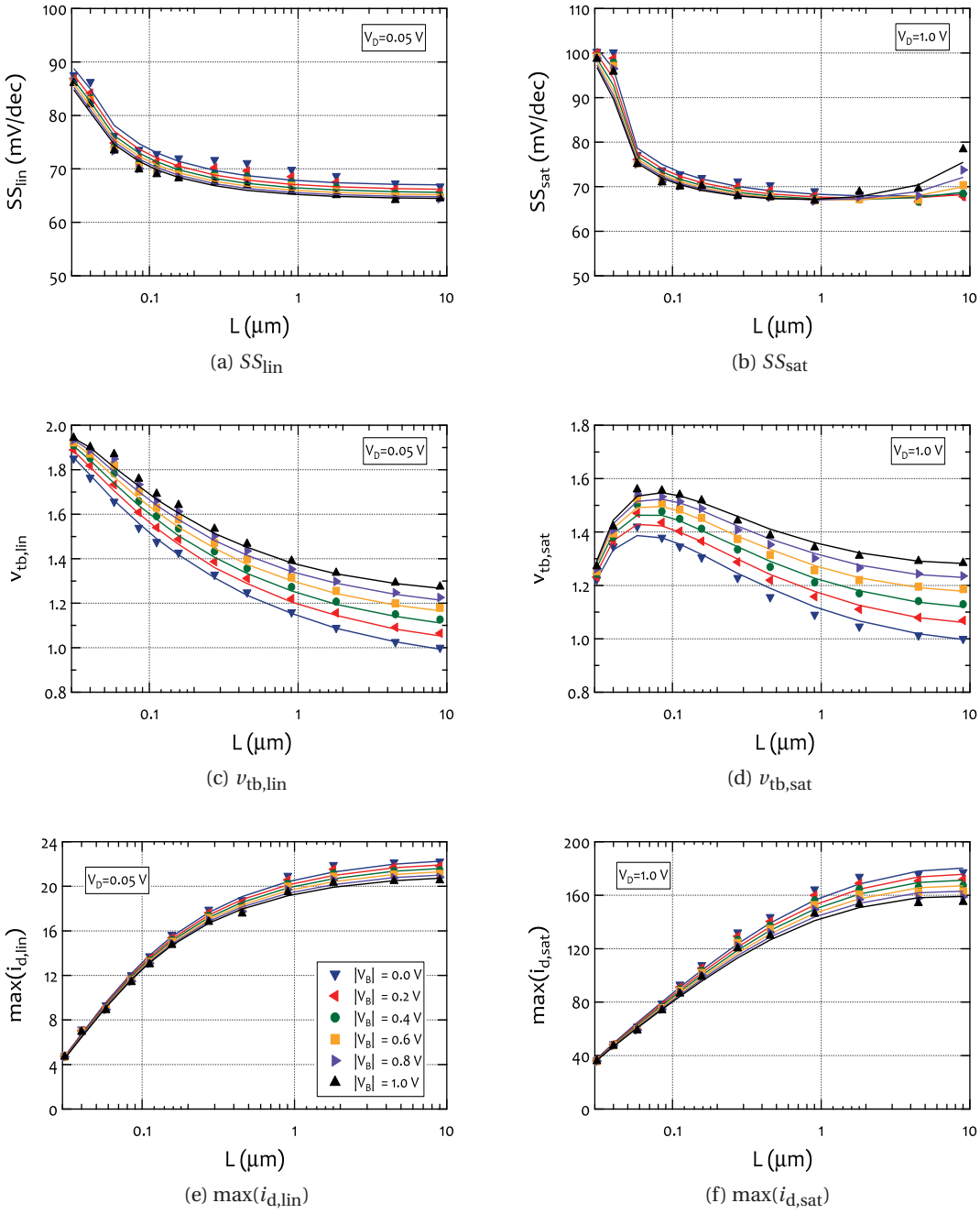


(a) i_d



(b) g_{ds}

Figure 8.21 – Normalized drain current i_d and its derivative g_{ds} vs. V_D for $V_G - V_{T0} = -0.1$ V (weak-inversion) and $V_G - V_{T0} = [0.26, 0.8]$ V (strong-inversion) at $V_S = V_D = V_B = 0$ V, for nMOS DC DUTs, with $W = 10$ µm and $L = [30n, 60n, 120n, 300n, 10\mu]$ m, of a 28 nm CMOS process. (a) i_d and (b) g_{ds}.



(a) SS_{lin} (b) SS_{sat}

(c) $v_{\mathrm{tb,lin}}$ (d) $v_{\mathrm{tb,sat}}$

(e) $\max(i_{\mathrm{d,lin}})$ (f) $\max(i_{\mathrm{d,sat}})$

Figure 8.22 – Length Scaling of: (a) SS_{lin} , (b) SS_{sat}, (c) $v_{\mathrm{tb,lin}}$, (d) $v_{\mathrm{tb,sat}}$, (e) $\max(i_{\mathrm{d,lin}})$, (f) $\max(i_{\mathrm{d,sat}})$ for nMOS DC DUTs of a 28 nm CMOS process, at $V_{\mathrm{D}} = 0.05$ V (linear mode) and $V_{\mathrm{D}} = 1.0$ V (saturation) and for $|V_{\mathrm{SB}}| = [0.0, 0.2, 0.4, 0.6, 0.8, 1.0]$ V.

Part III Analytical Modeling of FoMs

9 Device Level Figures-of-Merit as Design Guidelines

9.1 Introduction

The necessity of both compact and analytical models for the design process is unquestionable. The former are the built-in tools in simulators without which the IC design cannot be carried out easily and the latter provide us with a deep understanding of the behavior of the devices; an important step that should precede any design activity. And, although both types of models are used widely by circuit designers, they rarely enjoy a treatment as anything more than a black box.

Nevertheless, analog/RF IC design is a demanding and complex task which requires from the designer to identify the optimum choices according to the degrees of freedom available and at the same time achieve the highest performance in metrics such as gain, current efficiency, bandwidth, noise, and linearity [125]. On top of that, the “changes” imposed by the aggressive downscaling of CMOS technology (reduction of voltage supply, increase in the leakage current, profound variability etc.) complicate further the design of analog circuits [126], especially at RF frequencies where parasitics start to dominate. Apparently, any guidance to navigate within this multi-variable design space would be valuable.

In the last years, there is an increased interest in the concept of the inversion coefficient IC as the main design parameter even for very advanced technologies [127, 128]. The IC based design methodology [39] can prove to be especially useful for the design of low-power analog/RF circuits, where the operating point is pushed from the traditional strong-inversion region towards the moderate- or even the weak-inversion.

In this "design-oriented" chapter we present simple analytical expressions which may not account for all the physical phenomena present in nanoscale technologies but are able to predict certain aspects of the behavior of the MOS transistor with sufficient accuracy. Their simplicity makes them an attractive option for designers, who would like to have a first estimation of their design variables before they turn to a circuit simulator. Different FoMs

are modeled across all levels of inversion in **saturation**, using only very few parameters. A very easy and handy procedure for the extraction of these parameters from measurements is additionally presented. The models are then compared against measurements of two commercial bulk CMOS processes, namely, 28 nm and 40 nm, and with the BSIM6 compact model, focusing on short-channel devices. At last, different aspects on the design of a simple single-MOS CS amplifier, which is chosen as case study based on the *IC* design methodology, are discussed.

9.2 Modeling the Gate Transconductance - G_{m}

One of the most important small-signal parameters of a MOSFET is the gate transconductance G_{m} defined as:

$$G_{\mathrm{m}} = \frac{\partial I_{\mathrm{D}}}{\partial V_{\mathrm{G}}}. \tag{9.1}$$

Not only is it a significant characteristic of the device, but also several other design metrics such as $G_{\mathrm{m}}/I_{\mathrm{D}}$, G_{ds}, A_{v}, F_{t}, F, etc. strongly depend on it. As a result, modeling accurately the G_{m} is important if not mandatory. The gate transconductance can be expressed by the formula [39]:

$$G_{\mathrm{m}} = \frac{G_{\mathrm{ms}}}{n} = \frac{g_{\mathrm{ms}} \cdot G_{\mathrm{spec}}}{n} = \frac{g_{\mathrm{ms}} \cdot I_{\mathrm{spec}}}{n \cdot U_{\mathrm{T}}} = \frac{g_{\mathrm{ms}} \cdot I_{\mathrm{spec}_{\square}} \cdot \dfrac{W}{L}}{n \cdot U_{\mathrm{T}}}. \tag{9.2}$$

9.2.1 Modeling the Normalized Source Transconductance - g_{ms}

In advanced nanoscale CMOS devices the impact of velocity saturation is prominent. Velocity saturation effect appears when the longitudinal electric field E_{x} within the device increases beyond a certain value, called the critical electric field E_{c}, above which the drift velocity of the carriers, v_{drift}, starts to saturate to a maximum value v_{sat} [39]. The normalized source conductance in saturation, including the effect velocity saturation is given by [88, 129]:

$$g_{\mathrm{ms}} = \frac{G_{\mathrm{ms}}}{G_{\mathrm{spec}}} = \frac{\sqrt{\lambda_{\mathrm{c}}^2\, IC^2 + 2\,\lambda_{\mathrm{c}}\, IC + 4\, IC + 1} - 1}{\lambda_{\mathrm{c}}^2\, IC + \lambda_{\mathrm{c}} + 2}, \tag{9.3}$$

where λ_{c} is the velocity saturation parameter. λ_{c} depends on the length of the device through the relation:

$$\lambda_{\mathrm{c}} = \frac{L_{\mathrm{sat}}}{L}, \tag{9.4}$$

with $L_{\mathrm{sat}} = 2\,\mu_0\, U_{\mathrm{T}} \,/\, v_{\mathrm{sat}}$ being the length of the part of the channel that the velocity of the drift carriers is saturated, where μ_0 is the low-field mobility constant [39]. λ_{c} tends to zero for

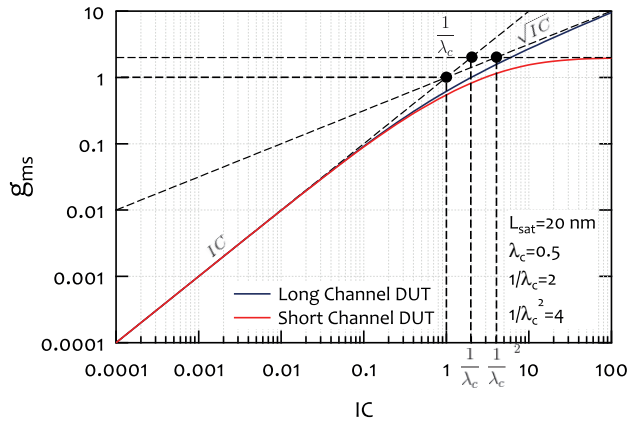


Figure 9.1 – g_{ms} vs. IC in saturation for a short and a long-channel device, having L =40 nm and L =10 µm, respectively. The g_{ms} of the short-channel device saturates at a value equal to $1/\lambda_c$. The velocity saturation effect is strongly limiting the short channel performance, especially in comparison with the long-channel, above $IC = 1/\lambda_c{}^2$.

long-channel devices, while it can reach a value of up to 0.7 for a 30 nm MOSFET. The value of λ_c actually shows the part of the channel length that velocity saturation effect prevails, e.g. a value of $\lambda_c = 0.7$ means that in 70% of the device channel length the drift carriers cannot increase their velocity any further, even if V_D is increased.

In Fig. 9.1, an example of the source transconductance g_{ms} vs. IC for a short (L =40 nm) and a long (L =10 µm) channel device, is presented. The value of L_{sat} was chosen to be 20 nm, meaning that for the long device $\lambda_c \approx 0$, while for the short-channel device $\lambda_c = 0.5$. From the graph we see that in weak-inversion g_{ms} is proportional to IC for both the devices. However, when moving to strong-inversion, the long-channel device demonstrates a g_{ms} proportional to $\sqrt{IC}$, whereas for the short-channel device the g_{ms} saturates to a value equal to $1/\lambda_c$. The strong-inversion asymptotes for both devices can be obtained through Eq. (9.3). For the long-channel device, where there is no velocity saturation effect, we assume $\lambda_c \simeq 0$ and $IC \gg 1$:

$$g_{ms}\Big|_{\text{SI asymptote w/o. VS}} = \sqrt{IC} = \sqrt{i_d}, \tag{9.5}$$

knowing that the normalized drain current in **saturation** is [39]:

$$i_d = \frac{I_{D_{sat}}}{I_{spec}} = IC. \tag{9.6}$$

To calculate the asymptote for the short-channel device, where the velocity saturation effect dominates, we assume $\lambda_c\, IC \gg 1$:

$$g_{ms}\Big|_{\text{SI asymptote w. VS}} = \frac{1}{\lambda_c}. \tag{9.7}$$

The velocity saturation parameter proves to be useful when it comes to understand at which inversion level the velocity saturation effect starts to be important. In Fig. 9.1, we see that the weak- and strong-inversion asymptotes for the short-channel device cross at $IC = 1/\lambda_c$, meaning that for an IC value above $1/\lambda_c$ the velocity saturation effect emerges. In addition, we see that the strong-inversion asymptotes of both devices cross at $IC = 1/{\lambda_c}^2$, so for any value of the IC beyond that point the velocity saturation effect becomes significant and affects strongly the behavior of short-channel devices. Actually, $IC = 1/\lambda_c^2$ can serve as a critical inversion coefficient [130] to define the onset of velocity saturation. Therefore, an effective inversion coefficient that includes the effect of velocity saturation [125, 129] can be defined as:

$$IC_{\text{eff}} = IC \cdot \left(1 + \frac{IC}{4IC_{\text{crit}}}\right), \tag{9.8}$$

with $IC_{\text{crit}} = 1/(\lambda_c^2)$.

9.2.2 Extraction of Parameters - n, I_{spec}, λ_c

Combining Eq. (9.2), Eq. (9.3) and Eq. (9.4) we observe that G_m can be modeled with the use of only three parameters namely, n, $I_{\text{spec}_\square}$ and L_{sat}. In order to facilitate the extraction procedure the transconductance efficiency G_m/I_D will be used.

Modeling the Normalized Transconductance Efficiency - g_{ms}/i_d

The transconductance efficiency G_m/I_D can be normalized as follows:

$$\frac{G_m}{I_D} = \frac{g_m \cdot G_{\text{spec}}}{i_d \cdot I_{\text{spec}}} = \frac{g_m}{i_d \cdot U_T} = \frac{g_{ms}}{i_d}\frac{1}{n \cdot U_T}. \tag{9.9}$$

In saturation, the normalized transconductance efficiency g_{ms}/i_d is found if we combine Eq. (9.3) and Eq. (9.6):

$$\frac{g_{ms}}{i_d} = \frac{\sqrt{\lambda_c^2\, IC^2 + 2\,\lambda_c\, IC + 4\, IC + 1} - 1}{\lambda_c^2\, IC^2 + \lambda_c\, IC + 2\, IC}. \tag{9.10}$$

The normalized transconductance efficiency g_{ms}/i_d for a short- and long-channel device is shown in Fig. 9.2. We see that g_{ms}/i_d remains invariant and equal to 1 for both devices in weak-inversion, whereas in strong-inversion it degrades much faster for the short-channel device due to the effect of velocity saturation. The strong-inversion asymptotes for both devices can be obtained through Eq. (9.10) using the same assumptions that where used for

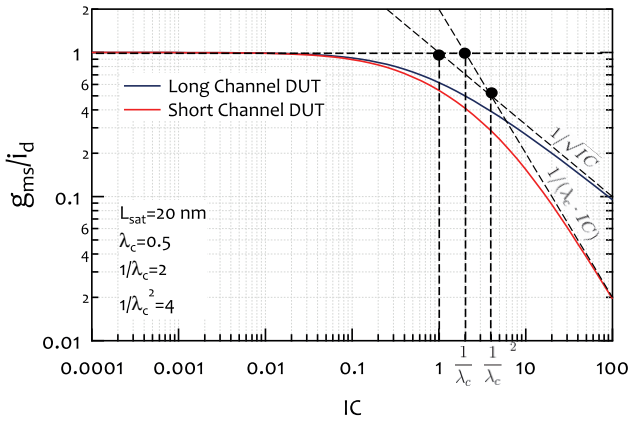


Figure 9.2 – g_{ms}/i_d vs. IC in saturation for a short and a long-channel device, having L =40 nm and L =10 µm, respectively. We see that the g_{ms}/i_d of the short-channel device degrades much faster in SI compared to the long-channel device, due to the effect of velocity saturation.

Eq. (9.5) and Eq. (9.7). So, for the long-channel device, without velocity saturation, we get:

$$\left.\frac{g_{ms}}{i_d}\right|_{\text{SI asymptote w/o. VS}} = \frac{1}{\sqrt{IC}} = \frac{1}{\sqrt{i_d}}, \tag{9.11}$$

while for the short-channel device, with velocity saturation, we get:

$$\left.\frac{g_{ms}}{i_d}\right|_{\text{SI asymptote w. VS}} = \frac{1}{\lambda_c\, IC} = \frac{1}{\lambda_c\, i_d}. \tag{9.12}$$

Similar to Fig. 9.1, in Fig. 9.2, we see that the different asymptotes cross at the same points, meaning that the weak and strong-inversion asymptotes for the short-channel device cross at $IC = 1/\lambda_c$, and that the strong-inversion asymptotes of both devices cross at $IC = 1/\lambda_c{}^2$.

After the description of the G_m/I_D FoM, the extraction procedure of the three parameters n, $I_{spec_\square}$ and L_{sat} will be presented. The parameters are extracted accounting for the characteristics of the G_m/I_D in different regions. This procedure can be used for any DUT and at any temperature (as long as the impact of the T on the U_T is accounted for) irrespectively of the process. It should be mentioned though that a different parameters should be extracted for nMOS and pMOS devices. In order to describe the procedure, I_D vs. V_G measurements of nMOS RF DUTs of a 40 nm CMOS process at V_D = 1.1 V will be used.

▷ Extraction of n

As described above, in weak-inversion $g_{ms}/i_d = 1$. So, using the Eq. (9.9) and solving for

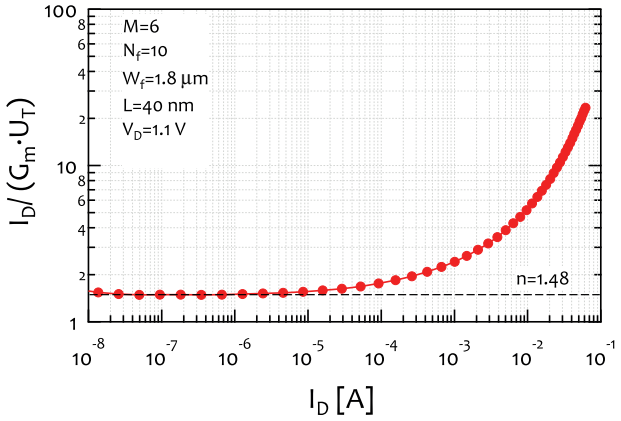


Figure 9.3 – Extraction of the slope factor n, which corresponds to the minimum value of the curve $I_\mathrm{D}/(G_\mathrm{m}\cdot U_\mathrm{T})$. The measurements correspond to an RF DUT having $M = 6$, $N_\mathrm{f} = 10$, $W_\mathrm{f} = 1.8$ µm and $L = 40$ nm, at $V_\mathrm{D} = 1.1$ V.

the slope factor n, we get:

$$n = \frac{1}{\frac{G_\mathrm{m}}{I_\mathrm{D}}\cdot U_\mathrm{T}}. \tag{9.13}$$

The slope factor can be then easily extracted if we plot $I_\mathrm{D}/(G_\mathrm{m}\cdot U_\mathrm{T})$ vs. I_D, set both axes in logarithmic scale, so that weak-inversion region is displayed better, and find its minimum value. The extraction of the n of a short-channel device is shown in Fig. 9.3.

▷ Extraction of $I_{\mathrm{spec}_\square}$

The extraction of the $I_{\mathrm{spec}_\square}$ should be carried out in a wide/long-channel device so that there are no short-channel effects and the device behaves in a manner close to ideal. From Eq. (9.11) we know that the SI asymptote of $g_\mathrm{ms}/i_\mathrm{d}$ for a long-channel device is equal to $1/\sqrt{i_\mathrm{d}}$, which in terms of non-normalized quantities can be translated into:

$$\left.\frac{G_\mathrm{m}\cdot n\cdot U_\mathrm{T}}{I_\mathrm{D}}\right|_{\text{SI asymptote w/o. VS}} = \sqrt{\frac{I_\mathrm{spec}}{I_\mathrm{D}}}. \tag{9.14}$$

So, we can extract the I_spec if we plot $(G_\mathrm{m}\cdot n\cdot U_\mathrm{T})/I_\mathrm{D}$ vs. I_D, with both axes in logarithmic scale, draw the asymptote of $(G_\mathrm{m}\cdot n\cdot U_\mathrm{T})/I_\mathrm{D}$ in strong-inversion and calculate I_spec as the current I_D at which the asymptote is equal to 1. Of course the extraction of n for this device must have preceded this step. The extraction of the I_spec is shown graphically in Fig. 9.4. After extracting I_spec, we can then calculate $I_{\mathrm{spec}_\square}$ from:

$$I_\mathrm{spec} = I_{\mathrm{spec}_\square}\cdot\frac{W}{L}. \tag{9.15}$$

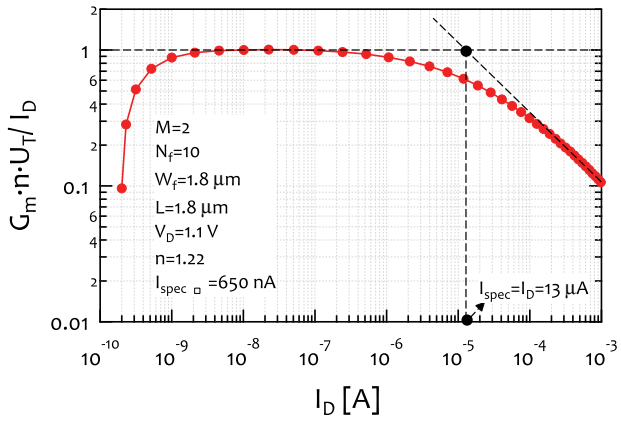


Figure 9.4 – Extraction of the specific current I_{spec}, which corresponds to the value of I_D at which the SI asymptote is equal to 1. The measurements correspond to an RF DUT having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 1.8$ µm, at $V_D = 1.1$ V. The extraction of n should have been carried out already.

