NOISE CHARACTERIZATION AND MODELING OF NANOSCALE MOSFETS

NOISE CHARACTERIZATION AND MODELING OF NANOSCALE MOSFETS

By

Xuesong Chen, B.Sc.

A Thesis Submitted to the School of Graduate Studies in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

McMaster University

Hamilton, Ontario, Canada

Doctor of Philosophy (2017) McMaster University

(Electrical and Computer Engineering) Hamilton, Ontario

TITLE: Noise Characterization and Modeling of Nanoscale MOSFETs

AUTHOR: Xuesong Chen, B.Sc.,

University of Science and Technology of China, Hefei, China

SUPERVISOR: Prof. Chih-Hung (James) Chen

NUMBER OF PAGES: xvii, 155

Abstract

High-frequency noise modeling and characterization of nanoscale MOSFETs are essential driving forces for highly scaled CMOS technology to be used in radio-frequency applications. Continuous downscaling increases the operating frequency of the MOSFETs, reduces the power supply voltage but does not scale noise accordingly. This makes the noise issue of future low-power technology more prominent and therefore accurate noise modeling more important.

In this thesis, several important issues regarding noise modeling and characterization for nanoscale MOSFETs are studied. First, a new noise factor deembedding algorithm is proposed for on-wafer noise measurements. It solves the problem of noise factor deembedding in which the active two-port device is surrounded by a four-port parasitic network. Based on it, a new deembedding-first and optimization-last noise parameter deembedding approach is proposed and its performance is evaluated using experimental data.

Second, the noise performance of modern sub-100-nm MOSFETs are evaluated using the noise sheet resistance as a figure of merit. It shows that future technologies generally have degraded noise performance. In addition, two accuracy issues regarding the calibration of noise receiver for high-frequency noise measurements are investigated and methods to mitigate these issues are discussed.

Third, a novel *Z*-parameter based approach to extract the gate resistance is proposed for MOSFET characterization. It is evaluated against other published methods using experimental data. In addition, the extraction of the resistance of the lightly-doped-drain region and the gate contact is also performed and discussed.

Finally, a new perspective to interpret the MSOFET channel noise as suppressed shot noise is presented. An easy-to-use analytical expression for the suppression factor is derived and it only relies on two process parameters – threshold voltage and effective oxide thickness – to predict the level of suppression for the channel noise of MOSFETs. It is evaluated using published experimental data on various CMOS technology nodes.

Acknowledgements

First, I would like to express my sincere gratitude and deep appreciation to my supervisor, Professor Chih-Hung (James) Chen, for his support and guidance over the past five years of my pursuit of the Ph.D. degree. His thoughtful and visionary advices have enlightened me in both my quest in scientific research and my journey in life. His encouragement has helped me overcome my own frustrations and work through many hard problems. His rigorous attitude and enthusiastic passion for research has influenced and motivated me to become an ever better researcher.

I would also like to sincerely thank my committee members – Professor M. Jamal Deen and Professor Yaser M. Haddara – for their valuable and insightful advices and comments on my research. Discussions with them have greatly helped advancing my research to a higher level and better quality.

My thanks also go to my colleagues in the ECE department – Dr. Zhiyun Li, Dr. Zeng Cheng, Xiaoqing Zhen, and Dr. Yiheng Qin – who are all outstanding researchers and working together with them gives me a great pleasure and memorable experience.

I would like to express my gratitude to the Natural Sciences and Engineering Research Council of Canada (NSERC), as well as the Government of Ontario, Dr. Chandra and Mrs. Wendy Kudsia through the Ontario Graduate Scholarship (OGS), for funding my research. I am also grateful to Focus Microwaves Inc. for research funding through the NSERC

Engage (PLUS) Grant. I would like to thank United Microelectronics Corporation (UMC) and Focus Microwaves Inc. for providing experimental data for my research.

Last but not the least, I sincerely thank my family for their unconditional support and unceasing encouragement during my long years of academic pursuit.