The above method cannot be used for short-channel devices since in strong-inversion $(G_m \cdot n \cdot U_T)/I_D$ no longer varies as $1/\sqrt{IC}$, but instead it varies as $1/(\lambda_c\, IC)$ and λ_c is extracted in the next step.

Keeping $I_{spec_\square}$ as a constant parameter for all devices of a process (note that different $I_{spec_\square}$ parameters should be calculated for nMOS and pMOS devices) is a very rough approximation. $I_{spec_\square}$ is expressed as [39]:

$$I_{spec_\square} = 2\, n\, \mu\, C_{ox}\, U_T^2, \tag{9.16}$$

so, it is affected by any change in n, μ and C_{ox}. Assuming that μ and C_{ox} remain almost constant (the effective low-field mobility μ_{eff} and the effective gate capacitance per unit area $C_{ox_{eff}}$ do change due to the short-channel effects but μ and C_{ox} will be considered constant for simplicity), we can take into account the change of n. Ideally, we would like to have a slope factor that is the same for all devices and equal to the n of long-channel device, however we observe an increase of the slope factor as the channel length is decreased. A simple approach to calculate $I_{spec_\square}$ of each device, taking into account the change in the slope factor, is by using as a reference the $I_{spec_\square}$ of a long-channel device, using the following formula:

$$I_{spec_{\square_{DUT}}} = I_{spec_{\square_{long}}} \cdot \frac{n_{short}}{n_{long}}. \tag{9.17}$$

▷ **Extraction of L_{sat}**

Parameters n and $I_{spec_\square}$ are sufficient to model the G_m of a long-channel device. The extraction of L_{sat}, is only required for short-channel devices where velocity saturation effect manifests. For the extraction of L_{sat} we follow a similar procedure as the one that

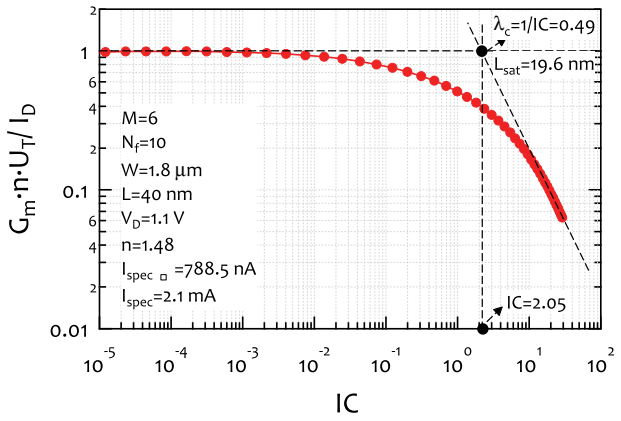


Figure 9.5 – Extraction of the velocity saturation parameter λ_c, which corresponds to the inverse value of the IC at which the SI asymptote is equal to 1. The measurements correspond to an RF DUT having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, at $V_D = 1.1$ V. The values of n and $I_{spec\square}$ should have been extracted already.

was presented for the extraction of $I_{spec\square}$, but in that case using data for a short-channel DUT. In details, from Eq. (9.12) we know that the SI asymptote of g_{ms}/i_d of a short-channel device is equal to $1/(\lambda_c\, IC)$, which in terms of non-normalized quantities is translated into:

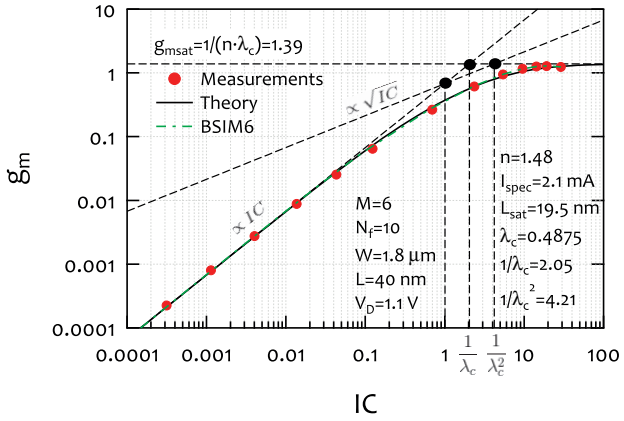
$$\left.\frac{G_m \cdot n \cdot U_T}{I_D}\right|_{\text{SI asymptote w. VS}} = \frac{I_{spec}}{\lambda_c \cdot I_D}. \tag{9.18}$$

Having already extracted from the previous steps n and $I_{spec\square}$, that are needed for the calculation of I_{spec}, we can now proceed to the extraction of the λ_c. For this we plot $(G_m \cdot n \cdot U_T)/I_D$ vs. IC, with both axes in logarithmic scale, draw the asymptote of $(G_m \cdot n \cdot U_T)/I_D$ in strong-inversion and calculate λ_c as the inverse of the IC at which the asymptote is equal to 1. The extraction of the λ_c is shown graphically in Fig. 9.5. Then we can calculate L_{sat} from Eq. (9.4).

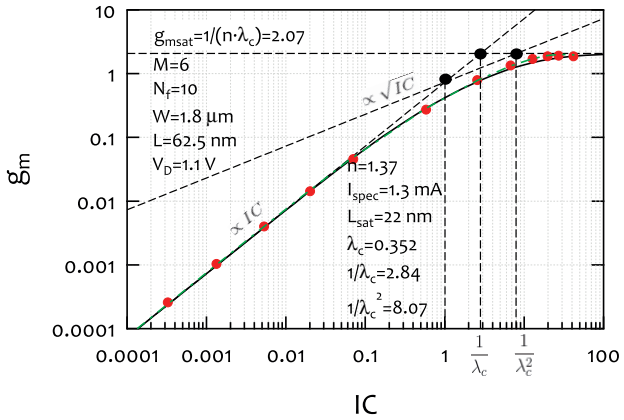
Eq. (9.4) offers a quite simplistic approximation for the calculation of λ_c, assuming a constant L_{sat} value for all the devices of a process. However, as a second order effect, L_{sat} also demonstrates a length dependence, and as a result can be considered constant only for a small range of channel lengths. In a case that we would like to have a model able to capture the effect of velocity saturation for all the DUTs of the process, we might need to take into account a length dependent L_{sat}, either through binning ot by introducing a scaling formula.

9.2.3 Model Verification

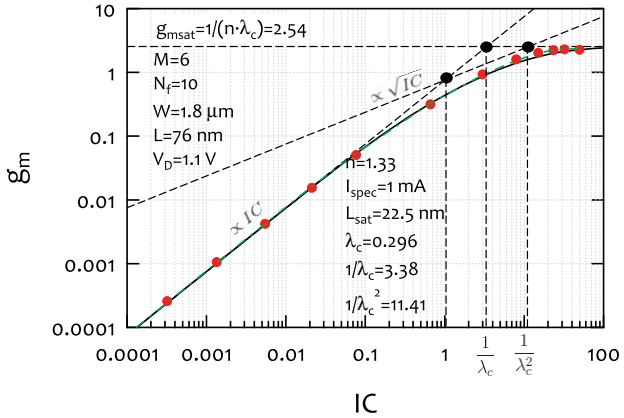
Now that all the parameters required to model G_m, as described by Eq. (9.2), have been extracted, we can proceed to the comparison of the model against measurements. In Fig. 9.6



(a) $L = 40$ nm

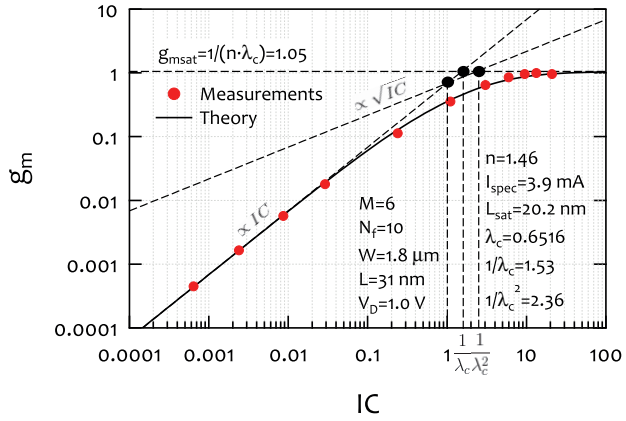


(b) $L = 62.5$ nm

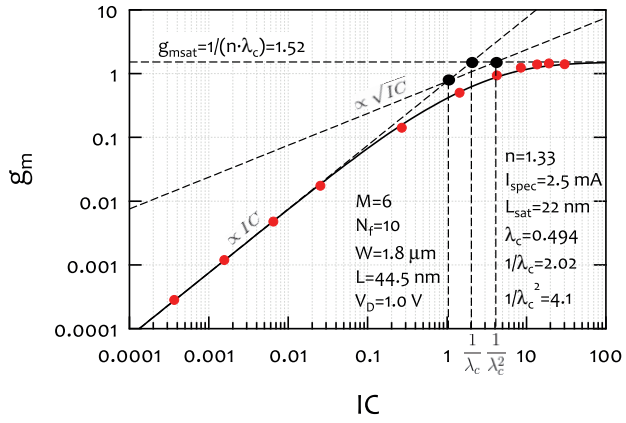


(c) $L = 76$ nm

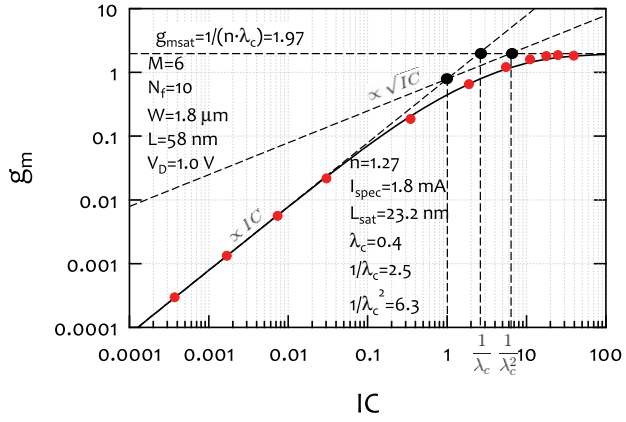
Figure 9.6 – Normalized transconductance g_m vs. IC for three RF DUTs of a 40 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, 62.5 nm, 76 nm, at $V_D = 1.1$ V. Here, a comparison between theory, measurements and BSIM6 model is made. The extracted values of the parameters of each device are also shown. All the devices have $I_{spec_\square} = 650$ nA.



(a) $L = 30$ nm



(b) $L = 44.5$ nm



(c) $L = 58$ nm

Figure 9.7 – Normalized transconductance g_m vs. IC for three RF DUTs of a 28 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 31$ nm, 44.5 nm, 58 nm, at $V_D = 1.0$ V. Here, a comparison between theory and measurements is made. The extracted values of the parameters of each device are also shown. All the devices have $I_{spec\square} = 870$ nA.

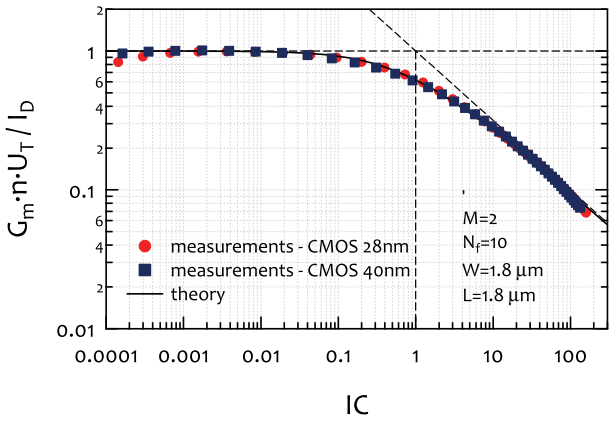


Figure 9.8 – Transconductance efficiency $G_m \cdot n \cdot U_T/I_D$ vs. IC for long RF DUTs of 28 nm and 40 nm CMOS process. We see that after proper normalization, $I_{spec_\square}$ and n being different for each device, all the measured points nicely fall on the same analytical curve.

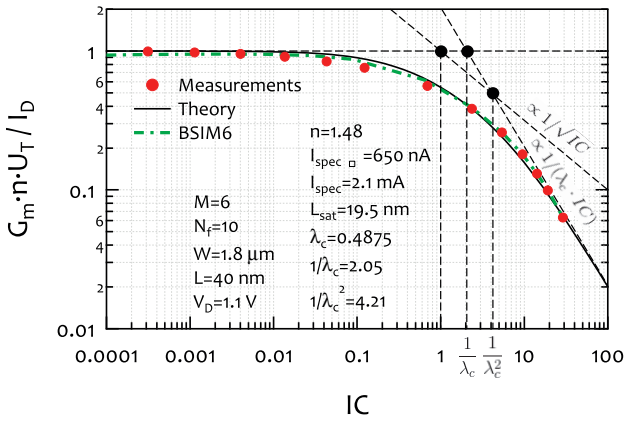
the analytical model is compared with measurements of a 40 nm CMOS process and BSIM6, whereas in Fig. 9.7 the model is compared with measurements of a 28 nm CMOS process. In both figures the comparison is done for three short-channel devices having different channel lengths. To achieve better accuracy, a local parameter extraction is selected, meaning that n and L_{sat} are extracted for each device, whereas $I_{spec_\square}$ is kept the same for all the devices of each process. We see that the analytical model is able to capture with precision the G_m of nanoscale devices, with the use of only three parameters. Since $g_m = g_{ms}/n$, the g_m saturates at a value equal to $g_{m_{sat}} = 1/(n \cdot \lambda_c)$ for $IC > 1/\lambda_c^2$.

9.3 Modeling the Transconductance Efficiency - G_m/I_D

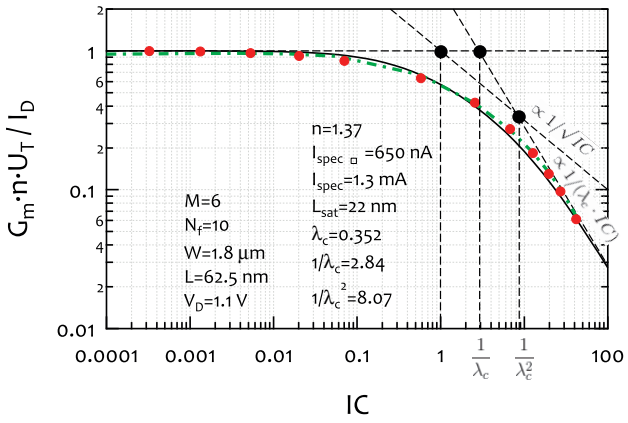
The transconductance efficiency G_m/I_D FoM is one of the most important performance metrics for analog circuit design. It is a measure of how much gain can be produced for a given bias current and it can be expressed as a function of the inversion coefficient as described earlier with Eq. (9.10). G_m/I_D is a quantity that is also very useful in the context of general circuit sizing [131, 132]. Recently, it was shown that G_m/I_D can be used even to calculate the harmonic distortion of a MOSFET [133].

For long-channel devices, where there are no short-channel effects, the g_{ms}/i_d characteristic is almost invariant to the technology. This means that all the measured points nicely fall on the analytical curve after proper normalization as shown in Fig. 9.8 for two CMOS processes, namely, 28 nm and 40 nm.

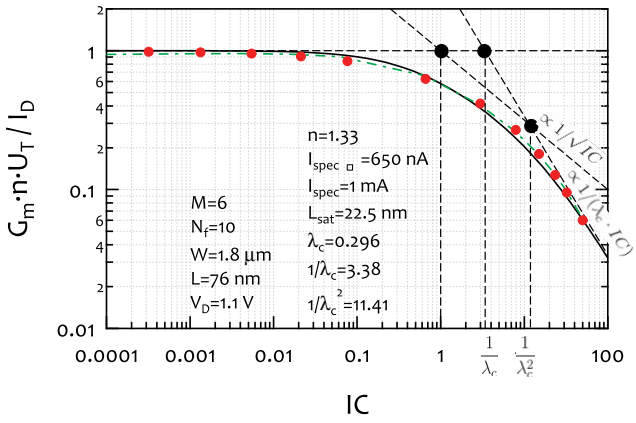
On the contrary, when we move to short-channel devices, λ_c is dependent on the channel length and thus the strong-inversion asymptote of the g_{ms}/i_d curve is no more geometry and technology invariant, as it is for the long-channel devices. In Fig. 9.9 and Fig. 9.10, we see the



(a) $L = 40$ nm

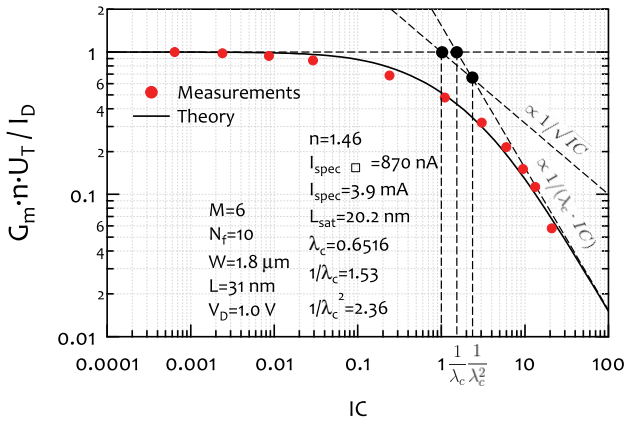


(b) $L = 62.5$ nm

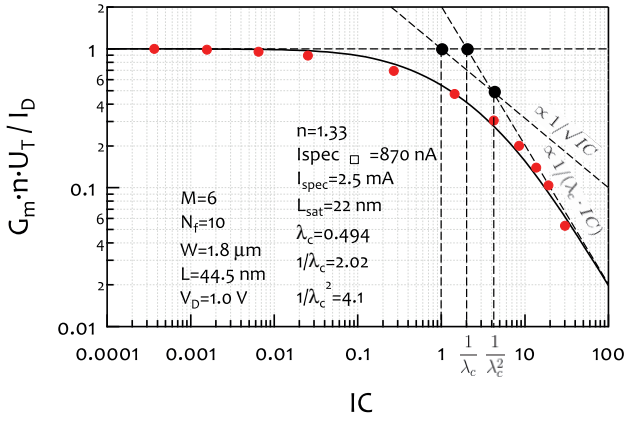


(c) $L = 76$ nm

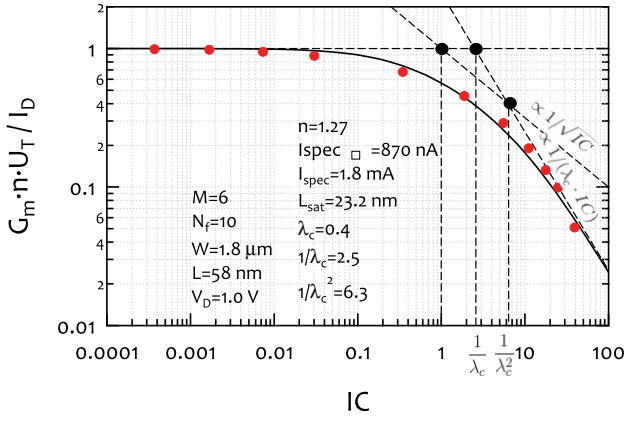
Figure 9.9 – Transconductance efficiency $G_m \cdot n \cdot U_T / I_D$ vs. IC for three RF DUTs of a 40 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, 62.5 nm, 76 nm, at $V_D = 1.1$ V. Here, a comparison between theory, measurements and the BSIM6 model is made. The extracted values of the parameters of each device are also shown.



(a) $L = 30$ nm



(b) $L = 44.5$ nm



(c) $L = 58$ nm

Figure 9.10 – Transconductance efficiency $G_m \cdot n \cdot U_T/I_D$ vs. IC for three RF DUTs of a 28 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 31$ nm, 44.5 nm, 58 nm, at $V_D = 1.0$ V. Here, a comparison between theory and measurements is made. The extracted values of the parameters of each device are also shown.

$G_m \cdot n \cdot U_T / I_D$ vs. *IC* for the same devices that were employed for the g_m vs. *IC* plots, using the same parameters. As it can be seen from these figures, the effect of velocity saturation is degrading the transconductance efficiency in strong-inversion, meaning that more current is required to obtain the same transconductance than the one that would be obtained without velocity saturation. Nevertheless, irrespectively of the channel length, the transconductance efficiency $G_m \cdot n \cdot U_T / I_D$ (or the normalized value g_{ms}/i_d) remains invariant ($g_{ms}/i_d = 1$) in weak-inversion. And although the short-channel effects have a strong impact on the drain current I_D and the gate transconductance G_m, e.g. DIBL impacts the weak-inversion region and velocity saturation the strong-inversion region, their ratio remains unaffected in WI. This can be explained if we take into account that G_m is proportional to I_D in weak-inversion and thus I_D and G_m are affected in the same way in this region [134].

9.4 Modeling the Output Conductance - G_{ds}

In analog circuit design a small output conductance G_{ds} is desirable. As it will be explained in Sec. 9.5, G_{ds} is directly linked to the intrinsic voltage gain of the transistor A_{v_i} through [39, 125, 135]:

$$A_{v_i} = \frac{G_m}{G_{ds}} = \frac{g_m}{g_{ds}}, \tag{9.19}$$

and, thus, the smaller the G_{ds} the larger the A_{v_i} that is achieved. In short-channel devices, there are different effects that affect the output conductance with the Drain Induced Barrier Lowering (DIBL) and the Channel Length Modulation (CLM) to impact G_{ds} more profoundly. In order to model G_{ds} in a simplistic way we will start with:

$$G_{ds} = \frac{\partial I_D}{\partial V_D} = \frac{\partial I_D}{\partial V_{T0}} \cdot \frac{\partial V_{T0}}{\partial V_D}. \tag{9.20}$$

Since, I_D can be regarded as a function of $V_G - V_{T0}$, using Eq. (9.1), we can write:

$$\frac{\partial I_D}{\partial V_{T0}} = -\frac{\partial I_D}{\partial V_G} = -G_m. \tag{9.21}$$

G_m is already modeled in Sec. 9.2, so the remaining part left to be evaluated in Eq. (9.20) is the calculation of $\partial V_{T0}/\partial V_D$, which is the V_{T0} shift due to the impact of V_D or in other words the effect of DIBL. We can therefore define a parameter:

$$\alpha_{dibl} = -\left(\frac{\partial V_{T0}}{\partial V_D}\right)^{-1}, \tag{9.22}$$

with the minus sign in the above equation coming from Eq. (9.21), and thus G_{ds} becomes:

$$G_{ds} = \frac{G_m}{\alpha_{dibl}}. \tag{9.23}$$

Note that DIBL causes a reduction of V_{T0} with the increase of V_D, which means that $(\partial V_{T0}/\partial V_D)^{-1}$ will be negative and consequently α_{dibl} will be positive.