Table of Contents

List of Figures

List of Abbreviations

BJT	Bipolar Junction Transistor
CLM	Channel Length Modulation
CMOS	Complementary Metal-Oxide-Semiconductor
DUT	Device-Under-Test
DIBL	Drain-Induced Barrier Lowering
ENR	Excess Noise Ratio
FET	Field Effect Transistor
FoM	Figure of Merit
FinFET	Fin Field Effect Transistor
HVT	High Threshold Voltage
JLT	Junction-Less Transistor
LDD	Lightly-Doped Drain
LNA	Low Noise Amplifier
LVT	Low Threshold Voltage
MOSFET	Metal-Oxide-Semiconductor Field Effect Transistor
NF	Noise Figure
NQS	Non-Quasi-Static
NFET	N-type Field Effect Transistor
PSD	Power Spectral Density
PFET	P-type Field Effect Transistor
QW	Quantum Well
RF	Radio-Frequency

RFIC	Radio-Frequency Integrated Circuits
RVT	Regular Threshold Voltage
SNR	Signal-to-Noise Ratio
SOI	Silicon-on-Insulator
UMC	United Microelectronics Cooperation

Chapter 1.

Introduction

1.1 Electrical Noise

Noise, in this work, refers to the spontaneous fluctuations that appear in currents and voltages of the system under consideration. It originates from the dynamic nature of the microscopic world where elementary particles keep moving randomly. Noise is a crucial problem in electrical engineering as it fundamentally limits the signal transmitting capability of the system, i.e., the signal power has to be strong enough to be distinguishable from the noise. Consequently, research of the noise properties of the electrical system is important to study its limits and also to possibly improve the system's performance.

The macroscopic noise in currents and voltages generally follows Gaussian distribution centered at the mean value of the current or voltage [1], which does not reveal much information of the noise itself. To study the behavior of the noise, Fourier analysis can be used to transform it to the frequency domain. The most commonly used physical quantity to describe the noise in the frequency domain is the power spectral density (PSD), i.e., the power of noise signal in one Hertz. For voltage fluctuations, the noise PSD has a unit of V^2/Hz, and for current fluctuations, it has A^2/Hz.

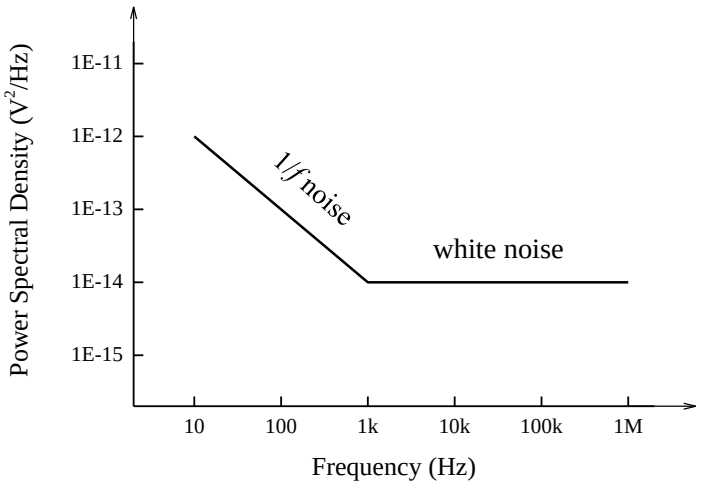


Fig. 1.1. Noise spectrum when both white noise and 1/*f* noise are present.

According to the noise spectrum's dependence on frequency, noise can be categorized as white noise, which has a constant value up to extremely high frequency where quantum correction needs to be considered; and the flicker noise, which has a 1/*f*, or "pink", dependence on the frequency, thus the other name 1/*f* noise [1]. For the latter, due to its inverse proportionality with respect to the frequency, it dominates at low frequency but decreases quickly when the frequency increases. Fig. 1.1 shows an example noise spectrum when both white noise and 1/*f* noise are present, in which the white noise dominates after a corner frequency of 1 kHz.

Thermal noise and shot noise are the major white noise sources in semiconductor devices [1]. The thermal noise comes from the random thermal motion of carriers, which shares similar properties with Brownian motion that is of much larger scale. Shot noise

occurs when the carriers travel in a random but independent way so that their arrival time follows Poissonian distribution, e.g., the current flow in a Schottky-barrier diode.

1.2 Future Nanoscale MOSFETs

The metal-oxide-semiconductor field-effect transistor (MOSFET) has been the most dominant fundamental building block of modern integrated circuits due to its low cost of production and ease of integration. Driven by the economic demand, the semiconductor industry over the past thirty years has been following the Moore's law [2], which is an empirical observation that states the number of transistors per chip doubles every 2 to 3 years when a new technology node is introduced, as shown in Fig. 1.2. The trend has slowed down in recent years when the scaling has started to hit the limits caused by physics as well as economics. For example, the scaling may finally run up against the limits of optical lithography. Although it might be overcome due to technological advance, the cost could be prohibitive [3],[4].

The scaling of the MOSFETs has been the main engine of driving the improvement of device performance. With reduced channel length, not only can the chip hold more transistors per unit area, the individual transistor can also perform faster due to the reduced travel time for the carriers in the channel. In modern technology nodes, this results in very high cutoff frequency (f_T) and maximum oscillation frequency (f_{max}) in the hundreds of GHz. According to 2012 International Technology Roadmap for Semiconductors [6], multi-gate MOSFETs with 15.3-nm gate length can achieve 710 GHz and 622 GHz for f_T

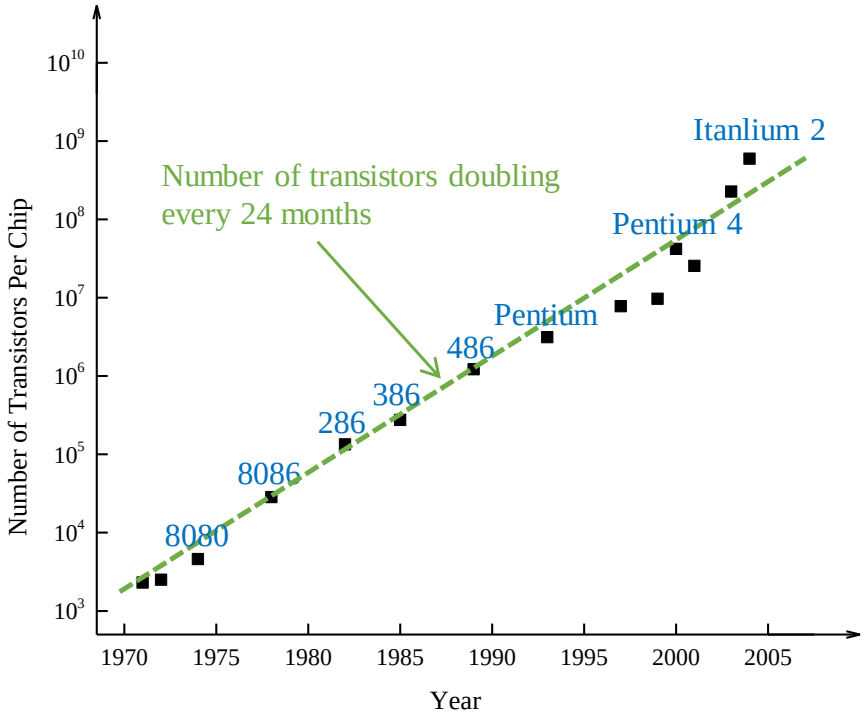


Fig. 1.2. Moore's law in action: number of transistors on an integrated circuit vs. year [5].

and f_{max}, respectively. This opens up the possibility for nanoscale CMOS technology as a potential candidate for sub-THz applications [7].