Using Eq. (9.23) to model G_{ds} we found out that although it is valid in weak-inversion, in strong-inversion G_{ds} does not saturate as fast as G_m. This can be attributed to the impact of CLM effect which causes an increase in I_D as V_D increases. Nevertheless, we can use a similar formula to G_m in order to model G_{ds}, but in that case we will use a different "λ_c" parameter, which will be defined as λ_{sat_gds}. As a result, the normalized output conductance can be calculated through:

$$g_{ds} = \frac{G_{ds}}{G_{spec}} = \frac{1}{n \cdot \alpha_{dibl}} \cdot \frac{\sqrt{\lambda_{sat\_gds}^2 IC^2 + 2\,\lambda_{sat\_gds}\, IC + 4\, IC + 1} - 1}{\lambda_{sat\_gds}^2 IC + \lambda_{sat\_gds} + 2}. \tag{9.24}$$

9.4.1 Extraction of Parameters - α_{dibl}, λ_{sat_gds}

Now that we have Eq. (9.24) to model g_{ds}, we need to extract the parameters α_{dibl} and λ_{sat_gds}. Considering that DIBL and CLM become more intense with the reduction of L, both parameters will depend on the channel length and, hence, they need to be extracted separately for each device. Below, the extraction procedure for these parameters will be demonstrated for a 40 nm nMOS RF DUT.

Starting with the α_{dibl} parameter, we need to mention that the shift in V_{T0} due to the DIBL effect is not a linear function of V_D. As a result, α_{dibl} is better to be extracted close to the V_D bias of interest. In our case $V_D = 1.1$, so we will focus in that region of operation.

An easy way to extract the value of α_{dibl}, requires to estimate first the V_{T0} in at least two different V_D points close to the V_D bias of interest and then calculate α_{dibl} through Eq. (9.22). Taking this into account, we estimated the V_{T0} at $V_D = [0.9, 1.1]\,V$, using the constant current method proposed in [115](different methods also exist e.g. [136]). Then we found:

$$\alpha_{dibl} = -\frac{\Delta V_D}{\Delta V_{T0}} = -\frac{0.2V}{-0.023V} = 8.7. \tag{9.25}$$

Since, G_{ds} has a very similar expression to G_m (only shifted by a factor α_{dibl} and having its own λ_{sat_gds}), it is reasonable to follow the same procedure to extract the value of λ_{sat_gds} as for the extraction of λ_c parameter in Sec. 9.2.2. In more details, we will now use G_{ds}/I_D, which in its normalized form is given by:

$$\frac{G_{ds}}{I_D} = \frac{g_{ds} \cdot G_{spec}}{i_d \cdot I_{spec}} = \frac{g_{ds}}{i_d} \cdot \frac{1}{n \cdot U_T \cdot \alpha_{dibl}}, \tag{9.26}$$

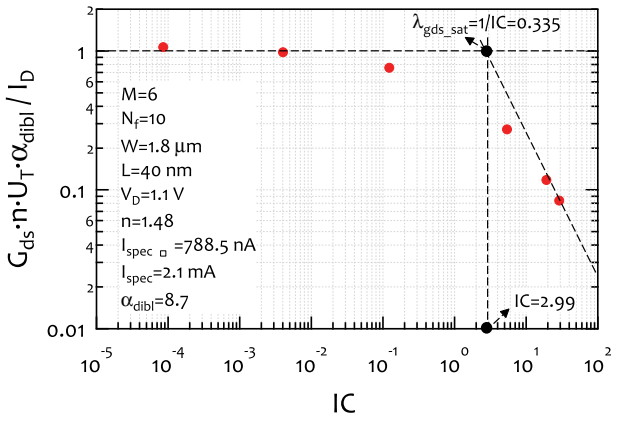


Figure 9.11 – Extraction of the parameter $\lambda_{\text{sat\_gds}}$, which corresponds to the inverse value of the IC at which the SI asymptote is equal to 1. The measurements correspond to an RF DUT having $M = 6$, $N_{\text{f}} = 10$, $W_{\text{f}} = 1.8$ µm and $L = 40$ nm, at $V_{\text{D}} = 1.1$ V. The values of n and $I_{\text{spec}\square}$ should have been extracted already.

where

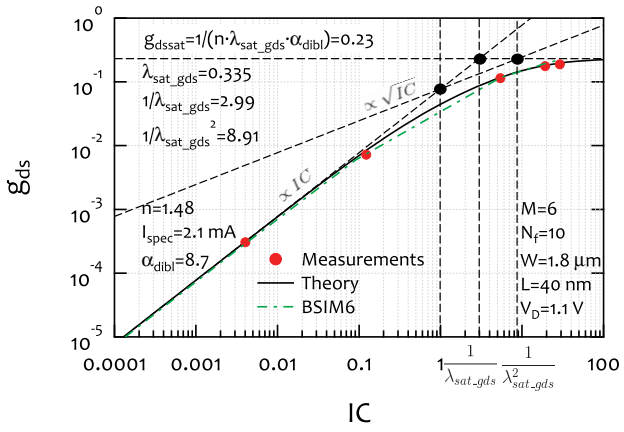
$$\frac{g_{\text{ds}}}{i_{\text{d}}} = \frac{\sqrt{\lambda_{\text{sat\_gds}}^2\, IC^2 + 2\,\lambda_{\text{sat\_gds}}\, IC + 4\, IC + 1} - 1}{\lambda_{\text{sat\_gds}}^2\, IC^2 + \lambda_{\text{sat\_gds}}\, IC + 2\, IC}. \tag{9.27}$$

From the above equation, we expect that the SI asymptote of $g_{\text{ds}}/i_{\text{d}}$ is the line $1/(\lambda_{\text{sat\_gds}}\, IC)$, which in terms of non-normalized quantities is translated into:

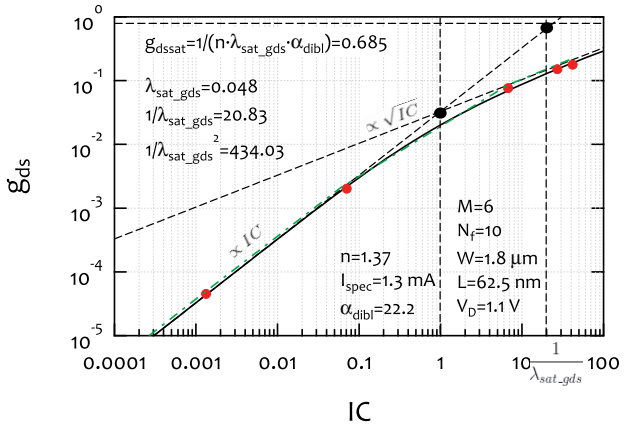
$$\left.\frac{G_{\text{ds}} \cdot n \cdot U_{\text{T}} \cdot \alpha_{\text{dibl}}}{I_{\text{D}}}\right|_{\text{SI asymptote}} = \frac{I_{\text{spec}}}{\lambda_{\text{sat\_gds}} \cdot I_{\text{D}}}. \tag{9.28}$$

Having already extracted in Sec. 9.2.2 n and $I_{\text{spec}\square}$, needed for the calculation of I_{spec}, we can now proceed to the extraction of the $\lambda_{\text{sat\_gds}}$. For this we plot $(G_{\text{ds}} \cdot n \cdot U_{\text{T}} \cdot \alpha_{\text{dibl}})/I_{\text{D}}$ vs. IC, with both axes in logarithmic scale, draw the asymptote of $(G_{\text{ds}} \cdot n \cdot U_{\text{T}} \cdot \alpha_{\text{dibl}})/I_{\text{D}}$ in strong-inversion and calculate $\lambda_{\text{sat\_gds}}$ as the inverse of the IC at which the asymptote is equal to 1. The extraction of the $\lambda_{\text{sat\_gds}}$ is shown graphically in Fig. 9.11.

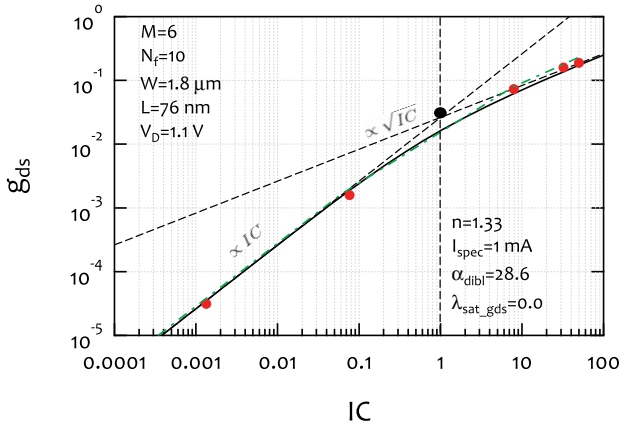
The α_{dibl} parameter shows how much lower is G_{ds} compared to G_{m} in the weak-inversion region. This is the maximum difference that can be achieved between G_{m} and G_{ds}. In strong-inversion where G_{ds} does not saturate as fast as G_{m} due to CLM, this difference is expected to be lower. Therefore, in short-channel devices where the α_{dibl} is already quite low (in the example of Fig. 9.11 α_{dibl} is only 8.7), it means that G_{ds} degrades A_{v_i} significantly and in strong-inversion this degradation becomes even higher.



(a) $L = 40$ nm

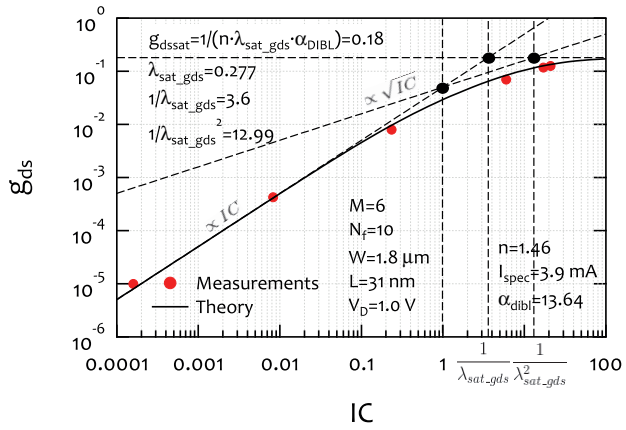


(b) $L = 62.5$ nm

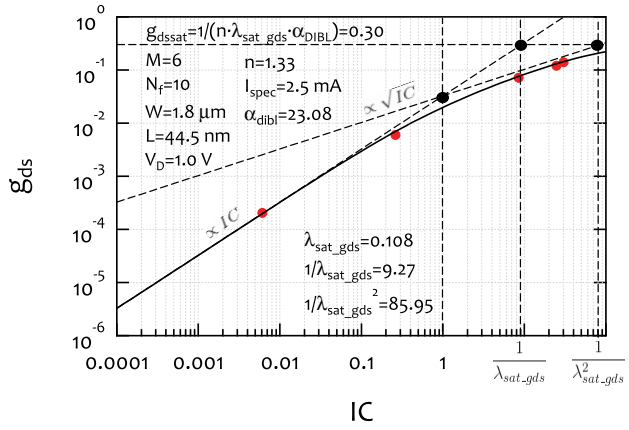


(c) $L = 76$ nm

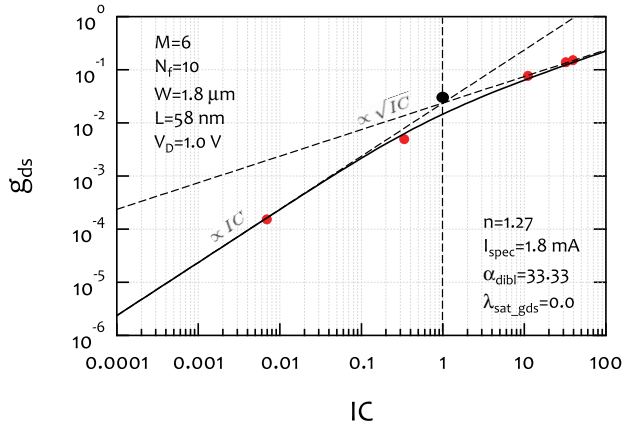
Figure 9.12 – Normalized output conductance g_{ds} vs. IC for three RF DUTs of a 40 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, 62.5 nm, 76 nm, at $V_D = 1.1$ V. Here, a comparison between theory, measurements and the BSIM6 model is made. The extracted values of the parameters of each device are also shown.



(a) $L = 30$ nm



(b) $L = 44.5$ nm



(c) $L = 58$ nm

Figure 9.13 – Normalized output conductance g_{ds} vs. IC for three RF DUTs of a 28 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 31$ nm, 44.5 nm, 58 nm, at $V_D = 1.0$ V. Here, a comparison between theory and measurements is made. The extracted values of the parameters of each device are also shown.

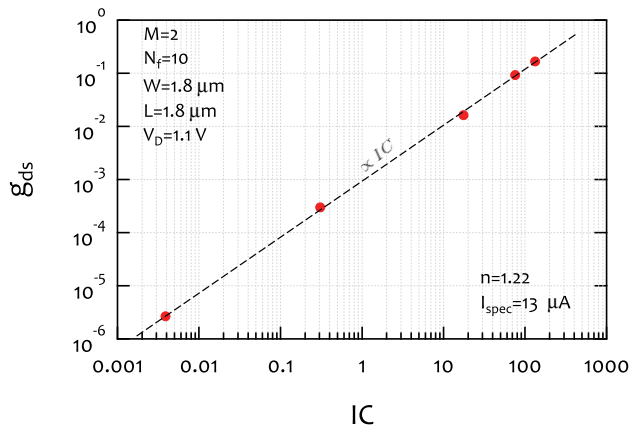


Figure 9.14 – Normalized output conductance g_{ds} vs. IC for a long-channel RF DUT of a 40 nm CMOS process having $M = 2$, $N_f = 10$, $W_f = 1.8$ µm and $L = 1.8$ µm, at $V_D = 1.1$ V. Here, we see that the g_{ds} of a long-channel device is proportional to IC.

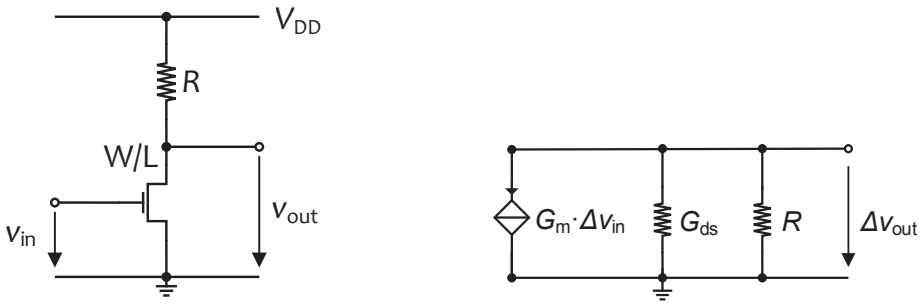
9.4.2 Model Verification

After the extraction of all the parameters required to model G_{ds}, as described by Eq. (9.26), we can proceed to the comparison of the model against measurements. In Fig. 9.12 the analytical model is compared with measurements of a 40 nm CMOS process and BSIM6, whereas in Fig. 9.13 the model is compared with measurements of a 28 nm CMOS process. The model is capturing accurately the output conductance behavior for all the DUTs demonstrated.

From the figures we see that the saturation of g_{ds} mainly happens for short channel devices, whereas for longer ones, for non-minimum length DUTs with $L < 2 \cdot L_{min}$, where L_{min} is the nominal channel length of the process, there is almost no saturation and g_{ds} becomes $\propto \sqrt{IC}$ in strong-inversion. For devices with $L \approx 2 \cdot L_{min}$, λ_{sat_gds} becomes zero and thus g_{ds} can be modeled only with α_{dibl} parameter. Despite the fact that G_{ds} is quite high for short-channel devices, the saturation of G_{ds} in strong-inversion can be considered at least beneficial, since the degradation of A_{v_i} would be worse if the G_{ds} of short-channel devices would follow the same trend as the longer ones. For long-channel devices, g_{ds} is no longer varying proportionally to $\sqrt{IC}$ but instead it varies proportional to IC as shown in Fig. 9.14.

9.5 Modeling the Intrinsic Voltage Gain - A_{v_i}

The intrinsic voltage gain A_{v_i} of a MOSFET is defined as the low-frequency, small-signal, gate-to-drain voltage gain of a MOSFET in a CS (common-source) configuration, when the drain is connected to an infinite resistance [39, 125, 135]. To explain the derivation of the A_{v_i}, we will use a simple CS MOS amplifier as shown in Fig. 9.15a, with its simplified small-signal equivalent circuit in saturation displayed in Fig. 9.15b. If a small-signal input, v_{in}, is applied



(a) Simple MOS Amplifier in CS Configuration (b) Small-Signal Equivalent Circuit in Saturation

Figure 9.15 – Simple MOS amplifier in a CS configuration (a), with its simplified small-signal equivalent circuit in saturation (b). This configuration is ideal to explain the derivation of A_{v_i}.

then:

$$A_v = \frac{v_{out}}{v_{in}} = -\frac{G_m}{G_{ds} + \dfrac{1}{R}}, \tag{9.29}$$

which means that the magnitude of the output voltage is the input voltage amplified by a gain factor of $G_m/(G_{ds} + 1/R)$. Even when R approaches infinity, the voltage gain cannot exceed the intrinsic or maximum voltage gain of the transistor $A_{v_i} = G_m/G_{ds}$ given in Eq. (9.19).

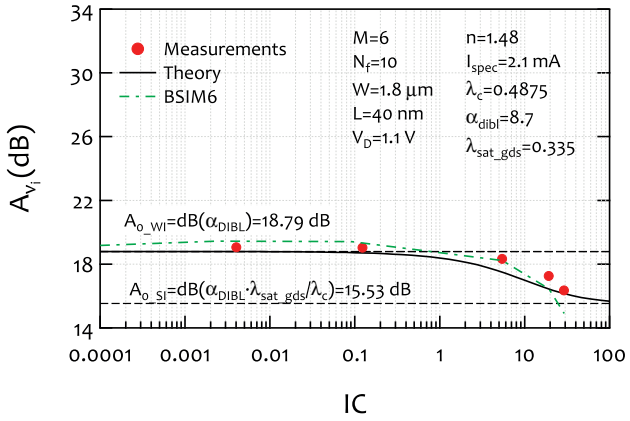
In Sec. 9.2 and Sec. 9.4, we demonstrated how we can model G_m and G_{ds}. Since, the intrinsic voltage gain is just the ratio of these two quantities we can model A_{v_i} as well. In Fig. 9.16 and Fig. 9.17 the A_{v_i} vs. IC is demonstrated. The analytical model is compared against measurement and the BSIM6 model for the 40 nm CMOS process and against measurements for the 28 nm CMOS process. As it was expected in weak-inversion the A_{v_i} is equal to α_{dibl} parameter resulting from:

$$A_{v_i}\Big|_{\text{WI asymptote}} = \frac{G_m|_{\text{WI asymptote}}}{G_{ds}|_{\text{WI asymptote}}} = \frac{g_m|_{\text{WI asymptote}}}{g_{ds}|_{\text{WI asymptote}}} = \frac{IC/n}{IC/(n \cdot \alpha_{dibl})} = \alpha_{dibl}, \tag{9.30}$$

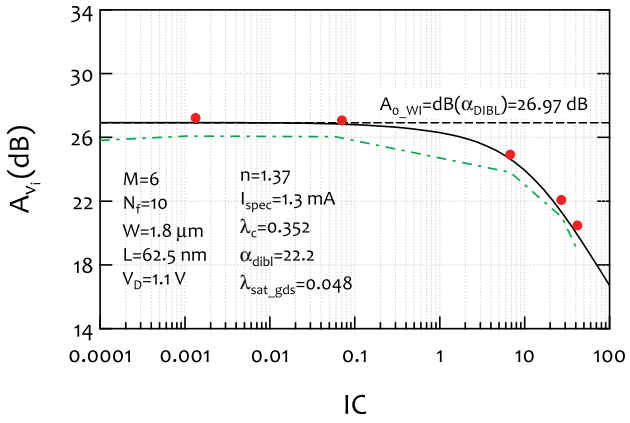
whereas the strong-inversion asymptote can be calculated as:

$$A_{v_i}\Big|_{\text{SI asymptote}} = \frac{G_m|_{\text{SI asymptote}}}{G_{ds}|_{\text{SI asymptote}}} = \frac{g_m|_{\text{SI asymptote}}}{g_{ds}|_{\text{SI asymptote}}} = \frac{1/(n \cdot \lambda_c)}{1/(n \cdot \lambda_{sat\_gds} \cdot \alpha_{dibl})} = \alpha_{dibl} \cdot \frac{\lambda_{sat\_gds}}{\lambda_c}. \tag{9.31}$$

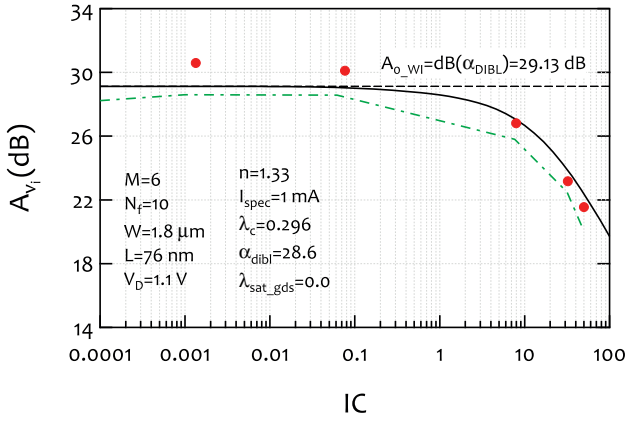
Since $\lambda_{sat\_gds} < \lambda_c$, it is obvious that the A_{v_i} will degrade in strong-inversion but how fast this will happen depends on the difference between the parameters λ_{sat_gds} ans λ_c. The bigger the



(a) $L = 40$ nm

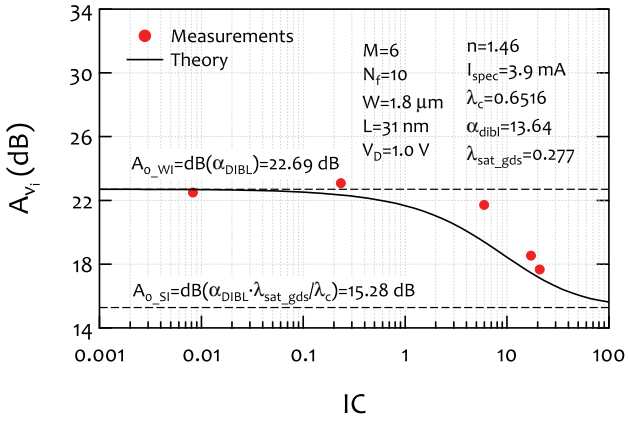


(b) $L = 62.5$ nm

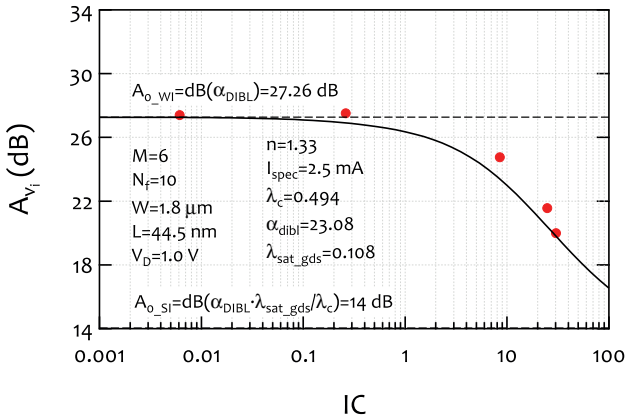


(c) $L = 76$ nm

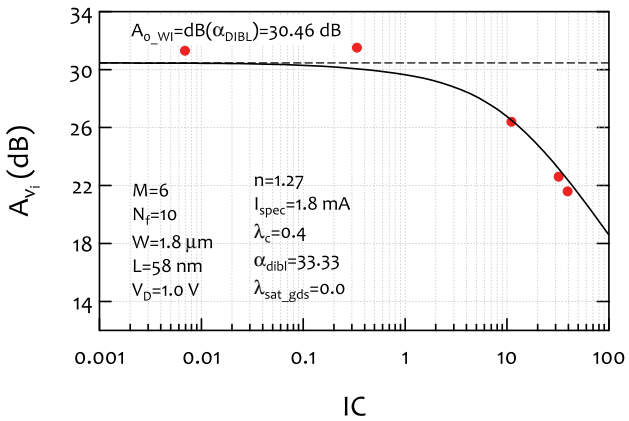
Figure 9.16 – Intrinsic voltage gain A_{V_i} vs. IC for three RF DUTs of a 40 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, 62.5 nm, 76 nm, at $V_D = 1.1$ V. Here, a comparison between theory, measurements and the BSIM6 model is made. The extracted values of the parameters of each device are also shown.