However, the scaling does not come without a price. The long-channel theory starts to deviate from the actual device performance as the channel length enters the sub-μm regime. As the channel length shrinks further, various short channel effects are introduced in the modeling of the transistors. The most prominent ones include the channel length modulation (CLM) effect, the mobility degradation effect due to the high field, and the drain-induced barrier lowering (DIBL) effect, most of which have deleterious effects on the transistor performance [8]. For example, the CLM effect causes the reduction of the

channel resistance, which contributes to the decreasing intrinsic gain of shorter-channel devices. Researchers have made great progress in finding new device designs that can alleviate, if not overcome, those effects. Some of the new designs, which are still based on the MOSFET structure, are already highly successful, like FinFET [9] and the Silicon-on-Insulator (SOI) MOSFET [10]. There are also silicon-based designs that are still in a research and development stage, like the junction-less transistor (JLT) using silicon nanowires [11]. Alternative materials that have significantly higher carrier mobility have also been considered as the channel material, e.g., the 2-dimensional MoS_2 sheets [12] and 1-dimensional carbon nanotubes [13].

For very short-channel MOSFETs, as well as these novel transistors, recent modeling research has replaced classical drift-diffusion transport with ballistic transport [14],[15]. When the transistor channel is so short and the carriers move so fast, there is not enough time for carriers to reach thermal equilibrium before reaching the drain terminal. This undermines the thermal equilibrium assumption which is the foundation of the traditional drift-diffusion model. The transition from drift-diffusion transport to ballistic transport marks a major shift, although still ongoing, in understanding how modern transistors work.

1.3 High-Frequency Noise Modeling for MOSFETs

The understanding and modeling of high-frequency noise sources in MOSFETs are essential for the CMOS technology to be used in radio-frequency (RF) applications. Research in high-frequency noise modeling of MOSFET has been a driving force to address the signal-to-noise ratio dilemma in deep sub-micron MOSFETs for radio-

frequency integrated circuits [16],[17]. The continuous downscaling enables the MOSFETs to operate at faster speed and smaller power supply voltage, but somehow does not improve, or even degrades, their noise performance. This will cause the noise issue to be more prominent and noise modeling more important in the future of RF microelectronics.

Noise modeling for MOSFETs usually lags behind the DC and AC modeling because of the fact that it has to be based on the latter, as well as the complexity of noise measurements and noise source extraction. Based on the high-frequency small-signal equivalent circuit of a MOSFET, e.g., Fig. 1 in [18], the accompanying noise sources include the thermal noise generated by the linear resistive elements, including the gate resistance R_g, the series resistance at the source and the drain, and the resistive elements in the substrate network. They also include the two noise sources generated by the internal part of the device, i.e., the channel noise that appears in the drain current and the induced gate noise. The induced gate noise is caused by coupling of the channel noise through the gate oxide capacitance to the gate terminal when the MOSFET is working at high frequencies and is therefore usually correlated to the channel noise and dependent on the frequency. As the channel lengths of the MOSFETs scale down, the coupling becomes weaker and thus the induced gate noise less important at the same frequency. Among all these noise sources, the channel noise is the most dominant and complicated one [19] that attracts most attention from researchers. In this work, we will focus on the modeling of the channel noise of MOSFETs.

The thermal noise current generated by a linear resistor is proportional to its conductance according to the Nyquist's formula [1]. However, it cannot be directly applied to the drain current of MOSFETs due to the non-linear current-voltage characteristics. Nonetheless, initial noise modeling of MOSFETs still started from the Nyquist's formula, but by applying it on the microscopic resistive segment of the channel. An integration of the noise contribution from all the channel segments can then be taken to find the overall drain current noise. For long-channel MOSFETs working in strong inversion and saturation region, the drain current noise is derived to be 2/3 of the thermal noise level generated by the channel resistance when the drain-to-source voltage is 0 V and the channel behaves like a linear resistor [1].

However, the noise modeling of MOSFETs is not independent from the previously mentioned short-channel effects. Frequently challenged by experimental data showing higher noise level than the theoretical prediction, noise modeling has gradually started to take into account those short-channel effects, most of which increase the channel noise level. Still, recent experimental data suggested that those short-channel effects may not be enough to explain the "excess noise" [20] for sub-100-nm MOSFETs. The cause is arguably the onset of non-equilibrium transport, which essentially indicates that the noise is no longer of thermal origin.