(a) $L = 30$ nm



(b) $L = 44.5$ nm



(c) $L = 58$ nm

Figure 9.17 – Intrinsic voltage gain A_{v_i} vs. IC for three RF DUTs of a 28 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 31$ nm, 44.5 nm, 58 nm, at $V_D = 1.0$ V. Here, a comparison between theory and measurements is made. The extracted values of the parameters of each device are also shown.

difference is, the faster A_{v_i} degrades.

For long-channel devices, we saw that G_m is proportional to IC in weak-inversion and proportional to $\sqrt{IC}$ in strong inversion, whereas G_{ds} is almost proportional to IC from weak- to strong-inversion. This means that the intrinsic gain A_{v_i} should be constant in weak-inversion and proportional to $1/\sqrt{IC}$ in strong-inversion.

Although in Fig. 9.16 and Fig. 9.17 the model seems not to demonstrate sufficient accuracy, we should mention that the difference between the model and the measurements does not exceed an error of 10%. Similar results for the A_{v_i} trend vs. IC were demonstrated in case of DG (double-gate) MOSFETs [137].

9.6 Modeling the Transit Frequency - F_t

The transit frequency F_t is a metric widely used for characterizing the RF behavior of a MOSFET. Many other performance metrics, such as the RF gain and the minimum noise factor F_{min}, are directly linked to F_t [39]. The downscaling of modern CMOS processes has resulted in an impressive boost of the transit frequency. This is especially in favor of low-power RF circuit design since it allows for operation within the frequency range in the order of tens of GHz while the transistor is still biased in moderate-inversion region. A good approximation of the F_t is given by [5, 39]:

$$F_t = \frac{G_m}{2\pi \cdot C_{GG}}. \tag{9.32}$$

with

$$C_{GG} = C_{GG_i} + C_{GG_e} \tag{9.33}$$

being the total gate capacitance comprising of the intrinsic C_{GG_i} and the extrinsic C_{GG_e} part. C_{GG_e} includes the contribution of both the overlap and fringing capacitances and in weak-inversion it can regarded as a linearly scaling quantity with respect to the overall width of the device that can be approximated using [39, 138]:

$$C_{GG_e} = C_{G_{eW}} \cdot W. \tag{9.34}$$

In Eq. (9.32) both G_m and C_{GG} are bias dependent and as a result F_t is bias dependent too. In WI though, C_{GG} can be considered almost constant and thus the variation is coming from G_m, meaning that F_t is proportional to IC. Similarly to G_m/I_D, F_t can be normalized to $F_{t_{spec}}$ defined as the value of the F_t on the WI asymptote at $IC = 1$. The exact calculation of $F_{t_{spec}}$ will

follow. The normalized transit frequency is then given by:

$$f_t = \frac{F_t}{F_{t_{spec}}} = g_{ms} = \frac{\sqrt{\lambda_c^2\, IC^2 + 2\,\lambda_c\, IC + 4\, IC + 1} - 1}{\lambda_c^2\, IC + \lambda_c + 2}. \tag{9.35}$$

In agreement with the g_{ms}, in strong-inversion and under the effect of velocity saturation (i.e. for $IC > 1/\lambda_c^2$), f_t saturates to $1/\lambda_c$. For longer devices, where λ_c has also a lower value, velocity saturation prevails at higher IC values and thus there is a region between $IC = 1/\lambda_c$ and $IC = 1/\lambda_c^2$ where f_t follows the ideal (no velocity saturation) SI asymptote $\sqrt{IC}$. We can calculate the analytical expression of $F_{t_{spec}}$ in WI, assuming that $C_{GG} \approx C_{GG_e}$, as follows:

$$\begin{aligned}
&f_t = g_{ms} \Rightarrow \\
&\frac{F_t}{F_{t_{spec}}} = g_{ms} \Rightarrow \\
&g_{ms} \cdot F_{t_{spec}} = \frac{G_m}{2\pi \cdot C_{GG}} \overset{C_{GG} \approx C_{GGe}}{\Rightarrow} \\
&g_{ms} \cdot F_{t_{spec}} = \frac{g_{ms} \cdot G_{spec}}{2\pi \cdot n \cdot C_{GG_e}} \Rightarrow \\
&F_{t_{spec}} = \frac{G_{spec}}{2\pi \cdot n \cdot C_{G_{eW}} \cdot W} \Rightarrow \\
&F_{t_{spec}} = \frac{I_{spec_\square} \cdot \frac{W}{L}}{2\pi \cdot n\, U_T \cdot C_{G_{eW}} \cdot W} \Rightarrow \\
&F_{t_{spec}} = \frac{I_{spec_\square}}{2\pi \cdot n\, U_T \cdot C_{G_{eW}} \cdot L},
\end{aligned} \tag{9.36}$$

which scales roughly as $1/L$. It is also interesting to point out that the absolute value of the maximum value of F_t is given by:

$$F_{t_{max}} = \frac{F_{t_{spec}}}{\lambda_c} \overset{Eq.\,(9.4)}{=} \frac{v_{sat} \cdot C_{ox}}{2\pi \cdot C_{G_{eW}}}, \tag{9.37}$$

showing that $F_{t_{max}}$ does not scale with $1/L$ anymore, meaning that the only way to increase $F_{t_{max}}$ is to increase C_{ox} and/or decrease $C_{G_{eW}}$ [138]. This is an explanation why the improvement of the $F_{t_{max}}$ is slowing down since a few recent technology nodes compared to earlier generations.

9.6.1 Extraction of Parameters - $C_{GG_{eW}}$

Since, $F_t = g_{ms} \cdot F_{t_{spec}}$ and $F_{t_{spec}}$ is given by Eq. (9.36), we have all the elements to model F_t, (the expression of g_{ms} and the values of n, $I_{spec_\square}$ and λ_c), except for the value of the parameter $C_{G_{eW}}$. We can extract the value of $C_{G_{eW}}$ if we plot F_t vs. IC, with both axes in logarithmic scale, draw the weak-inversion asymptote and calculate $F_{t_{spec}}$ as the point at which the asymptote meets the $IC = 1$ line. The extraction of $F_{t_{spec}}$ is shown graphically in Fig. 9.18. After extracting

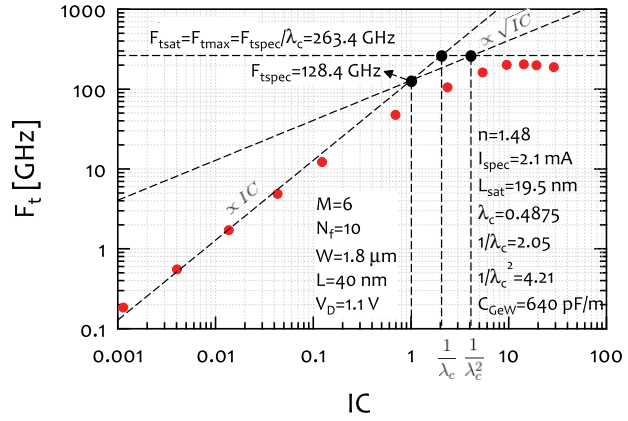


Figure 9.18 – Transit frequency F_t vs. IC and extraction of the parameter $C_{G_{eW}}$. First, $F_{t_{spec}}$, which corresponds to the value of the WI asymptote at $IC = 1$, is extracted and then $C_{G_{eW}}$ is calculated through Eq. (9.38). The measurements correspond to an RF DUT having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, at $V_D = 1.1$ V. The values of n, $I_{spec_\square}$ and λ_c should have been extracted already.

$F_{t_{spec}}$ we can calculate $C_{G_{eW}}$ through:

$$C_{G_{eW}} = \frac{F_{t_{spec}} \cdot 2\pi \cdot n\, U_T \cdot L}{I_{spec_\square}}. \tag{9.38}$$

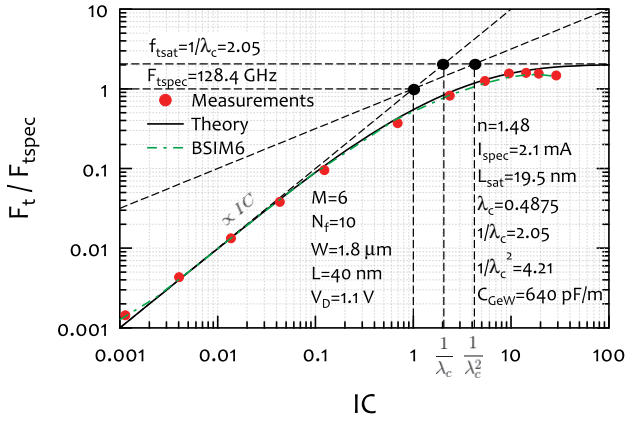
In the same figure, we see clearly that F_t is proportional to IC in WI, while it follows the asymptote $\sqrt{IC}$ between $IC = 1/\lambda_c$ and $IC = 1/\lambda_c^2$ and saturates at the value $F_{t_{max}} = F_{t_{spec}}/\lambda_c$ in strong-inversion.

9.6.2 Model Verification

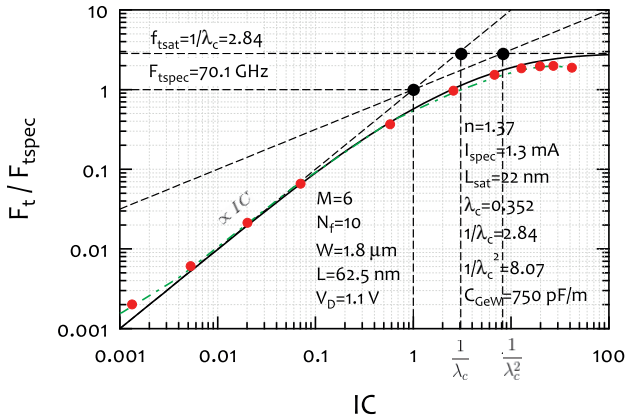
After the extraction of the $C_{G_{eW}}$ parameter we can now proceed to the comparison of the analytical model against measurements. In Fig. 9.19 and Fig. 9.20 we see the normalized transit frequency f_t vs. IC. The model is compared against measurements of three devices for the two processes under study namely, 40 nm and 28 nm. For the 40 nm process the BSIM6 model is also shown. From the figures we see that the analytical model is able to capture the transit frequency with sufficient accuracy for both technologies, with only one additional parameter to those extracted for modeling the G_m.

9.7 Modeling the $G_m/I_D \cdot F_t$ RF FoM

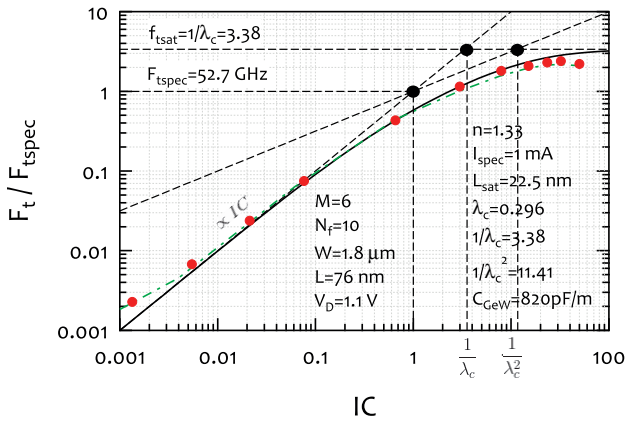
Both G_m/I_D and F_t are very important metrics/FoMs from an analog/RF design perspective: the former characterizes the DC performance of a device while the latter characterizes its high-frequency performance. In low-power operation we should target for high G_m/I_D meaning



(a) $L = 40$ nm

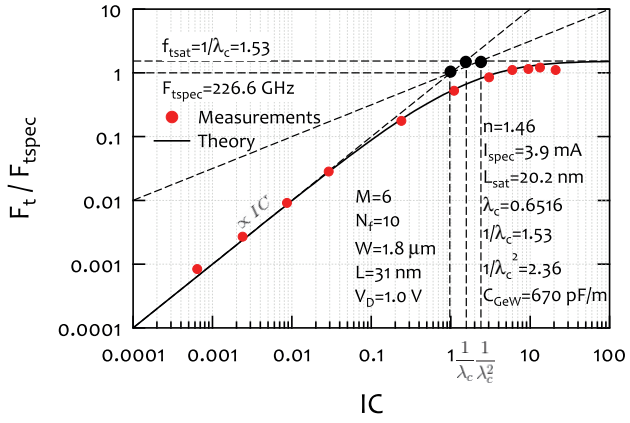


(b) $L = 62.5$ nm

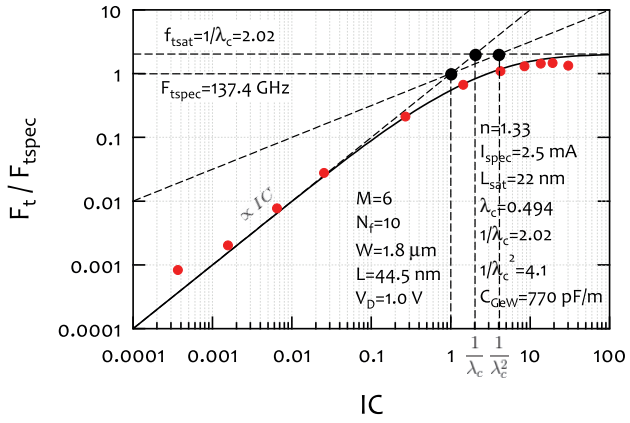


(c) $L = 76$ nm

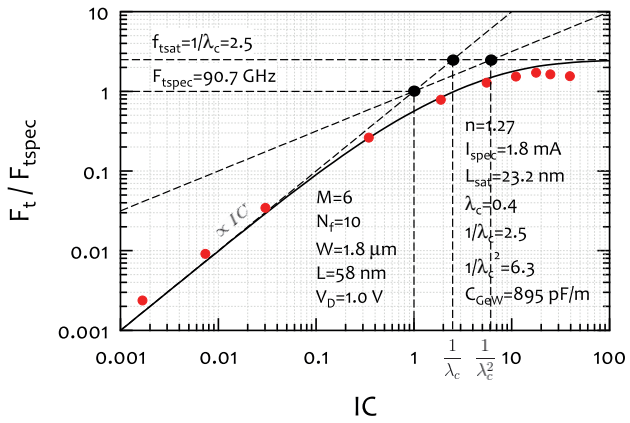
Figure 9.19 – Normalized transit frequency f_t vs. IC for three RF DUTs of a 40 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, 62.5 nm, 76 nm, at $V_D = 1.1$ V. Here, a comparison between theory, measurements and the BSIM6 model is made. The extracted values of the parameters of each device are also shown.



(a) $L = 30$ nm



(b) $L = 44.5$ nm



(c) $L = 58$ nm

Figure 9.20 – Normalized transit frequency f_t vs. IC for three RF DUTs of a 28 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 31$ nm, 44.5 nm, 58 nm, at $V_D = 1.0$ V. Here, a comparison between theory and measurements is made. The extracted values of the parameters of each device are also shown.

small values of *IC* in order to maximize the efficiency. This inevitably means a compromise in speed (gain-bandwidth) since F_t remains quite low at low *IC* values. This fundamental trade-off between the two quantities is revealed if we take a closer look at Fig. 9.5 and Fig. 9.18. So, we could define an FoM by combining these two quantities that have their maxima on the opposite ends of the *IC* axis. This FoM is then given by the product of the G_m/I_D and F_t and serves as a guide to locate the optimum *IC* [139].The normalized $G_m/I_D \cdot F_t$ FoM is defined as [88]

$$fom_{rf} = \frac{g_{ms} \cdot f_t}{i_d} = \frac{g_{ms}^2}{i_d}. \tag{9.39}$$

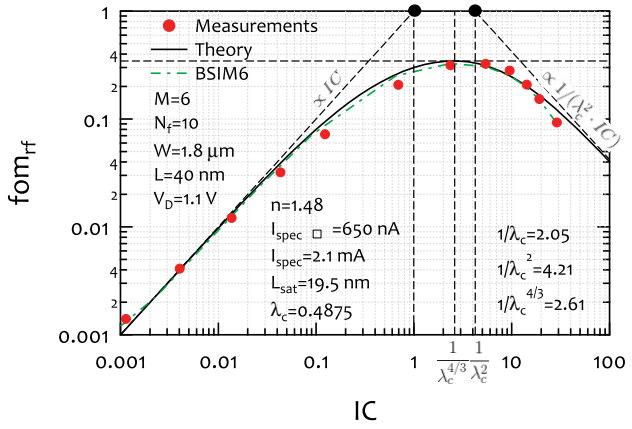
In Fig. 9.21 and Fig. 9.22 the normalized FoM fom_{rf} vs. *IC* is shown. As it can be seen from these figures, fom_{rf} shows a peaking behavior [54, 138, 139] and thus can be used to locate the optimum *IC*. This is because of the degradation of G_m and G_m/I_D in SI due to the effect of velocity saturation [88, 130]. In the absence of velocity saturation (for long-channels), in SI, G_m (consequently f_t) and G_m/I_D are respectively proportional to $\sqrt{I_D}$ and $1/\sqrt{I_D}$, implying that the FoM would simply saturate in this region. The peak of the fom_{rf} lies at the higher end of the moderate-inversion region for longer devices [88, 134], and moves deeper into the moderate-inversion region with decreasing channel lengths, as shown in Fig. 9.21 and Fig. 9.22. In the same figures we see also that in WI, $g_{ms}/i_d = 1$ and $f_t = IC$, so $fom_{rf} = IC$, while in SI inversion accounting for the effect of velocity saturation, $g_{ms}/i_d = 1/(\lambda_c \cdot IC)$, $f_t = 1/\lambda_c$ resulting in $fom_{rf} = 1/(\lambda_c^2 \cdot IC)$. In [129] it was shown that the peak of the fom_{rf} lies around $IC = 1/\lambda_c^{4/3}$, which is also verified by Fig. 9.21 and Fig. 9.22. That means that we could define an optimum *IC* as $IC_{opt} = 1/\lambda_c^{4/3}$. The fact that this IC_{opt} lies in moderate-inversion is attractive for the low-power RF design.

From a low-power RF design perspective, it was shown that this FoM can be interpreted as the ratio of the small-signal voltage gain over the product of noise times current of a simple common-source (CS) stage loaded by an identical stage [88]. Maximizing this FoM therefore means maximizing the gain while minimizing the noise and current.

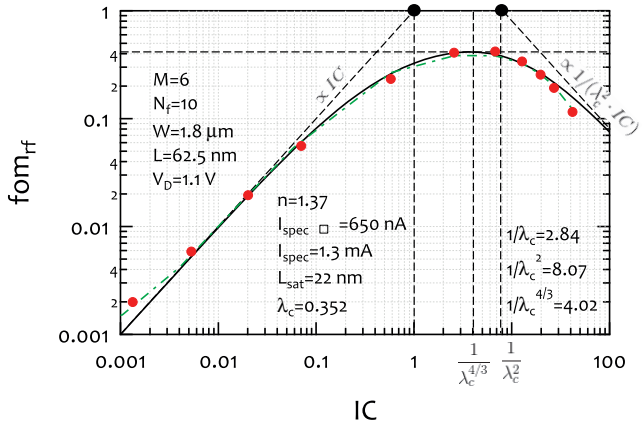
9.8 Modeling the Noise Factor - *F*

High F_t is not the only desired aspect. Low-noise is also a requirement for realistic RF circuits, especially when operating in low-power. Modeling the RF noise, thus, can prove to be a useful feature. In this section, we will try to provide a simple analytical expression to model the noise factor - *F*. The noise factor is the ratio of the total output noise to the input noise resulting only from the noise at the two-port input and can be written in terms of the four RF noise parameters as [39, 140]:

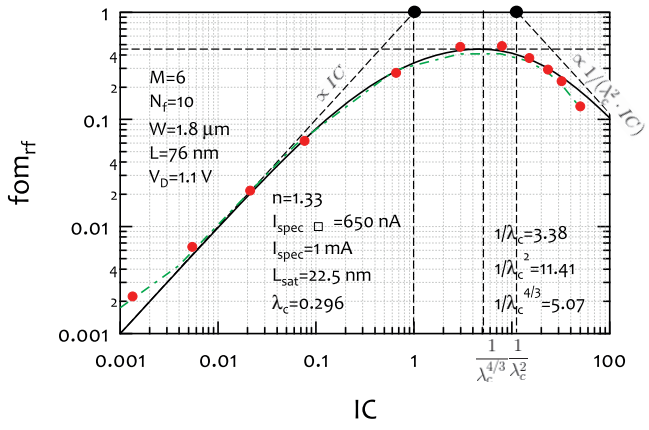
$$F = F_{min} + \frac{R_n}{G_s} \cdot \left[\left(G_s - G_{opt} \right)^2 + \left(B_s - B_{opt} \right)^2 \right]. \tag{9.40}$$



(a) $L = 40$ nm

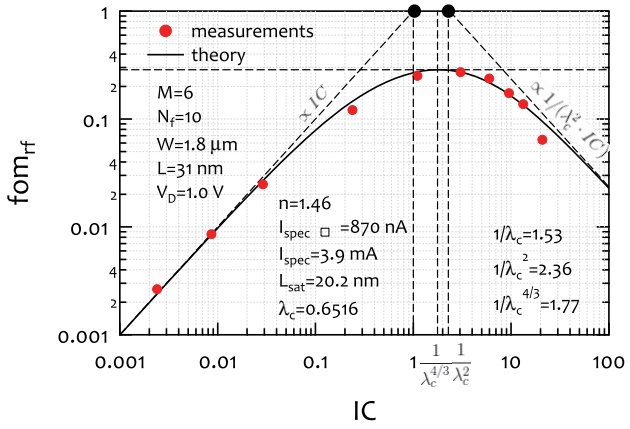


(b) $L = 62.5$ nm

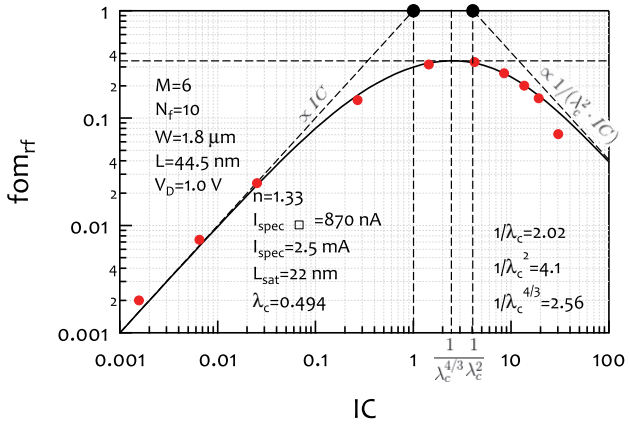


(c) $L = 76$ nm

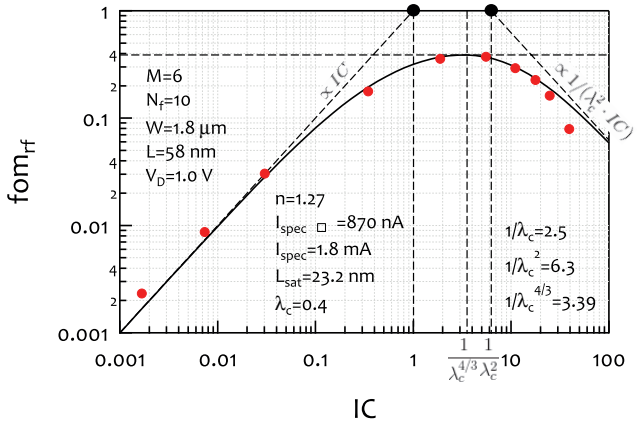
Figure 9.21 – fom_{rf} vs. IC for three RF DUTs of a 40 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, 62.5 nm, 76 nm, at $V_D = 1.1$ V. Here, a comparison between theory, measurements and the BSIM6 model is made.



(a) $L = 30$ nm



(b) $L = 44.5$ nm



(c) $L = 58$ nm

Figure 9.22 – fom_{rf} vs. IC for three RF DUTs of a 28 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 31$ nm, 44.5 nm, 58 nm, at $V_D = 1.0$ V. Here, a comparison between theory and measurements is made.

Combining Eq. (9.40) with the expressions of the four RF noise parameters, namely, R_{n}, G_{opt}, B_{opt}, F_{min}, as they where presented in Ch. 4 with Eq. (4.31) [5] and assuming a constant source impedance $Z_S = R_S = 1/G_{\mathrm{s}} = 50\ \Omega$, we can calculate a simplified expression for the noise factor as:

$$F = 1 + \frac{1}{50\ \Omega} \cdot \left(\frac{\gamma_{\mathrm{nD}}}{G_{\mathrm{m}}} + R_{\mathrm{G}} \right). \tag{9.41}$$

From the above expression, we see that in order to analytically model F we lack the values of γ_{nD} and R_{G} (G_{m} has been already calculated in Sec. 9.2).

9.8.1 Extraction of Parameters - R_{G}, γ_{nD}

The methodology to extract the values of R_{G} and γ_{nD} directly from measurements, at any V_{G} bias in saturation, was thoroughly demonstrated in Ch. 3, Ch. 4 and [5]. In Fig. 3.3c and Fig. 4.4a, we saw that R_{G} remains almost constant with respect to *IC*, but γ_{nD} demonstrates an *IC* dependence.

Since in this chapter we do not aim for excellent accuracy, but for simple, handy expressions, we can make even more simplifications. Regarding R_{G} we can use a constant value equal to the mean value of the extracted R_{G} parameter across *IC* or even use one extracted value of the R_{G} at any *IC* using Eq. (3.12a). The latter might decrease the accuracy of the analytical simplified expression, but it requires only one extracted value of the R_{G}. In our case we preferred to use the mean value of the R_{G} as demonstrated in Fig. 3.3c, which would provide us with better accuracy. Finally, we need to model the *IC* dependence of γ_{nD}. From Fig. 4.4 we see that from moderate- to strong-inversion the γ_{nD} parameter shows an almost linear *IC* dependence, which at low levels of inversion tends to 1, and can be approximated by:

$$\gamma_{\mathrm{nD}} = 1 + \alpha_{\gamma_{\mathrm{nD}}} \cdot IC. \tag{9.42}$$

So, in order to have a simple expression of γ_{nD} across *IC*, we need to extract the value of the $\alpha_{\gamma_{\mathrm{nD}}}$ parameter. To do so, we can extract the value of γ_{nD} at any *IC* using Eq. (4.33a) and then calculate $\alpha_{\gamma_{\mathrm{nD}}}$ using:

$$\alpha_{\gamma_{\mathrm{nD}}} = \frac{\gamma_{\mathrm{nD}} - 1}{IC}, \tag{9.43}$$

which results from Eq. (9.42).

In our case we used the extracted value of γ_{nD} at the highest available *IC* point in strong inversion. In Fig. 9.23, we demonstrate how we extracted the $\alpha_{\gamma_{\mathrm{nD}}}$ parameter and we compare γ_{nD} extracted from measurements (the values are the same as in Fig. 4.4a) against the γ_{nD} given by Eq. (9.42). We see that assuming a linear *IC* dependence for γ_{nD} is a quite good approximation.

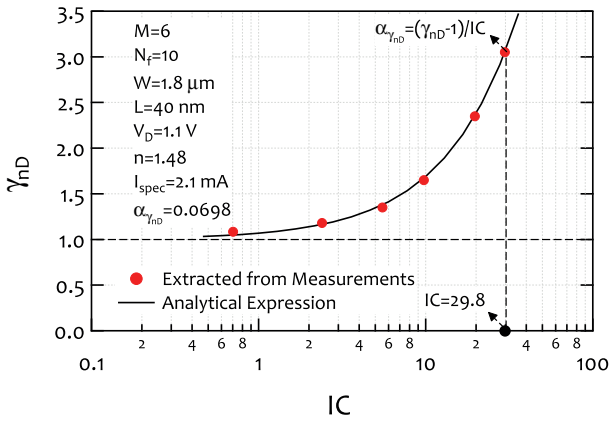


Figure 9.23 – γ_{nD} vs. IC and extraction of the parameter $\alpha_{\gamma_{\mathrm{nD}}}$. First, γ_{nD} at an IC_{spot} needs to be extracted using CITE EQUATION, and then $\alpha_{\gamma_{\mathrm{nD}}}$ can be calculated. The RF DUT which was used has $M = 6$, $N_{\mathrm{f}} = 10$, $W_{\mathrm{f}} = 1.8$ µm and $L = 40$ nm, at $V_{\mathrm{D}} = 1.1$ V. The values of n and $I_{\mathrm{spec}_{\square}}$ should have been extracted already.

Combining all the elements of the above discussion, F becomes:

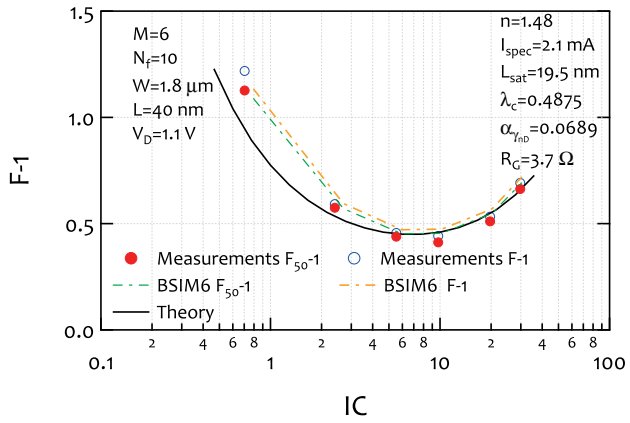
$$F = 1 + \frac{1}{50\ \Omega} \cdot \left(\frac{1 + \alpha_{\gamma_{\mathrm{nD}}}|_{\mathrm{IC_{spot}}} \cdot IC}{G_{\mathrm{m}}} + R_{\mathrm{G}}|_{\mathrm{IC_{spot}}} \right), \tag{9.44}$$

which requires the extraction of two parameters, namely R_{G} and $\alpha_{\gamma_{\mathrm{nD}}}$, at an $IC = IC_{\mathrm{spot}}$ (note that we do not have to use the same IC_{spot} for both R_{G} and $\alpha_{\gamma_{\mathrm{nD}}}$), in order to model the noise factor between moderate and strong-inversion.

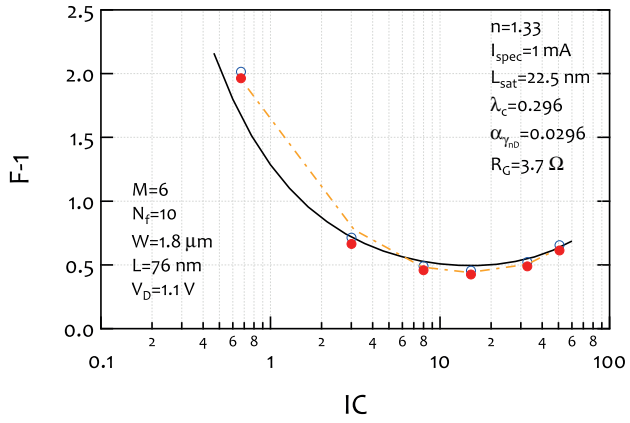
9.8.2 Model Verification

After the simplification of the noise factor F, we can proceed to the comparison of the analytical model against measurements. In Fig. 9.24 we see the $F-1$ vs. IC. We have chosen to show $F-1$ because this is the noise that is generated by the transistor itself. Thus it is a good way to see how a single device contributes to the total output noise. The model is compared against measurements and the BSIM6 model for two devices of a 40 nm CMOS process for $F-1$ and $F_{50}-1$. The difference between $F-1$ and $F_{50}-1$, is that the former is measured at a varying Z_{S} around 50 Ω (as shown in Fig. 4.7b), while the latter is calculated at a constant source impedance $Z_{\mathrm{S}} = R_{\mathrm{S}} = 50\ \Omega$ using (4.34). Nevertheless, we see that there is not a big discrepancy between the two. For the calculation of $F-1$ and $F_{50}-1$ the mean value across frequency at each bias point was used.

In Fig. 9.24, we clearly see that that the minimum $F-1$ is located at the lower end of strong-inversion for longer devices and moves towards moderate-inversion as the channel length is decreased. This adds one more argument in favor of biasing the transistor in MI for RF circuits, when targeting for low-power and low-noise operation [5, 54, 86–88, 141].



(a) $L = 40$ nm



(b) $L = 76$ nm

Figure 9.24 – RF noise $F-1$ vs. IC for two RF DUTs of a 40 nm CMOS process having $M = 6$, $N_f = 10$, $W_f = 1.8$ µm and $L = 40$ nm, 76 nm, at $V_D = 1.1$ V. Here, a comparison between theory, measurements and the BSIM6 model is made. For measurements and simulations both $F-1$ and $F_{50}-1$ are demonstrated, while the values at each *IC* correspond to the mean values across frequency. The extracted values of the parameters of each device are also shown.

9.9 Single-Transistor Common-Source Amplifier

Advanced nanoscale transistors may offer an impressive F_t but there are applications that operate in the low GHz range and do not necessarily require such high transit frequencies. For these applications it is possible to employ longer than the minimum length transistors which, even though exhibit lower speed, are less affected by short-channel effects. The question that arises is how to choose the optimal L of the transistor whilst meet the specifications posed by the application. In an attempt to define some design guidelines that can help to answer the above question, a simple single-MOS capacitively loaded common-source amplifier, as shown in Fig. 9.25, is used. From the following analysis certain useful remarks, related to the behavior of the mos transistor, are drawn.

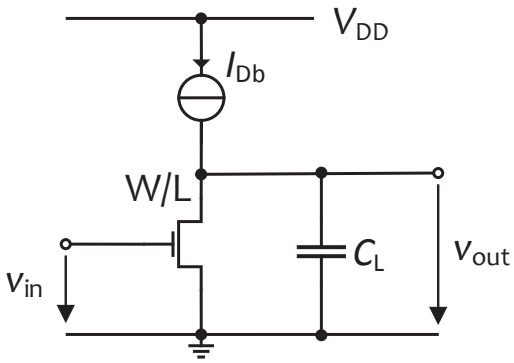


Figure 9.25 – Single-MOS capacitively loaded CS amplifier.

Using the BSIM6 RF model card, which was extracted for the 40 nm CMOS process, several simulations for this CS amplifier were carried out. Based on the *IC* design methodology, I_{D_b} (the drain bias current), *IC* and *L* are considered the unknown design parameters. When these three design parameters are defined, *W* can be calculated using:

$$W = \frac{I_{D_b} \cdot L}{I_{spec_\square} \cdot IC}. \tag{9.45}$$

For short-channel devices the output conductance G_{ds} and the parasitic capacitances, especially those that affect directly the output, have an important contribution in the behavior of the transistor and consequently they cannot be ignored. Along these line, and neglecting the rest of the elements of the circuit, the small-signal voltage gain A_v of the circuit, that is loaded with a capacitance C_L, can be approximated as:

$$A_v = \frac{A_{v_i}}{\frac{j\omega \cdot C_{tot}}{G_{ds}} + 1}, \tag{9.46}$$

where A_{v_i} is given by Eq. (9.19) and C_{tot} is:

$$C_{tot} = WL \cdot C_{GD_{WL}} + C_L. \tag{9.47}$$

In Eq. (9.47), $C_{GD_{WL}}$ is the total gate-to-drain capacitance per unit area.

To validate that Eq. (9.46) is indeed able to predict the A_v of this simple CS amplifier, in Fig. 9.26 the A_v vs. frequency, at a specific I_{D_b}, *IC* and *L*, that results from the simulation of the circuit using BSIM6 (with the extracted RF model card) is compared with the mathematical

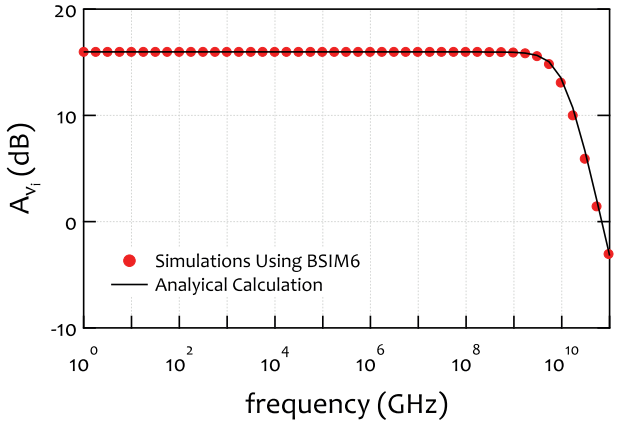


Figure 9.26 – A_{v} vs. frequency of a single-MOS capacitively loaded CS amplifier at $IC_{\mathrm{spot}} = 9.558$ and $I_{\mathrm{D_{bspot}}} = 1.778$ mA for a short channel device with $L = 40$ nm. Here a comparison between simulations using BSIM6 and the mathematical calculation based on Eq. (9.46) is made.

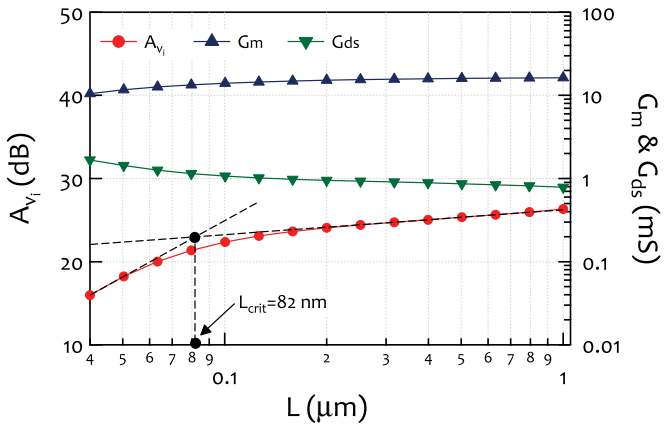


Figure 9.27 – $A_{\mathrm{v_i}}$, G_{m}, G_{ds}vs. L of a single-MOS capacitively loaded CS amplifier at $IC_{\mathrm{spot}} = 9.558$ and $I_{\mathrm{D_{bspot}}} = 1.778$ mA.

calculation.

For a specific $I_{\mathrm{D_b}}$ and IC the intrinsic gain $A_{\mathrm{v_i}}$ ($\omega = 0$) is decreasing as we move to shorter channel devices. This occurs due to the impact of the short-channel effects that degrade both the transconductance and the output conductance of the transistor, as shown in Fig. 9.27. In that plot, G_{m}, G_{ds} and $A_{\mathrm{v_i}}$ vs. L are shown. Since $I_{\mathrm{D_b}}$ and IC need to be kept constant for all points, the W changes proportionally to L, meaning that the W/L ratio is also constant. Taking this into account, in the G_{m} and G_{ds} vs. L curves, we observe the actual influence of the short-channel effects on the device behavior. Regarding $A_{\mathrm{v_i}}$, we can draw the long- and the short-channel asymptotes and we see that they cross at a point that corresponds to a channel length which we call L_{crit}. It is obvious from Fig. 9.27 that this L_{crit} defines the channel length below which the $A_{\mathrm{v_i}}$ starts to be affected strongly by short-channel effects.

If we now plot the L_{crit} vs. *IC* under the same I_{D_b} as shown in Fig. 9.28, we notice that L_{crit} does not vary significantly for different levels of inversion. Interestingly, this results from the fact that the two different slopes of the long- and the short-channel asymptotes of A_{v_i} vs. L remain almost constant across the *IC* range, although this is not the case for the slopes of the long- and the short-channel asymptotes of G_m and G_{ds} vs. L. The above can be further explained graphically through Fig. 9.29, which presents the different slopes that are calculated from the long- and short-channel asymptotes of A_{v_i}, G_m and G_{ds} when they are plotted vs. L with both axes in logarithmic scale, across *IC*. In Fig. 9.29a we see the long-channel asymptote slopes and we observe that for long-channel devices both G_m and G_{ds} scale, as expected, proportionally to $1/L$ across *IC*, while A_{v_i} remains almost length independent. The situation becomes more complicated when we analyze the short-channel asymptote slopes shown in Fig. 9.29b. There we see that the length scaling properties of G_m and G_{ds} differ between weak- and strong-inversion. As far as G_m is concerned in weak-inversion it scales proportionally to $1/L$ behaving similarly to the long-channel devices. In strong-inversion though, it becomes length independent due to the effect of velocity saturation. On the other hand G_{ds} is demonstrating a strong length dependence equal to $1/L^2$ in weak-inversion due to the effects of DIBL and CLM, while this dependence is progressively reduced to $1/L$ for strong-inversion. Nevertheless, and no matter what are the mechanisms that affect differently weak- and strong-inversion regions of the short-channel asymptote slopes of G_m and G_{ds}, the intrinsic gain A_{v_i} of short-channel devices scales always almost proportionally to $1/L$ across the whole *IC* range, resulting only in slight variations of L_{crit}. As it was mentioned earlier, W changes proportionally to L, so in order to derive, in Fig. 9.29, these slopes and isolate only the length-scaling contribution, the proportional dependence of W on L has been ignored. This is especially important for G_m and G_{ds} but not for A_{v_i} in which the W/L ratio that is inherently in both G_m and G_{ds} is eliminated. This is explained further using the following example. In Fig. 9.27 we observe that for long-channel devices G_m vs. L remains almost constant. This is due to the fact that no matter the value of the L, W is always proportional to it (remember $G_m \propto W/L$). However, this does not reflect the real L dependence of G_m. So, in Fig. 9.29a, in which the fact that $W \propto L$ is ignored, we end up with a value of -1 for the long-channel asymptote slope of G_m, meaning that G_m shows a length dependence equal to $1/L$.

In a real design, a circuit should provide a desired intrinsic gain and operate under a specific bias current. From L_{crit} we can know below which channel length the A_{v_i} starts to be affected by short-channel effects, but how to size the transistor when a specific A_{v_i} and I_{D_b} are defined is a different question. In Fig. 9.30 the A_{v_i} contours for various *IC* and L is demonstrated. In this plot we see that in order to achieve a larger intrinsic gain we would require a longer, and therefore a wider transistor (W and L scale proportionally). But, we further observe, that the minimum of each contour corresponds to an *IC* in the moderate-inversion region. This means that there is a minimum L for which the desired A_{v_i} can be achieved and if this L is chosen, the transistor should be biased in low moderate-inversion region, and this appears to be true regardless of the A_{v_i} level.