1.4 Research Contributions

Based on the discussions in the previous sections, we address several important issues on noise characterization and modeling of nanoscale MOSFETs in this thesis. The major contributions of this work can be summarized as follows.

- A noise factor deembedding algorithm that can work with on-wafer device-under-tests (DUTs) which are surrounded by four-port parasitic networks is presented. The algorithm extends existing noise factor deembedding theories to work in situations when there are feedback paths between the output and input ports of the DUT. Based on the algorithm, a new noise parameter deembedding approach that performs deembedding first and optimization last is presented, together with statistical evaluations in comparison to the traditional approach. The proposed approach outperforms the traditional approach by being statistically more robust and producing less non-physical results. This work is published in [21].
- The trend of noise performance for future technologies, including planar MOSFETs and III-V quantum well FETs, is presented using the noise sheet resistance as a figure of merit. The impact of different process engineering techniques on noise performance is predicted. In addition, two accuracy issues on noise receiver calibration for high-frequency noise measurement are studied and methods to mitigate them are provided. This work is published in [22] and [23].
- An improved *Z*-parameter based approach to extract gate resistance at low frequencies is presented, as well as statistical evaluation with comparison to other

Y-parameter based approaches. Compared to other published methods, the proposed approach demonstrates statistical robustness and stability which are particularly important for future technology nodes with smaller feature size and larger process variations. A method to determine the gate contact resistance is also proposed. This work is published in [24].

- A consistent interpretation of the channel noise of MOSFETs as suppressed shot noise is proposed for both long- and short-channel devices. An easy-to-use analytical equation for the suppression factor is presented for MOSFETs working in strong inversion and saturation. It can predict the level of shot noise suppression for MOSFETs with only two process parameters – threshold voltage and effective oxide thickness, which can be easily obtained from DC modeling and experiments. This work is published in [25].

1.5 Thesis Organization

This thesis is organized as follows. After the general introduction in Chapter 1, we first introduce the noise theory for linear networks in the beginning of Chapter 2. We then present a noise factor deembedding algorithm that can deal with an active two-port device surrounded by a four-port parasitic network. It solves the problem of noise factor deembedding when there are feedback paths between the output and input of a device-under-test. The algorithm also leads to a new approach to obtain intrinsic noise parameters for on-wafer noise measurements by performing deembedding first and optimization last. Experimental data of n-type MOSFETs in UMC's 28- and 90-nm CMOS technologies are

used to evaluate the effect of the noise factor deembedding, as well as the optimization-last noise parameter deembedding approach.

In Chapter 3, we present the future trends in noise performance and challenges in noise characterization of semiconductor devices. A detailed evaluation of noise performance for future nanoscale technologies is carried out using the figure of merit – noise sheet resistance. Experimental data are based on devices fabricated using various modern nanoscale technologies, including UMC's 65 nm, 40 nm and 28 nm CMOS technology and technologies using III-V materials and quantum-well structures from IBM, MIT, etc. The trend for their noise performance is then discussed and recommendations are given for improving noise performance. In the second part, the calibration of noise receiver is studied on the receiver gain variations caused by the impedance difference of the noise source in "hot" and "cold" states, as well as the impedance mismatch between the noise source and the tuner. Possible solutions to mitigating the problems are also evaluated and discussed.

In Chapter 4, we present an improved *Z*-parameter based approach to extract the gate resistance at low frequencies. The effectiveness of this approach, compared with other Y-parameter based approaches, is verified using 430 samples fabricated in 40-, 55-, 90-, and 110-nm CMOS technology nodes. The extraction of the channel resistance at zero drain-source bias and the resistance of the lightly-doped-drain (LDD) region is performed. The dependence of the extracted gate resistance on the channel length and process of devices is analyzed and discussed. A method to determine the gate contact resistance is also demonstrated.