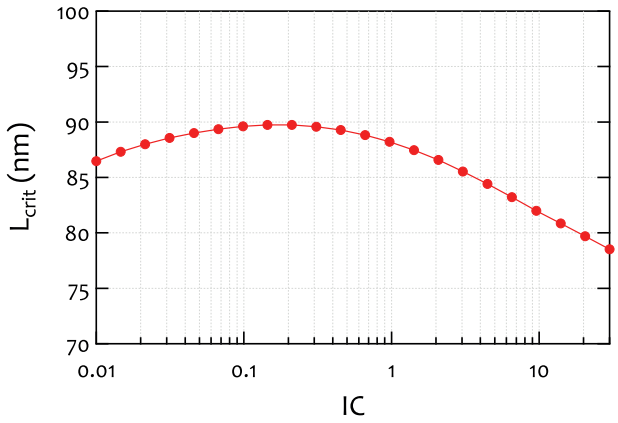
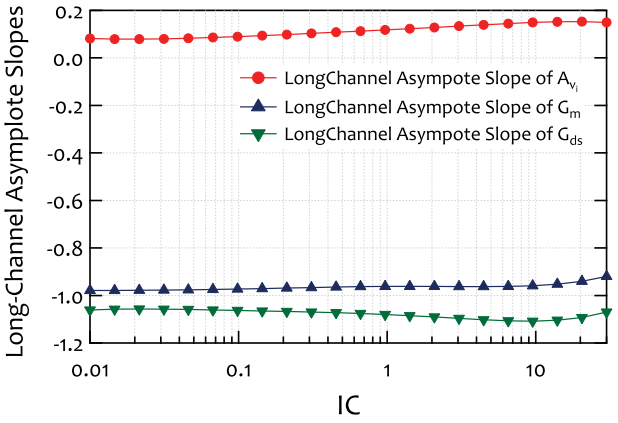
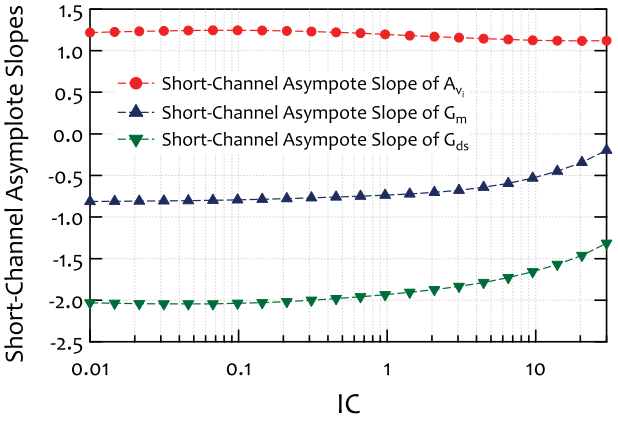


Figure 9.28 – L_{crit} vs. IC of a single-MOS capacitively loaded CS amplifier at $I_{\text{D}_{\text{bspot}}}$ = 1.778 mA.



(a) long-channel



(b) short-channel

Figure 9.29 – Slopes of the (a) long- and (b) short-channel asymptotes of A_{v_i}, G_{m} and G_{ds} vs. L in log-log scale across IC of a single-MOS capacitively loaded CS amplifier at IC_{spot} = 1.778 mA.

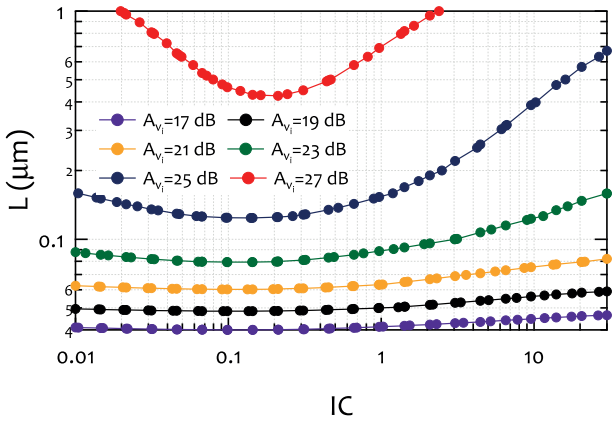


Figure 9.30 – A_{v_i} contours for different IC and L of a single-MOS capacitively loaded CS amplifier at $I_{D_{bspot}}$ = 1.778 mA. We observe that the minima of the contours always fall in the lower part of the moderate-inversion region.

In addition to achieving the required A_{v_i}, an RF circuit has also limitations regarding its bandwidth (BW) or gain-bandwidth (GBW). In this case study, where we investigate circuits that operate in the low GHz range, we set the specifications for the BW and GBW to 2.4 GHz and 16.8 GHz (7 x BW), respectively and we opt for operation in the $A_v = A_{v_i}$ part. The BW is defined by the 3-dB corner frequency F_c, calculated by:

$$F_c = \frac{G_{ds}}{2\pi \cdot C_{tot}}, \tag{9.48}$$

while the GBW is defined by the transit frequency F_t, given by:

$$F_t = \frac{G_m}{2\pi \cdot C_{tot}}. \tag{9.49}$$

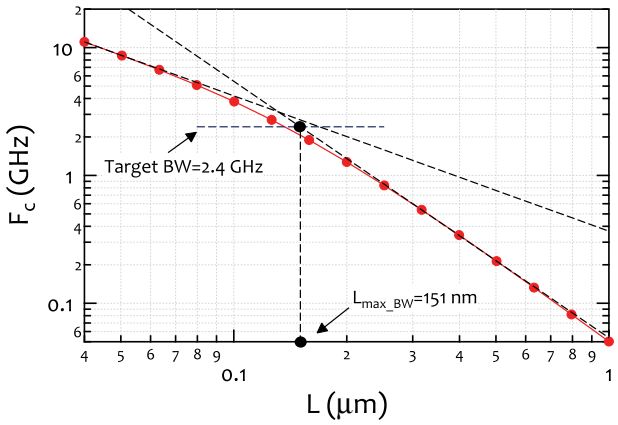
Similar to the analysis followed for the A_{v_i}, we can plot the F_c and F_t vs. L for a specific I_{D_b} and IC, as shown in Fig. 9.31. For both F_c and F_t, we can draw the long- and short-channel asymptotes and then check for which length the corresponding asymptote crosses the target BW and GBW, respectively. We can then define L_{max_BW} and L_{max_GBW} as the maximum lengths that can be selected to size the transistor in order to achieve the specified BW or GBW. In Fig. 9.32 the L_{max_BW} and L_{max_GBW} vs. IC are plotted and we observe that the minimum IC_{min} for both metrics lies in the limit between weak- and moderate-inversion. An IC lower than IC_{min} is not an option since it would require a transistor shorter than the L_{min} of the process. We also see that the BW specification sets more strict limits compared to the GBW, since the L_{max_BW} remains considerably lower compared to L_{max_GBW}, providing less design flexibility. Taking L_{crit} into account, which is also plotted in Fig. 9.32, we notice that in the lower part of moderate-inversion region it is necessary to choose a transistor with $L < L_{crit}$ in order to achieve the required BW or GBW. Nonetheless, for $IC > 1$ we can choose a transistor that

its A_{v_i} is not influenced by short-channel effects. In other words, it is possible to achieve a specific, relatively low BW or GBW without designing using minimum length devices, avoiding in this way the degradation that results from short-channel effects, and still remain in the moderate-inversion region.

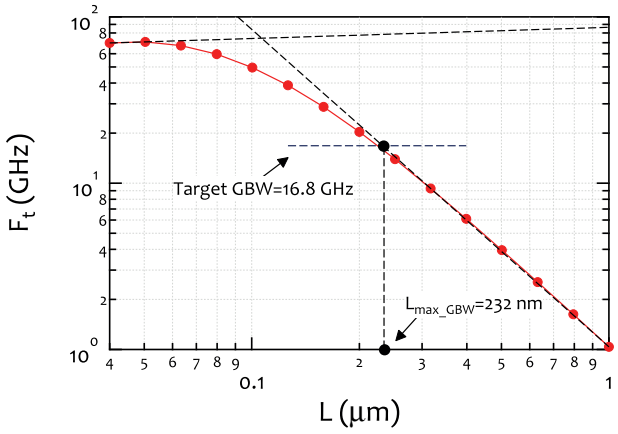
In this last part of the analysis, the impact of C_{tot} on F_c and F_t should not be underestimated. For small devices the load capacitance C_L is mainly contributing to C_{tot}, but for larger devices, usually when the operation of the transistor is pushed towards very low levels of inversion or when the required intrinsic gain is high (L should be high), $WL \cdot C_{GD_{WL}} \gg C_L$ and so the parasitic and internal MOS capacitances have a leading role over the load. This of course influences the scaling properties of F_c and F_t. For small devices F_c and F_t follow the scaling properties of G_{ds} and G_m respectively, but for larger ones the L factor introduced by the C_{GD} capacitance should be added. To give an example, and taking into account the length dependencies of G_m an G_{ds} presented in Fig. 9.29, the short-channel asymptote of F_c vs. L in weak-inversion would demonstrate $1/L^3$ length dependence, whereas in strong-inversion it would maintain the $1/L$ length dependence of G_{ds} (in this example and in order to isolate the length dependence the proportionality of W is not accounted for).

9.10 Conclusions

In this chapter, the concept of the inversion coefficient IC as the main design parameter covering the whole range of points from weak- to strong-inversion was used. Several simple analytical expressions that are able to model different analog/RF design FoMs with only a few parameters are presented. These expressions might not account for all the physical phenomena that exist in nanoscale devices, but they are easy to use and offer a handy procedure for the extraction of their parameters. The comparison of the models with measurements from two commercial bulk CMOS processes, namely, 28 nm and 40 nm, and the BSIM6 compact model and the excellent results that are demonstrated, prove that these simple formulas can be indeed used for advanced nanoscale devices providing a valuable guidance during the design procedure. As a final part of this chapter, a simple case of a single-MOS capacitively loaded CS amplifier was investigated based on the IC design methodology. Different metrics were studied for their dependencies on L and IC, demonstrating once more the advantages of moderate-inversion when the speed requirements of an RF circuit are not that high.



(a) F_c



(b) F_t

Figure 9.31 – F_c and F_t vs. L of a single-MOS capacitively loaded CS amplifier at IC_{spot} = 9.558 and $I_{D_{b_{spot}}}$ = 1.778 mA.

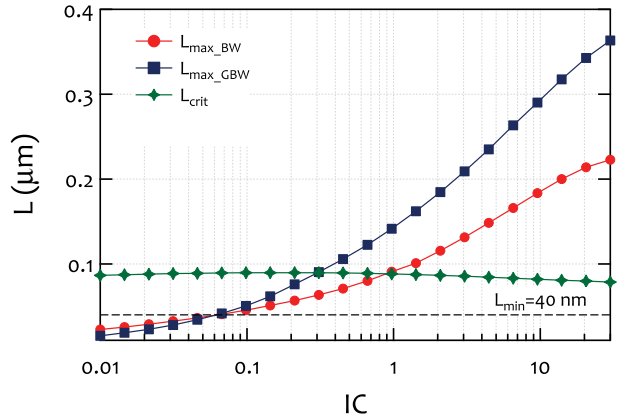


Figure 9.32 – L_{crit}, L_{max_BW} and L_{max_GBW} vs. IC of a single-MOS capacitively loaded CS amplifier at $I_{D_{b_{spot}}}$ = 1.778 mA.

10 Conclusion

Moore's law, in spite of being an observation based on the information and the trends of the initial steps of the development of microelectronics, managed to dictate the pace that was obeyed by the semiconductor industries for the following decades up to today. To adhere to this law was not a trivial task that merely required the downscaling of CMOS technologies, but several innovative fabrication techniques and technology boosters had to come to the rescue. Despite all the difficulties, the semiconductor industries succeeded to surpass all the obstacles and reach for the contemporary bulk CMOS process a nominal channel gate length of 28 nm, which was inconceivable when Moore's law appeared.

And although such nanoscale devices are at the disposal of the IC designers, the societal and industrial present needs define a particular set of constraints and specifications that cannot be easily met. There is a wide range of applications that demand extra low-power operating conditions and are used in rapidly emerging fields, such as RF wireless communications and autonomous sensing systems. For these kinds of applications the speed and size of the device, which are the focal points of the technological development road-map, are not of principal importance. Nonetheless, they can also benefit from state-of-the-art record high performance, by trading-off this exceptional speed with lower power consumption. This can be achieved by shifting the operating point from the traditional maximum performance strong-inversion towards moderate- or even weak-inversion region. Even though designing in lower inversion levels does not exploit fully the speed capabilities of advanced nanoscale devices, the transit frequency remains high enough for the specifications posed by an ample range of low-power applications. At the same time, the power consumption at these operating conditions is reduced drastically.

Given the aforementioned facts, the target of this thesis has been to investigate, analyze, expand and evaluate the capabilities of the state-of-the-art modeling approaches that are used under this particular RF low-power perspective. Even though an extensive characterization and modeling effort has been made already, the particular intersection where RF operation meets weak-inversion for advanced nanoscale technologies has not been explored sufficiently.

10.1 Summary of Results

The conclusions of this work have been documented already in the main body of this manuscript and the more important points are summarized below.

10.1.1 Analytical Small-Signal RF Modeling

One main objective of this work was to model analytically the RF behavior of nanoscale devices, focusing especially on low levels of inversion. For this, a simple, first-order small-signal equivalent circuit of the transistor was employed. At first, the Y-parameters were modeled and then this circuit was expanded so as to include RF noise and model directly the four RF noise parameters. The comparison between the RF analytical expressions and the measurements attested that such a circuit, even though similar approaches have been long used for much older technologies, can still provide accurate results from weak- to strong-inversion region for state-of-the art devices. The analytical expressions were used to provide a step-by-step procedure to extract both the RF circuit components and the RF noise model parameters. Especially, for the extraction of the RF noise model parameters, such a procedure was presented for the first time. The overall structure of the analytical model and the parameter extraction procedure remains consistent with existing standard advanced compact MOSFET models. Using these analytical expressions the differences in the operation of the MOSFET between weak- and strong-inversion were highlighted, analyzed and successfully modeled. This analysis resulted into showing for the first time analytical expressions able to capture different RF FoMs for minimum channel length devices from deep weak-inversion to strong-inversion .

10.1.2 Compact Modeling with BSIM6

The functionality of the analog/RF circuit simulators has been drastically increased by the built-in integration of compact models. Yet, a compact model is not only useful from a design perspective, it is, furthermore, an invaluable tool when it comes to the characterization of a process. Another main task of this thesis has been the evaluation of the behavior of the state-of-the-art BSIM6 compact model by comparing it with measurements of advanced CMOS technologies. This study has resulted into the identification of its weaknesses and led to the development of a series of enhancements for displaying improved accuracy for nanoscale technologies. Under this framework, an important contribution to the model was the development of the geometrical scaling equations. It was shown that using one single empirical scaling expression for all the parameters except one (the parameter related to the channel doping) and thanks to the flexibility offered by the model, it was possible to capture with high accuracy the full geometrical scaling characteristics of advanced technology nodes. Note that in such technologies the channel length of the device extents over almost three

decades. It is also important to mention that the scaling equations that were adopted have no impact on the wide/long channel devices, minimizing this way the iterations required during the parameter extraction procedure. Furthermore, the subset of the model parameters on which the scaling rules applied was determined. The scaling equations that were chosen boosted the behavior of the model with respect to its previous versions. Besides, an efficient and almost straight-forward step-by-step parameter extraction procedure was established, which facilitates importantly the usage of the model. A section of the BSIM6 technical manual is dedicated on these parameter extraction guidelines. Finally, the model went through a broad evaluation in all modes of operation (CV, IV and RF), and its suitability for analog and RF design using state-of-the-art devices, even in very low-power conditions, was demonstrated.

10.1.3 Analytical Modeling of FoMs

In the last years there is an increased interest in the concept of the inversion coefficient as the main design parameter even for very advanced technology nodes. Within the scope of this thesis, several first-order analytical expressions that can be used to model different analog/RF design FoMs were either employed or developed. These expressions might not account for all the physical phenomena that exist in modern devices, but their beauty lays in the fact that they remain simple and they use only a few parameters that can be extracted easily from measurements, while also being sufficiently accurate. The comparison of the different analytical expressions against measurements, proved that these models can be indeed used even for very short devices. Their simplicity and their straightforwardness are their major advantages which allow them to be easily used by a wide range of related engineers other than model developers, who do not always master all the physics of the device, but they just need an overall insight in the device behavior. In this way such analytical expressions can offer a valuable guidance in various cases such as during the process characterization or design procedure.

10.1.4 General Remark

In overall, this work has demonstrated the promising potential of biasing nanoscale devices in moderate- or even weak- inversion in order to achieve low-power operation in RF. Based on the conclusions of this study, the key suggestion is that moderate-inversion region has been hiding a well-balanced trade-off in terms of gain, transit frequency, power consumption, noise and linearity, and promises an uncharted land to the designers to explore and exploit in order to meet the exhausting specifications of the advanced ultra-low-power RF systems.

10.2 Suggestions for Future Work

Based on the work carried out for this thesis, the following areas of research can be proposed as its natural continuation which would strengthen further its value and expand its results beyond the scope of a single Ph.D. dissertation. To make the points of future work that can stem out from this thesis more clear to the reader, a set of question has been added to the following discussion. *Which are the key questions that emerged after the completion of this study?*

- **Geometrical Scaling in RF**

  The geometrical scaling of the RF behavior is an aspect that would benefit from further investigation. The presented RF analytical expressions of the first part of this thesis could be enhanced to include the geometrical scaling properties of nanoscale RF devices. *How do the RF components and the RF noise model parameters scale across the width/length plane? Do they scale in the same way in all levels of inversion?* In addition, using the simple analytical expressions for the FoMs presented in the last part of the thesis, it would be interesting to study the geometrical scaling of the model parameters. *Are there simple expressions, even if empirical, to represent the scaling profiles of the different parameters?* BSIM6 could also be evaluated for its ability to model accurately the scaling properties of the transistor in RF operation. *Do the scaling equations regarding the parasitic RF components and the RF noise model parameters in BSIM6 need to be enhanced?*

- **Large-Signal RF Modeling**

  The large-signal RF behavior of nanoscale devices does also form an interesting field for research. *What are the modifications needed in the equivalent small-signal RF circuit so that it can be used for modeling the large-signal RF behavior? How are the Y-parameters and the four RF noise parameters influenced under large-signal operation? Is there a way to analytically model their expressions?*

- **Evaluation of Analytical Expression in Design Level**

  Although the analytical expressions in this thesis have been compared against measurements of real nanoscale devices, and this consists a strong and typical evaluation step for models, they could also be evaluated when used in design level. *Do they result in circuits that meet the required specifications when they are used to size and bias the transistor/transistors? Are there any guidelines that could be provided to designers when they need to make specific decisions during the design procedure? Among these guidelines are any of them generic enough so that they can be applied to any circuit?*

- **Extension of the Analytical Expressions to more Advanced Devices**

There is a popular estimation that the contemporary bulk CMOS technologies will be used for many more years to come for a plethora of applications. Nevertheless the semiconductor industry has already expanded its path to more advanced CMOS processes in order to keep up with the Moore's law. The main successors of the conventional bulk MOS transistor that have been already selected for the technology nodes beyond 28 nm are the FD-SOI (Fully-Depleted Silicon-on-Insulator) and the FinFet (Fin Field-Effect-Transistor) transistors. Since the development of models should keep up with the technology advancements, an interesting research area would be the exploration of the RF behavior of these devices from weak- to strong-inversion. *What are the changes needed in the small-signal equivalent schematic presented in this thesis in order to be capable to provide accurate analytical expressions? How do these analytical expressions differ from the ones used for contemporary bulk CMOS devices? How do the Y-parameters, the four RF noise parameters and the different RF FoMs behave? Are there any different trends between weak- and strong-inversion observed? The progress in the technology should be also considered in the IC design methodology. Can the IC design methodology be extended and used for these more advanced structures? What are the enhancements needed to cover also the behavior of these devices?*

▷ **RF Parameter Extraction in BSIM6**

From a more practical point of view, it is also essential that the parameter extraction methodology of BSIM6 is extended to cover the RF small-signal operation. *Can the extraction methodology for the RF circuit components and the RF noise model parameters presented in the first part of the thesis be enhanced, if necessary, and then introduced into BSIM6? What happens with the extraction of the values of the RF parasitic components when a more complicated equivalent small-signal circuit is chosen? In that case, is it possible to extract the values of the components directly from measurements?*

Bibliography

[1] G. Moore, "Cramming More Components Onto Integrated Circuits," *Proceedings of the IEEE*, vol. 86, pp. 82–85, Jan 1998.

[2] P. Dautriche, "Analog design trends and challenges in 28 and 20nm CMOS technology," in *Solid-State Device Research Conference (ESSDERC), 2011 Proceedings of the European*, pp. 1–4, Sept 2011.

[3] L. Poulain, N. Waldhoff, D. Gloria, F. Danneville, and G. Dambrine, "Small signal and HF noise performance of 45 nm CMOS technology in mmW range," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) 2011*, pp. 1–4, 2011.

[4] C.-H. Jan, M. Agostinelli, H. Deshpande, M. El-Tanani, W. Hafez, U. Jalan, L. Janbay, M. Kang, H. Lakdawala, J. Lin, Y.-L. Lu, S. Mudanai, J. Park, A. Rahman, J. Rizk, W.-K. Shin, K. Soumyanath, H. Tashiro, C. Tsai, P. Vandervoorn, J.-Y. Yeh, and P. Bai, "RF CMOS technology scaling in High-k/metal gate era for RF SoC (system-on-chip) applications," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, pp. 27.2.1–27.2.4, Dec 2010.

[5] **M.-A. Chalkiadaki** and C. Enz, "RF Small-Signal and Noise Modeling Including Parameter Extraction of Nanoscale MOSFET From Weak to Strong Inversion," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 63, pp. 2173–2184, July 2015.

[6] B. Pullman, *L'Atome dans l'histoire de la pensée humaine*, pp. 48–49. 75006, Paris, France: Fayard, 1995.

[7] D. Vasilescu and M. Adrian-Scotto, "From Democritus to Schrödinger: a reflection on quantum molecular modeling," *Structural Chemistry*, vol. 21, no. 6, pp. 1289–1314, 2010.

[8] L. De Forest, "The Audion; A New Receiver for Wireless Telegraphy," *American Institute of Electrical Engineers, Transactions of the*, vol. XXV, pp. 735–763, Jan 1906.

[9] L. De Forest, "Space telegraphy.," Feb. 18 1908. US Patent 879532.

[10] P. B. Helsdon, "Home-Made Transistors, Inexpensive Conversion of Selected Germanium Diodes," *Wireless World*, pp. 20–23, Jan 1954.

[11] J. Bardeen and W. Brattain, "The Transistor, A semi-conducting Triode," *Physical Review*, vol. 74, no. 2, pp. 230–231, 1948.

[12] J. E. Lilienfeld, "Device for controlling electric current," Mar. 28 1928. US Patent 1900018.

[13] R.-Y. Sah, "Evolution of the MOS transistor-from conception to VLSI," *Proceedings of the IEEE*, vol. 76, pp. 1280–1326, Oct 1988.

[14] C. Svensson, "Forty years of feature-size predictions (1962-2002)," in *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International*, pp. 35–36, Feb 2003.

[15] A. Kaneko, A. Yagishita, K. Yahashi, T. Kubota, M. Omura, K. Matsuo, I. Mizushima, K. Okano, H. Kawasaki, S. Inaba, T. Izumida, T. Kanemura, N. Aoki, K. Ishimaru, H. Ishiuchi, K. Suguro, K. Eguchi, and Y. Tsunashima, "Sidewall transfer process and selective gate sidewall spacer formation technology for sub-15nm finfet with elevated source/drain extension," in *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, pp. 844–847, Dec 2005.

[16] J. J. Liou and F. Schwierz, "RF MOSFET: recent advances, current status and future trends," *Solid-State Electronics*, vol. 47, no. 11, pp. 1881 – 1895, 2003.

[17] S. M. Sze and K. N. Kwok, *Physics of Semiconductor Devices*. New York: Wiley, 3rd ed., 2006.

[18] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*. Oxford University Press, 3rd ed., 2010.

[19] G. Gildenblat, ed., *Compact Modeling, Principles, Techniques and Applications*. Springer, New-York, 2010.

[20] W. Grabinski and D. Tomaszewski, eds., *Open Source CAD for Compact Modeling*. Springer, 1st ed., 2015.

[21] M. Fernandez-Barciela, P. Tasker, Y. Campos-Roca, M. Demmler, H. Massler, E. Sanchez, M. Curras-Francos, and M. Schlechtweg, "A simplified broad-band large-signal nonquasi-static table-based FET model," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 48, pp. 395–405, Mar 2000.

[22] M. Graham, J. Paulos, and D. Nychka, "Compact table-based MOSFET model," in *University/Government/Industry Microelectronics Symposium, 1993., Proceedings of the Tenth Biennial*, pp. 137–140, May 1993.

[23] C. T. Sah, "A History of MOS Transistor Compact Modeling," in *WCM 2005, Technical Proceedings of the 2005 Workshop on Compact Modeling, Chapter 5: A History of MOS Transistor Compact Modeling*, 1993.

[24] B. N. W. Grabinski and D. Schreurs, eds., *Transistor Level Modeling for Analog/RF IC Design*. Springer, 2006.

[25] L. W. Nagel and D. O. Pederson, "SPICE (Simulation Program with Integrated Circuit Emphasis)," in *Memorandum No. ERL-M382, University of California, Berkeley*, April 1973.

[26] A. Vladimirescu, "SPICE-the third decade," in *Bipolar Circuits and Technology Meeting, 1990., Proceedings of the 1990*, pp. 96–101, Sep 1990.

[27] S. Terry, J. Rochelle, D. Binkley, B. Blalock, D. Foty, and M. Bucher, "Comparison of a BSIM3V3 and EKV MOSFET model for a 0.5 mu;m CMOS process and implications for analog circuit design," *Nuclear Science, IEEE Transactions on*, vol. 50, pp. 915–920, Aug 2003.

[28] N. Paydavosi, T. H. Morshed, D. D. Lu, W. M. Yang, M. V. Dunga, X. J. Xi, J. He, W. Liu, Kanyu, M. Cao, X. Jin, J. J. Ou, M. Chan, A. M. Niknejad, and C. Hu, "Bsim4v4.8.0 mosfet model," tech. rep., Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, California, USA, 94720, 2013.

[29] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. Smit, A. Scholten, and D. Klaassen, "PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation," *Electron Devices, IEEE Transactions on*, vol. 53, pp. 1979–1993, Sept 2006.

[30] G. Gildenblat, H. Wang, T.-L. Chen, X. Gu, and X. Cai, "SP: an advanced surface-potential-based compact MOSFET model," *Solid-State Circuits, IEEE Journal of*, vol. 39, pp. 1394–1406, Sept 2004.

[31] R. V. Langevelde and F. M. Klaassen, "An Explicit Surface-Potential-Based MOSFET Model for Circuit Simulation,," *Solid-State Electronics*, vol. 3, pp. 409–418, 2000.

[32] "Support page of PSP," 2015. Technical University of Delft, http://psp.ewi.tudelft.nl/.

[33] M. Miura-Mattausch, N. Sadachika, D. Navarro, G. Suzuki, Y. Takeda, M. Miyake, T. Warabino, Y. Mizukane, R. Inagaki, T. Ezaki, H. Mattausch, T. Ohguro, T. Iizuka, M. Taguchi, S. Kumashiro, and S. Miyamoto, "HiSIM2: Advanced MOSFET Model Valid for RF Circuit Simulation," *Electron Devices, IEEE Transactions on*, vol. 53, pp. 1994–2007, Sept 2006.

[34] H. Oguey and S. Cserveny, "Modele du transistor MOS valable dans un grand domaine de courants," *Bull. SEW VSE*, Feb 1982.

[35] H. Oguey and S. Cserveny, "MOS modelling at low current density," *Summer Course on "Process and Device Modelling," ESAT Leuven-Heverlee, Belgium*, June 1983.

[36] H. Agarwal, C. Gupta, S. Khandelwal, J. P. Duarte, Y. S. Chauhan, S. Salahuddin, and C. Hu, "BSIM6.1.1 MOSFET Compact Model. Technical Manual," 2015.

[37] C. Galup-Montoro, M. Schneider, A. Cunha, F. de Sousa, H. Klimach, and O. Siebel, "The Advanced Compact MOSFET (ACM) Model for Circuit Analysis and Design," in *Custom Integrated Circuits Conference, 2007. CICC '07. IEEE*, pp. 519–526, Sept 2007.

[38] C. C. Enz, "A Short Story of the EKV MOS Transistor Model," *IEEE SSCS*, July 2008. On-Line Issue.

[39] C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling: The EKV model for low-power and RF IC design.* John Wiley & Sons, Ltd, 2006.

[40] M. Bucher, A. Bazigos, F. Krummenacher, J.-M. Sallese, and C. Enz, *Chapter 3: EKV3.0: An Advanced Charge Based MOS Transistor Model, in book: Transistor Level Modeling for Analog/RF IC Design.* Springer, 2006.

[41] C. C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, pp. 186–201, 2000.

[42] C. Enz, "An MOS Transistor Model for RF IC Design Valid in All Regions of Operation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 1, pp. 342–359, 2002.

[43] Y. Cheng, M. Deen, and C.-H. Chen, "MOSFET Modeling for RF IC Design," *IEEE Transactions on Electron Devices*, vol. 52, no. 7, pp. 1286–1303, 2005.

[44] **M.-A. Chalkiadaki** and M. Bucher, "Large-signal RF modeling with the EKV3 MOSFET model," *Journal of Telecommunications and Information Technology (JTIT)*, vol. 1, pp. 25 – 28, 2010.

[45] N. Waldhoff, C. Andrei, D. Gloria, F. Danneville, and G. Dambrine, "Small Signal and Noise Equivalent Circuit for CMOS 65 nm up to 110 GHz," in *38th European Microwave Conference (EuMC) 2008*, pp. 321–324, Oct 2008.

[46] H. Wong, S. Siu, K. Kakusima, and H. Iwai, "Modeling and characterization of radio-frequency characteristics of multi-finger nanometer MOS transistors," in *Electron Devices and Solid-State Circuits, 2009. EDSSC 2009. IEEE International Conference of*, pp. 3–7, Dec 2009.

[47] A. Bazigos, M. Bucher, P. Sakalas, M. Schroter, and W. Kraus, "High-frequency scalable compact modelling of Si RF-CMOS technology," *physica status solidi (c)*, vol. 5, no. 12, pp. 3681–3685, 2008.

[48] S. Asgaran, M. Deen, and C.-H. Chen, "Analytical Modeling of MOSFETs Channel Noise and Noise Parameters," *IEEE Transactions on Electron Devices*, vol. 51, no. 12, pp. 2109–2114, 2004.

[49] S. Asgaran, M. Deen, C.-H. Chen, G. Rezvani, Y. Kamali, and Y. Kiyota, "Analytical Determination of MOSFET's High-Frequency Noise Parameters From NF50 Measurements and Its Application in RFIC Design," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 5, pp. 1034–1043, 2007.

[50] J. Shi, Y. Z. Xiong, K. Kang, L. Nan, and F. Lin, "RF Noise of 65-nm MOSFETs in the Weak-to-Moderate-Inversion Region," *Electron Device Letters, IEEE*, vol. 30, pp. 185–188, Feb 2009.

[51] S.-C. Wang, P. Su, K.-M. Chen, K.-H. Liao, B.-Y. Chen, S.-Y. Huang, C.-C. Hung, and G.-W. Huang, "Comprehensive Noise Characterization and Modeling for 65-nm MOSFETs for Millimeter-Wave Applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 4, pp. 740–746, 2010.

[52] A. Scholten, G. Smit, R. Pijper, L. Tiemeijer, A. Mercha, and D. Klaassen, "FinFET compact modelling for analogue and RF applications," in *IEEE International Electron Devices Meeting (IEDM) 2010*, pp. 8.4.1–8.4.4, Dec 2010.

[53] G. Crupi, A. Caddemi, D. M.-P. Schreurs, W. Wiatr, and A. Mercha, "Microwave noise modeling of FinFETs," *Solid-State Electronics*, vol. 56, no. 1, pp. 18 – 22, 2011.

[54] A. Antonopoulos, M. Bucher, K. Papathanasiou, N. Makris, R. Sharma, P. Sakalas, and M. Schroter, "CMOS RF Noise, Scaling, and Compact Modeling for RFIC Design," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) 2013*, pp. 53–56, 2013.

[55] F. Danneville, L. Poulain, Y. Tagro, S. Lepilliet, B. Dormieu, D. Gloria, P. Scheer, and G. Dambrine, "RF and broadband noise investigation in High-k/Metal Gate 28-nm CMOS bulk transistor," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 27, no. 5-6, pp. 736–747, 2014.

[56] J.-P. Colinge, "Trends in Silicon-on-insulator Technology," in *Solid State Device Research Conference, 1992. ESSDERC '92. 22nd European*, pp. 795–802, Sept 1992.

[57] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm," *Electron Devices, IEEE Transactions on*, vol. 47, pp. 2320–2325, Dec 2000.

[58] M. Koolen, J. Geelen, and M. Versleijen, "An improved de-embedding technique for on-wafer high-frequency characterization," in *Bipolar Circuits and Technology Meeting, 1991., Proceedings of the 1991*, pp. 188–191, Sep 1991.

[59] K. Aufinger and J. Bock, "A Straightforward Noise De-Embedding Method and its Application to High-Speed Silicon Bipolar Transistors," in *Solid State Device Research Conference, 1996. ESSDERC '96. Proceedings of the 26th European*, pp. 957–960, Sept 1996.

[60] G. Dambrine, H. Happy, F. Danneville, and A. Cappy, "A New Method for On Wafer Noise Measurement," *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, pp. 375–381, Mar 1993.

[61] A. Abidi, "RF CMOS Comes of Age," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 549–561, Apr 2004.

[62] G. Dambrine, C. Raynaud, D. Lederer, M. Dehan, O. Rozeaux, M. Vanmackelberg, F. Danneville, S. Lepilliet, and J. Raskin, "What are the limiting parameters of deep-submicron MOSFETs for high frequency applications?," *Electron Device Letters, IEEE*, vol. 24, pp. 189–191, March 2003.

[63] U. Gogineni, J. del Alamo, and C. Putnam, "RF power potential of 45 nm CMOS technology," in *Silicon Monolithic Integrated Circuits in RF Systems (SiRF), 2010 Topical Meeting on*, pp. 204–207, Jan 2010.

[64] B. Dormieu, C. Charbuillet, F. Danneville, N. Kauffmann, and P. Scheer, "Millimeter-wave modeling of isolated MOS substrate network through gate-bulk measurements," in *Radio Frequency Integrated Circuits Symposium (RFIC), 2011 IEEE*, pp. 1–4, June 2011.

[65] U. Gogineni, H. Li, J. del Alamo, S. Sweeney, J. Wang, and B. Jagannathan, "Effect of Substrate Contact Shape and Placement on RF Characteristics of 45 nm Low Power CMOS Devices," *Solid-State Circuits, IEEE Journal of*, vol. 45, pp. 998–1006, May 2010.

[66] A. Tong, W. M. Lim, K. S. Yeo, C. B. Sia, and W. C. Zhou, "A Scalable RFCMOS Noise Model," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 57, pp. 1009–1019, May 2009.

[67] M. Bucher, A. Bazigos, S. Yoshitomi, and N. Itoh, "A scalable advanced RF IC design-oriented MOSFET model," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 18, no. 4, pp. 314–325, 2008.

[68] J. Gao and A. Werthof, "Scalable Small-Signal and Noise Modeling for Deep-Submicrometer MOSFETs," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 57, pp. 737–744, April 2009.

[69] J. Cha, J.-Y. Cha, D.-H. Jung, and S. Lee, "An improved RF MOSFET model including scalable gate resistance and external inductances," in *SoC Design Conference, 2008. ISOCC '08. International*, vol. 01, pp. I–431–I–432, Nov 2008.

[70] B. Dormieu, P. Scheer, C. Charbuillet, S. Jan, and F. Danneville, "4-Port isolated MOS modeling and extraction for mmW applications," in *ESSCIRC (ESSCIRC), 2012 Proceedings of the*, pp. 38–41, Sept 2012.

[71] S. Venugopalan, K. Dandu, S. Martin, R. Taylor, C. Cirba, X. Zhang, A. Niknejad, and C. Hu, "A non-iterative physical procedure for RF CMOS compact model extraction

using BSIM6," in *Custom Integrated Circuits Conference (CICC), 2012 IEEE*, pp. 1–4, Sept 2012.

[72] **M.-A. Chalkiadaki** and C. Enz, "Accurate RF modeling of nanoscale MOSFET using BSIM6 including low levels of inversion," *Microelectronics Journal*, vol. 45, no. 9, pp. 1159 – 1167, 2014.

[73] J. Mueller, G. Caruyer, R. Thoma, C. Bernicot, and A. Juge, "Extrinsic, Parasitic Capacitance Contributions of MOSFETs, their Impact on Performance, and their Modeling," in *Solid-State Device Research Conference, 2006. ESSDERC 2006. Proceeding of the 36th European*, pp. 323–326, Sept 2006.

[74] M. M. Arshad, V. Kilchytska, M. Emam, F. Andrieu, D. Flandre, and J.-P. Raskin, "Effect of parasitic elements on UTBB FD SOI MOSFETs RF figures of merit," *Solid-State Electronics*, vol. 97, pp. 38 – 44, 2014. Selected papers from EuroSOI 2013.

[75] S. F. Tin and K. Mayaram, "Substrate Network Modeling for CMOS RF Circuit Simulation," in *Proceedings of the IEEE Custom Integrated Circuits 1999*, pp. 583–586, 1999.

[76] D. Fleury, A. Cros, G. Bidal, J. Rosa, and G. Ghibaudo, "A New Technique to Extract the Source/Drain Series Resistance of MOSFETs," *IEEE Electron Device Letters*, vol. 30, pp. 975–977, Sept 2009.

[77] D.-W. Lin, M.-L. Cheng, S.-W. Wang, C.-C. Wu, and M.-J. Chen, "A Novel Method of MOSFET Series Resistance Extraction Featuring Constant Mobility Criteria and Mobility Universality," *IEEE Transactions on Electron Devices*, vol. 57, pp. 890–897, April 2010.

[78] A. Bracale, V. Ferlet-Cavrois, N. Fel, D. Pasquet, J. Gautier, J. Pelloie, and J. du Port de Poncharra, "A New Approach for SOI Devices Small-Signal Parameters Extraction," *Analog Integrated Circuits and Signal Processing*, vol. 25, no. 2, pp. 157–169, 2000.

[79] H. Brut, A. Juge, and G. Ghibaudo, "New approach for the extraction of gate voltage dependent series resistance and channel length reduction in CMOS transistors," in *Microelectronic Test Structures, 1997. ICMTS 1997. Proceedings. IEEE International Conference on*, pp. 188–193, Mar 1997.

[80] E. Vincent, G. Ghibaudo, G. Morin, and C. Papadas, "On the oxide thickness extraction in deep-submicron technologies," in *Microelectronic Test Structures, 1997. ICMTS 1997. Proceedings. IEEE International Conference on*, pp. 105–110, Mar 1997.

[81] S. Walstra and R.-Y. Sah, "Thin oxide thickness extrapolation from capacitance-voltage measurements," *Electron Devices, IEEE Transactions on*, vol. 44, pp. 1136–1142, Jul 1997.

[82] G. Ghibaudo, S. Bruyere, T. Devoivre, B. DeSalvo, and E. Vincent, "Improved method for the oxide thickness extraction in MOS structures with ultra-thin gate dielectrics," in

Microelectronic Test Structures, 1999. ICMTS 1999. Proceedings of the 1999 International Conference on, pp. 111–116, 1999.

[83] B. Dormieu, P. Scheer, C. Charbuillet, H. Jaouen, and F. Danneville, "Revisited RF Compact Model of Gate Resistance Suitable for High-K/Metal Gate Technology," *Electron Devices, IEEE Transactions on*, vol. 60, pp. 13–19, Jan 2013.

[84] A. Sahoo, S. Fregonese, P. Scheer, D. Celi, A. Juge, and T. Zimmer, "Pulsed radio frequency characterisation on 28 nm complementary metal-oxide semiconductor technology," *Electronics Letters*, vol. 51, no. 1, pp. 71–72, 2015.

[85] J. M. Rollett, "The Measurement of Transistor Unilateral Gain," *IEEE Transactions on Circuit Theory*, vol. 12, no. 1, pp. 91–97, 1965.

[86] **M.-A. Chalkiadaki** and C. Enz, "Low-Power RF Modeling of a 40nm CMOS Technology Using BSIM6," in *Proceedings of the 20th International Conference Mixed Design of Integrated Circuits and Systems (MIXDES) 2013*, pp. 57–62, June 2013.

[87] T. Taris, J. Begueret, and Y. Deval, "A 60μW LNA for 2.4 GHz Wireless Sensors Network Applications," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) 2011*, pp. 1–4, June 2011.

[88] A. Mangla, **M.-A. Chalkiadaki**, F. Fadhuile, T. Taris, Y. Deval, and C. Enz, "Design methodology for ultra low-power analog circuits using next generation BSIM6 MOSFET compact model," *Microelectronics Journal*, vol. 44, no. 7, pp. 570 – 575, 2013.

[89] A. van der Ziel, "Noise in Solid-State Devices and Lasers," *Proceedings of the IEEE*, vol. 58, pp. 1178–1206, Aug 1970.

[90] J. Lee and G. Bosman, "Correlation Noise Measurements and Modeling of Nanoscale MOSFETs," in *Advanced Experimental Methods For Noise Research in Nanoscale Electronic Devices* (J. Sikula and M. Levinshtein, eds.), vol. 151 of *NATO Science Series II: Mathematics, Physics and Chemistry*, pp. 153–160, Springer Netherlands, 2005.

[91] H. Hillbrand and P. Russer, "An Efficient Method for Computer Aided Noise Analysis of Linear Amplifier Networks," *IEEE Transactions on Circuits and Systems*, vol. 23, pp. 235–238, Apr 1976.

[92] V. Rizzoli and A. Lipparini, "Computer-Aided Noise Analysis of Linear Multiport Networks of Arbitrary Topology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 33, pp. 1507–1512, Dec 1985.

[93] R. A. Pucel, W. Struble, R. Hallgren, and U. L. Rohde, "A General Noise De-embedding Procedure for Packaged Two-Port Linear Active Devices," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, pp. 2013–2024, Nov 1992.

[94] C.-H. Chen, M. Deen, Y. Cheng, and M. Matloubian, "Extraction of the Induced Gate Noise, Channel Noise, and Their Correlation in Submicron MOSFETs from RF Noise Measurements," *IEEE Transactions on Electron Devices*, vol. 48, pp. 2884–2892, Dec 2001.

[95] G. Smit, A. Scholten, R. Pijper, L. Tiemeijer, R. van der Toorn, and D. Klaassen, "RF-Noise Modeling in Advanced CMOS Technologies," *IEEE Transactions on Electron Devices*, vol. 61, pp. 245–254, Feb 2014.

[96] H. Rohdin, N. Moll, C. yi Su, and G. Lee, "Interfacial Gate Resistance in Schottky-Barrier-Gate Field-Effect Transistors," *IEEE Transactions on Electron Devices*, vol. 45, pp. 2407–2416, Dec 1998.

[97] A. Antonopoulos, M. Bucher, K. Papathanasiou, N. Mavredakis, N. Makris, R. Sharma, P. Sakalas, and M. Schroter, "CMOS Small-Signal and Thermal Noise Modeling at High Frequencies," *IEEE Transactions on Electron Devices*, vol. 60, no. 11, pp. 3726–3733, 2013.

[98] A. Scholten, L. Tiemeijer, R. van Langevelde, R. Havens, A. Zegers-van Duijnhoven, and V. Venezia, "Noise Modeling for RF CMOS Circuit Simulation," *IEEE Transactions on Electron Devices*, vol. 50, pp. 618–632, March 2003.

[99] A. Roy and C. Enz, "Compact modeling of thermal noise in the MOS transistor," *Electron Devices, IEEE Transactions on*, vol. 52, pp. 611–614, April 2005.

[100] "BSIM3 Homepage."

[101] H. Chenming and C. Enz, "Free Standard SPICE Model Supported by University Partners," *GSA Forum*, vol. 19, pp. 11–12, March 2012.

[102] M. Maher and C. Mead, "A physical charge-controlled model for MOS transistors," *Advanced Research in VLSI*, 1987.