In Chapter 5, we present a consistent interpretation of the channel noise of MOSFETs as suppressed shot noise for both long- and short-channel devices. An easy-to-use analytical equation for the shot noise suppression factor of MOSFETs working in the saturation region is derived. The expression only relies on two process parameters - threshold voltage and effective oxide thickness, to predict the level of suppression for devices in CMOS technology. Together with the shot noise limit, the suppression factor can accurately predict MOSFET channel noise. Moreover, the modeled suppression factors are compared with experimental values extracted from published experimental data in 180-nm and 40-nm technologies.

Finally, in Chapter 6, the thesis is concluded with a summary of the research. Discussions for future work that can further this study are also presented.

Chapter 2.

Noise Factor Deembedding

2.1 Introduction

Besides the device-under-test (DUT), noise measurement systems usually consist of many other noisy components such as low-noise amplifiers (LNA), tuners, etc., as well as unwanted parasitics that may appear at high frequencies. The raw noise factor measured by the noise receiver is contributed by all of the noisy components. In order to obtain the noise factor of the DUT, deembedding needs to be performed to remove those contributions from the raw reading. The deembedding of noise factors for a cascade of two-ports is governed by the well-known Friis' equation [1], which is further extended to work with three-ports, such as baluns, by Abidi *et al.* [26]. However, they are not suitable for on-wafer noise measurements where the intrinsic device is surrounded by parasitics that may cause feedback paths between the output and the input ports. The parasitics, which come from the probe pads and metal interconnections, have a significant impact on high-frequency noise measurements [27]. In this work, we model the parasitics as a general passive four-port network following [28], which serves the same role as the package in the problem formulated for packaged devices [29].

The existing deembedding theories developed in [29], and more generally in [30], do not allow us to deembed a single noise factor. They only work with full knowledge of the four noise parameters of the DUT, namely the minimum noise figure NF_{min}, the equivalent noise resistance R_n, and the magnitude and phase of the optimum source reflection coefficient Γ_{opt} [31]. The four noise parameters, which determine the noise factor of the DUT at any given source admittance, can be computed from at least four noise factor measurements. However, due to the large uncertainties in noise measurements, usually 20 ~ 30 noise factors are measured in practice. To achieve better results, various optimization algorithms [32]-[42] have been developed for the computation of the noise parameters from measured noise factors, most of which are based on the least-squares fitting technique. The deembedding of noise parameters also requires the admittance (Y) parameters of both the two-port DUT and the four-port parasitics. The admittance parameters of the DUT can be computed from the measured S-parameters whereas those of the parasitics from the measured S-parameters of properly designed dummy structures, either without an equivalent circuit model [28], or modeled as parallel-series [43]-[45] or cascade [46]-[48] networks. In summary, existing approaches for on-wafer noise parameter deembedding [45]-[48] invariably require to invoke the optimization procedure before the deembedding procedure.

In this section, we first review the noise theory of linear two-port networks, which serves as the foundation for the following derivation of the noise factor deembedding. The proposed noise factor deembedding algorithm extends the noise theory of linear networks to enable direct deembedding of noise factors for an active two-port surrounded by a

passive four-port. This algorithm also leads to a new approach to obtain the intrinsic noise parameters for on-wafer noise measurements by performing deembedding first and optimization last. Following the theoretical derivations, we present the verification of the noise factor deembedding algorithm using idealized data. The optimization-last approach is then evaluated using devices fabricated in 90-nm and 28-nm CMOS technology from United Microelectronics Corporation (UMC).

2.2 Noise Theory of Linear Two-Port Networks

In this section, we present the fundamental noise theory of linear two-port networks that has already been established in the literature. It serves as the basis for the subsequent noise deembedding theories, and it is presented here for the completeness of the derivation and the ease to follow through.