[103] "Compact Model Coalition," 2015.

[104] J.-M. Sallese, M. Bucher, F. Krummenacher, and P. Fazan, "Inversion charge linearization in MOSFET modeling and rigorous derivation of the EKV compact model," *Solid-State Electronics*, vol. 47, no. 4, pp. 677 – 683, 2003.

[105] **M.-A. Chalkiadaki**, A. Mangla, C. Enz, Y. Chauhan, M. Karim, S. Venugopalan, A. Niknejad, and C. Hu, "Evaluation of the BSIM6 compact MOSFET model's scalability in 40nm CMOS technology," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, pp. 50–53, Sept 2012.

[106] "EKV3 MOSFET Compact Model Documentation, Model Version 301.02," 2008.

[107] J. He, X. Xi, H. Wan, M. Dunga, M. Chan, and A. M. Niknejad, "BSIM5: An advanced charge-based {MOSFET} model for nanoscale {VLSI} circuit simulation," *Solid-State Electronics*, vol. 51, no. 3, pp. 433 – 444, 2007.

[108] C. McAndrew, "Validation of MOSFET model Source-Drain Symmetry," *Electron Devices, IEEE Transactions on*, vol. 53, pp. 2202–2206, Sept 2006.

[109] X. Li, W. Wu, A. Jha, G. Gildenblat, R. van Langevelde, G. Smit, A. Scholten, D. Klaassen, C. McAndrew, J. Watts, C. Olsen, G. Coram, S. Chaudhry, and J. Victory, "Benchmark Tests for MOSFET Compact Models With Application to the PSP Model," *Electron Devices, IEEE Transactions on*, vol. 56, pp. 243–251, Feb 2009.

[110] H. Agarwal, S. Venugopalan, **M.-A. Chalkiadaki**, N. Paydavosi, J. Duarte, S. Agnihotri, C. Yadav, P. Kushwaha, Y. Chauhan, C. Enz, A. Niknejad, and C. Hu, "Recent enhancements in BSIM6 bulk MOSFET model," in *Simulation of Semiconductor Processes and Devices (SISPAD), 2013 International Conference on*, pp. 53–56, Sept 2013.

[111] Y. Chauhan, S. Venugopalan, **M.-A. Chalkiadaki**, M. Karim, H. Agarwal, S. Khandelwal, N. Paydavosi, J. Duarte, C. Enz, A. Niknejad, and C. Hu, "BSIM6: Analog and RF Compact Model for Bulk MOSFET," *Electron Devices, IEEE Transactions on*, vol. 61, pp. 234–244, Feb 2014.

[112] S.-Y. Oh, D. Ward, and R. Dutton, "Transient analysis of MOS transistors," *Electron Devices, IEEE Transactions on*, vol. 27, pp. 1571–1578, Aug 1980.

[113] **M.-A. Chalkiadaki**, C. Valla, F. Poullet, and M. Bucher, "Why- and how- to integrate Verilog-A compact models in SPICE simulators," *International Journal of Circuit Theory and Applications*, vol. 41, no. 11, pp. 1203–1211, 2013.

[114] "PSP 103.1 Technical Manual," 2009.

[115] A. Bazigos, M. Bucher, J. Assenmacher, S. Decker, W. Grabinski, and Y. Papananos, "An Adjusted Constant-Current Method to Determine Saturated and Linear Mode Threshold Voltage of MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 58, no. 11, pp. 3751–3758, 2011.

[116] Y. A. El-Mansy and A. R. Boothroyd, "Authors' reply to "comments on 'a new approach to the theory and modeling of igfet's'"," *IEEE Transactions on Electron Devices*, vol. 25, pp. 393–394, March 1978.

[117] **M.-A. Chalkiadaki** and C. Enz, "Chapter 10: Parameter Extraction Procedure, in: BSIM6.1.1 MOSFET Compact Model. Technical Manual," 2015.

[118] R. Ionita and A. Vladimirescu, "Substrate-well modeling for DSM triple-well CMOS digital circuits with adjustable VT," in *Microelectronics, 2004. ICM 2004 Proceedings. The 16th International Conference on*, pp. 274–277, Dec 2004.

[119] W.-S. Feng, C.-I. Yeh, C.-M. Tsao, H.-H. Li, P. Chatterjee, and C.-H. Chen, "Low Power Design for RF Circuits," in *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC), 2011*, vol. 1, pp. 657–660, 2011.

[120] K. Okada, Y. Nomiyama, R. Murakami, and A. Matsuzawa, "A 0.114-mW Dual-Conduction Class-C CMOS VCO with 0.2-V Power Supply," in *VLSI Circuits, 2009 Symposium on*, pp. 228–229, 2009.

[121] A. Shirazi and S. Mirabbasi, "An Ultra-Low-Voltage CMOS Mixer Using Switched-Transconductance, Current-Reuse and Dynamic-Threshold-Voltage Gain-Boosting Techniques," in *New Circuits and Systems Conference (NEWCAS), 2012 IEEE 10th International*, pp. 393–396, 2012.

[122] U. Pfeiffer and E. Ojefors, "Terahertz Imaging with CMOS/BiCMOS Process Technologies," in *ESSCIRC, 2010 Proceedings of the*, pp. 52–60, 2010.

[123] D. Huang, T. LaRocca, M.-C. Chang, L. Samoska, A. Fung, R. Campbell, and M. Andrews, "Terahertz CMOS Frequency Generator Using Linear Superposition Technique," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 12, pp. 2730–2738, 2008.

[124] E. Seok, C. Cao, D. Shim, D. Arenas, D. Tanner, C.-M. Hung, and K. O, "A 410GHz CMOS Push-Push Oscillator with an On-Chip Patch Antenna," in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, pp. 472–629, 2008.

[125] D. M. Binkley, "Tradeoffs and Optimization in Analog CMOS Design," in *Mixed Design of Integrated Circuits and Systems, 2007. MIXDES '07. 14th International Conference on*, pp. 47–60, 21-23 June 2007 2007.

[126] L. Lewyn, T. Ytterdal, C. Wulff, and K. Martin, "Analog Circuit Design in Nanoscale CMOS Technologies," *Proceedings of the IEEE*, vol. 97, pp. 1687–1714, Oct 2009.

[127] W. Sansen, "Analog CMOS from 5 micrometer to 5 nanometer," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, pp. 1–6, Feb. 2015.

[128] W. Sansen, "Analog design procedures for channel lengths down to 20 nm," in *Electronics, Circuits, and Systems (ICECS), 2013 IEEE 20th International Conference on*, pp. 337–340, 8-11 Dec. 2013 2013.

[129] A. Mangla, *Modeling nanoscale quasi-ballistic MOS transistors.* PhD thesis, EPFL, Lausanne, 2014.

[130] A. Mangla, C. C. Enz, and J. M. Sallese, "Figure-of-merit for optimizing the current-efficiency of low-power RF circuits," in *Mixed Design of Integrated Circuits and Systems (MIXDES), 2011 Proceedings of the 18th International Conference*, pp. 85–89, 16-18 June 2011 2011.

[131] F. Silveira, D. Flandre, and P. Jespers, "A gm/ID based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA," *Solid-State Circuits, IEEE Journal of*, vol. 31, pp. 1314–1319, Sep 1996.

[132] P. Jespers, *The gm/ID Methodology, a sizing tool for low-voltage analog CMOS Circuits: The semi-empirical and compact model approaches.* Analog Circuits and Signal Processing, Springer, 2009.

[133] P. Jespers and B. Murmann, "Calculation of MOSFET distortion using the transconductance-to-current ratio (gm/ID)," in *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*, pp. 529–532, May 2015.

[134] C. Enz, **M.-A. Chalkiadaki**, and A. Mangla, "Low-power Analog/RF Circuit Design based on the Inversion Coefficient," in *Solid-State Circuit Conference (ESSCIRC), 2015 Proceedings of the European*, Sept 2015.

[135] C. C. Hu, *Modern Semiconductor Devices for Integrated Circuits.* Prentice Hall, 2010.

[136] T. Rudenko, M. M. Arshad, J.-P. Raskin, A. Nazarov, D. Flandre, and V. Kilchytska, "On the gm/ID-based approaches for threshold voltage extraction in advanced MOSFETs and their application to ultra-thin body SOI MOSFETs," *Solid-State Electronics*, vol. 97, pp. 52 – 58, 2014. Selected papers from EuroSOI 2013.

[137] R. Sharma, A. Antonopoulos, N. Mavredakis, and M. Bucher, "Analog/RF figures of merit of advanced DG MOSFETs," in *Devices, Circuits and Systems (ICCDCS), 2012 8th International Caribbean Conference on*, pp. 1–4, March 2012.

[138] C. Enz and **M.-A. Chalkiadaki**, "Nanoscale MOSFET Modeling for Low-power RF Design using the Inversion Coefficient," in *Asia-Pacific Microwave Conference (APCM), 2015 Proceedings of the*, Dec 2015.

[139] A. Shameli and P. Heydari, "Ultra-low power RFIC design using moderately inverted MOSFETs: an analytical/experimental study," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2006 IEEE*, p. 4 pp., 2006.

[140] V. Rizzoli and A. Lipparini, "Computer-Aided Noise Analysis of Linear Multiport Networks of Arbitrary Topology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 33, pp. 1507–1512, Dec 1985.

[141] G. Guitton, A. Mangla, **M.-A. Chalkiadaki**, F. Fadhuile, T. Taris, and C. Enz, "Design of ultra low-power RF oscillators based on the inversion coefficient methodology using BSIM6 model," *International Journal of Circuit Theory and Applications*, vol. 44, no. 2, pp. 382–397, 2016.

MARIA-ANNA CHALKIADAKI

✉ machalkiadaki@gmail.com

Key Skills – In my field of Expertise

Compact MOSFET Modeling	BSIM6 Model
Parameter Extraction Techniques	EKV & EKV3 MOSFET Model
Modeling at Radio Frequencies (RF)	Verilog-A, SPICE, ICCAP, ADS, SMASH

Education

May. 2016 **PhD in Microsystems and Microelectronics**

Title of PhD Thesis: "Characterization and Modeling of Nanoscale MOSFET for Ultra-low power RF IC Design"

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

Feb. 2011 **Master in Electronic and Computer Engineering**

(2 years of studies and research)

Title of Master Thesis: "Microwave Modeling and Parameter Extraction of MOSFETs"

Technical University of Crete (TUC), Chania, Greece.

Sep. 2008 **Diploma in Electronic and Computer Engineering (GPA: 7.94/10, very good)**

(5 years studies equivalent to master)

Title of Diploma Thesis: "Small and Large Signal Modeling of MOSFETs at High Frequencies"

Technical University of Crete (TUC), Chania, Greece.

Professional Experience

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland — **Research**

Model Developer — **Feb. 2012 to Present**

- ✓ Collaborating with the BSIM Team for the development of BSIM6 MOSFET Model
- ✓ Extraction of Model Parameters of BSIM6 MOSFET Model for Advanced 40nm & 28nm bulk CMOS Technologies
- ✓ Working on Compact & Analytical RF MOSFET Modeling for Advanced CMOS Technologies

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland — **Teaching Assistant**

Teaching Assistant at "Advanced Analog and RF IC Design I" & "Advanced Analog and RF IC Design II" Master Courses — **Sep. 2011 to Dec. 2014**

- ✓ Delivering Exercise Sessions

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland — **Internship**

Model Developer — **Sep. 2011 to Jan. 2012**

- ✓ Collaborating with the BSIM Team for the development of BSIM6 MOSFET Model

Dolphin Integration, Grenoble, France — **Internship**

Circuit Simulator Developer — **April to Aug., 2011**

- ✓ Implementation and Validation of Verilog-A Compact Models in SMASH (SPICE-like Circuit Simulator)
- ✓ EKV3 MOSFET Model and HV-MOSFET EPFL Model have been implemented

EM Microelectronic Marin-SA, Neuchatel, Switzerland — **Project Assignment**

Model Developer — **Oct. to Nov. 2010**

- ✓ Extraction of Model Parameters of EKV2.6 MOSFET Model for 180nm CMOS Technology

Technical University of Crete, Chania, Greece — **Teaching Assistant**

Teaching Assistant at "Electronics II" Undergraduate Course — **Semesters 2009 & 2010**

- ✓ Providing Assistance during Experiments in the Laboratory
- ✓ Delivering Exercise Sessions

Technical University of Crete, Chania, Greece — **Research**

Master & Undergraduate Student Researcher — **Feb. 2008 to Feb. 2011**

- ✓ Characterization of a series of Advanced CMOS Technologies down to 90nm (TOSHIBA, IBM, TSMC, ATMEL, EM)
- ✓ Work with- and Parameter Extraction of- different MOSFET Models (EKV3, PSP, EKV2.6, BSIM)

Languages

Greek (native)
English (fluent) — Certificate of Proficiency in English, University of Cambridge & University of Michigan
German (basic) — Zertifikat Deutsch, Goethe Institut
French (basic)

Programming Languages & Software

Verilog-A, Spice	IC-CAP-Agilent, ADS-Agilent, SMASH-Dolphin
C, C++, Java, MATLAB	Visual Studio-Microsoft, MATLAB-MathWorks, Mathematica, Mathcad
MySQL	MySQL Server
VHDL, Assembly (MIPS, ATMEL AVR)	MaxPlus-Altera, AVR Studio-ATMEL, Electric, Pro-Engineering-PTC

Publications

Journals:

- ✓ **M.-A. Chalkiadaki**, C.C. Enz, "Analytical RF and RF-Noise Modeling and Parameter Extraction of Nanoscale MOSFET from Weak to Strong Inversion", in *IEEE Trans. on Microwave Theory and Techniques*, vol. 63, num. 7, pp. 2173 - 2184, May 2015.
- ✓ G. Guitton, A. Mangla, **M.-A. Chalkiadaki**, F. Fadhuile, T. Taris, and C.C. Enz, "Design of Ultra Low-Power RF Oscillators based on the Inversion Coefficient Methodology using BSIM6 model", in *International Journal of Circuit Theory and Applications*, published online: Mar. 2015.
- ✓ **M.-A. Chalkiadaki**, C.C. Enz, "Accurate RF modeling of nanoscale MOSFET using BSIM6 including low levels of inversion", in *Microelectronics Journal*, vol. 45, num. 9, pp. 1159-1167, Sep. 2014.
- ✓ Y.S. Chauhan, S. Venugopalan, **M.-A. Chalkiadaki**, M.A. Karim, H. Agarwal, S. Khandelwal, N. Paydavosi, J.P. Duarte, C. Enz, A. Niknejad, C. Hu, "BSIM6: Analog and RF Compact Model for Bulk MOSFET", in *IEEE Trans. on Electron Devices*, vol. 61, no. 2, pp. 234-244, Feb. 2014.
- ✓ **M.-A. Chalkiadaki**, C. Valla, F. Poullet, M. Bucher, "Why-and how-to integrate Verilog-A compact models in SPICE simulators", in *International Journal of Circuit Theory and Applications*, vol. 41, no. 11, pp. 1203-1211, Nov. 2013.
- ✓ A. Mangla, **M-A. Chalkiadaki**, F. Fadhuile, T. Taris, Y. Deval, C.C. Enz, "Design methodology for ultra low-power analog circuits using next generation BSIM6 MOSFET compact model", in *Microelectronics Journal*, vol. 44, no. 7, pp. 570-575, July 2013.
- ✓ **M.-A. Chalkiadaki**, M. Bucher, "Large-signal RF modeling with the EKV3 MOSFET model", *Journal of Telecommunications and Information Technology (JTIT)*, vol.1, pp. 25-28, Jan. 2010.

Conferences / Workshops:

- ✓ C.C. Enz, **M.-A. Chalkiadaki**, "Nanoscale MOSFET Modeling for Low-power RF Design using the Inversion Coefficient", *Asia-Pacific Microwave Conference (APMC2015)*, Nanjing, China, Dec. 6-9, 2015.
- ✓ C.C. Enz, **M.-A. Chalkiadaki**, A. Mangla, "Low-Power Analog/RF Circuit Design Based on the Inversion Coefficient", *Proceedings at European Solid-State Circuits/Device Research Conference (ESSCIRC/ESSDERC)*, Graz, Austria, Sep. 14-18, 2015.
- ✓ **M.-A. Chalkiadaki**, C.C. Enz, "RF Characterization and Modeling of Nanoscale MOSFET from Weak to Strong Inversion", *MOS-AK Workshop at European Solid-State Circuits/Device Research Conference (ESSCIRC/ESSDERC)*, Venice Lido, Italy, Sep. 26, 2014.

- ✓ **M.-A. Chalkiadaki**, C.C. Enz, "RF Characterization and Modeling of Nanoscale MOSFET from Weak to Strong Inversion", *MOS-AK Workshop at European Solid-State Circuits/Device Research Conference (ESSCIRC/ESSDERC)*, Venice Lido, Italy, Sep. 26, 2014.
- ✓ H. Agarwal, S. Venugopalan, **M.-A. Chalkiadaki**, N. Paydavosi, J.P. Duarte, S. Agnihotri, C. Yadav, P. Kushwaha, Y.S. Chauhan, C.C Enz, A. Niknejad, C. Hu, "Recent Enhancements in BSIM6 Bulk MOSFET Model", *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Glasgow, Scotland, Sep. 3-5, 2013.
- ✓ **M.-A. Chalkiadaki**, C.C. Enz, "Low-Power RF Modeling of a 40nm CMOS Technology Using BSIM6", *20th International Conference Mixed Design of Integrated Circuits and Systems (MIXDES)*, Gdynia, Poland, June 20-22, 2013
- ✓ Y.S. Chauhan, **M.-A. Chalkiadaki**, S. Venugopalan, M.A. Karim, N. Paydavosi, S. Jandhyala, J.P. Duarte, C.C. Enz, A. Niknejad, C. Hu, "Global Geometrical Scaling in BSIM6", *MOS-AK Workshop*, San Francisco, USA, Dec. 12, 2012.
- ✓ **M.-A. Chalkiadaki**, A. Mangla, C.C. Enz, Y.S. Chauhan, M.A. Karim, S. Venugopalan, A. Niknejad, C. Hu, "Evaluation of the BSIM6 compact MOSFET model's scalability in 40nm CMOS technology", *Proceedings at European Solid-State Circuits/Device Research Conference (ESSCIRC/ESSDERC)*, Bordeaux, France, Sep. 17-21, 2012.
- ✓ C.C. Enz, A. Mangla, **M.-A. Chalkiadaki**, "Design Methodology for Ultra Low-Power Analog Circuits using Next Generation BSIM6 Compact Model", *Workshop on Compact Modeling (WCM) at the Nanotech 2012*, Santa Clara, California, USA, 18-21 June, 2012.
- ✓ Y.S. Chauhan, M. Karim, S. Venugopalan, P. Thakur, N. Paydavosi, A. Sachid, A. Niknejad, C. Hu, W. Wu, K. Dandu, K. Green, T. Krakowski, G. Coram, S. Cherepko, S. Sirohi, A. Dutta, R. Williams, J. Watts, **M.-A. Chalkiadaki**, A. Mangla, W. Grabinski , C.C. Enz, "Transitioning from BSIM4 to BSIM6", *International Workshop on Device Modeling for Microsystems*, Noida, India, Mar. 17, 2012.
- ✓ Y. S. Chauhan, M. A. Karim, S. Venugopalan, A. Sachid, A. Niknejad, C. Hu, W. Wu, K. Dandu, K. Green, G. Coram, S. Cherepko, J. Wang, S. Sirohi, J. Watts, **M.-A. Chalkiadaki**, A. Mangla, A. Bazigos, F. Krummenacher, W. Grabinski, C. Enz, "BSIM6: Symmetric Bulk MOSFET Model", *The Nano-Tera Workshop on the next generation MOSFET Compact Models*, Lausanne, Switzerland, Dec. 15-16, 2011.
- ✓ **M.-A. Chalkiadaki**, M. Bucher, A. Bazigos, "RF Parameter Extraction in 90 nm CMOS with EKV3 Compact Model", *4th Int. Conf. on Micro-Nanoelectronics, Nanotechnologies & MEMS (Micro&Nano 2010)*, Athens, Greece, Dec. 12-15, 2010.
- ✓ **M.-A. Chalkiadaki**, M. Bucher, A. Bazigos, "RF Modeling and Parameter Extraction with the EKV3 MOSFET Model", *8th Graduate Student Meeting on Electronic Engineering*, Tarragona, Spain, June 28-29, 2010.
- ✓ M. Bucher, **M.-A. Chalkiadaki**, A. Bazigos, "The EKV3 MOS Transistor Model for RF Circuit Simulation", *16th Int. Conf. Mixed Design of Integrated Circuits and Systems (MIXDES 2009)*, Lodz, Poland, June 25-27, 2009.
- ✓ **M.-A. Chalkiadaki**, M. Bucher, "Large Signal RF Modelling with the EKV3 MOSFET model", *8th Diagnostics & Yield Symposium*, Warsaw, Poland, June 22-24, 2009.
- ✓ M. Bucher, A. Bazigos, **M.-A. Chalkiadaki**, "Aspects of High-Frequency Modelling of MOSFETs with EKV3", *MOS-AK Workshop at European Solid-State Circuits/Device Research Conference (ESSCIRC/ESSDERC)*, Edinburgh, Scotland, Sep. 19, 2008.

Technical Manuals:

- ✓ **M.-A. Chalkiadaki**, C. C. Enz, "Parameter Extraction Procedure" *in BSIM6.1.1 MOSFET Compact Model Technical Manual*, pp. 81-99, 2014. Online: http://www-device.eecs.berkeley.edu/bsim/?page=BSIM6.

Training Courses

- ✓ "Teaching Toolkit", Lausanne, Switzerland, September 18, 2014.
- ✓ "Management of Innovation and Technology Transfer", Lausanne, Switzerland, June 03-04 & 11-12, 2014.
- ✓ "Nanoscale MOS transistors: Semi-classical modeling and applications", Udine, Italy, May 20-24, 2013.
- ✓ "Training Courses on Compact Modeling (TCCM 2010)", Tarragona, Spain, June 30, 2010 - July 1, 2010
- ✓ "Disruptive Technologies for More Moore", Athens, Greece, September 14, 2009.
- ✓ IDESA Training Course, "Advanced RF Implementation flow", Warsaw, Poland, June 15-19, 2009.

Activities and Interests

Scouting	Volunteer Scout Leader from 2001-2010
Sports	Hiking, Sea Sports (Sailing License, Speedboat License), Kick Boxing
Music	Guitar Playing