2.2.1 Noise Factor and Noise Parameters

A linear two-port network with internal noise sources, no matter how many and how complicated, can be simplified as a noise-free two-port network with two explicit noise current sources, as shown in Fig. 2.1 (a) [49]. To calculate the noise parameters of the two-port in Fig. 2.1 (a), we first convert it to Fig. 2.1 (b) with one noise voltage source and one noise current source. With the help of the admittance (Y) parameters of the noise-free two-port, the conversion of the two noise representations can be written as [49]

$$u = -\frac{1}{Y_{21}} i_2 \tag{2.1}$$

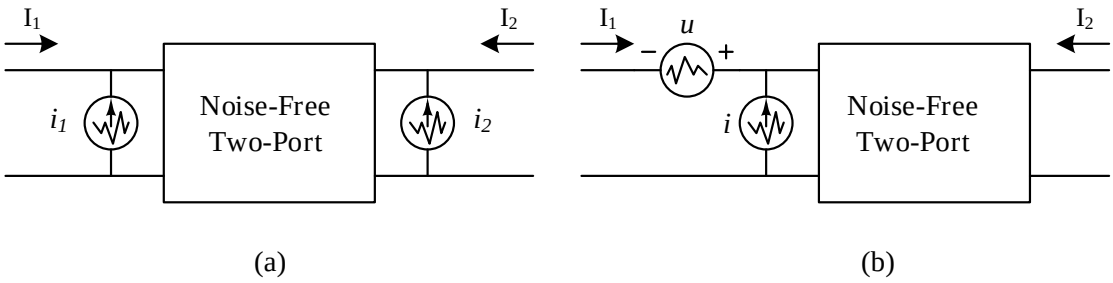


Fig. 2.1. Two representations of a noisy linear two-port network: (a) a noise-free two-port with two noise current sources i_1 and i_2; (b) a noise-free two-port with a noise voltage source u and a noise current source i [49].

and

$$i = i_1 - \frac{Y_{11}}{Y_{21}} i_2 . \tag{2.2}$$

The noise sources i and u can be correlated. We assume partial correlation and denote the uncorrelated part of i as i_{un} [49], i.e.,

$$i = i_{un} + uY_{cor} \tag{2.3}$$

and

$$\overline{i \cdot u^*} = Y_{cor} \overline{u^2} \tag{2.4}$$

where * means complex conjugate and the bar means average over time. Power spectral density of the noise sources, which are all thermal, can be expressed as the noise generated by an equivalent resistance or conductance, i.e.,

$$\overline{u^2} = 4kT_0 \Delta f \cdot R_u , \tag{2.5}$$

$$\overline{i^2} = 4kT_0 \Delta f \cdot G_i , \tag{2.6}$$

and

$$\overline{i_{un}^2} = 4kT_0 \Delta f \cdot G_{un} . \tag{2.7}$$

Based on (2.3) and (2.4), we have the following relation between the them,

$$G_i = G_{un} + |Y_{cor}|^2 R_u. \tag{2.8}$$

When the noisy two-port is connected to a signal source, it will add extra noise to the signal and thus degrade the signal-to-noise ratio (SNR). The noise factor is a measure for the degradation of the SNR when the signal is processed by a noisy component. By definition, the noise factor is the ratio of the total available noise power per unit bandwidth at the output port and its portion due to the noise at the input port at the standard temperature T_0 = 290 K [31], i.e.,

$$F = \left. \frac{N_{o,av}}{G_{av} N_i} \right|_{T_0 = 290\,\mathrm{K}} \tag{2.9}$$

where G_{av} is the available gain of the network, $N_{o,av}$ is the available noise power at the output port, and N_i is the noise power generated by the source. In engineering practice, the logarithmic scale of noise factor, which is called noise figure (NF), is also widely used, i.e.,

$$NF = 10 \cdot \log(F). \tag{2.10}$$

The theoretical lower limit of noise factor is 1, which corresponds to noise figure of 0 dB. In this work, noise factor is used in all sections for clarity.

Based on the definition in (2.9), to calculate the noise factor of a noisy two-port, we need to provide an input noise source. Fig. 2.2 shows a noisy two-port network with an input noise source that has an admittance Y_S and an equivalent noise current i_S [31]. Since the noise-free two-port does not alter the noise factor and the equivalent noise sources

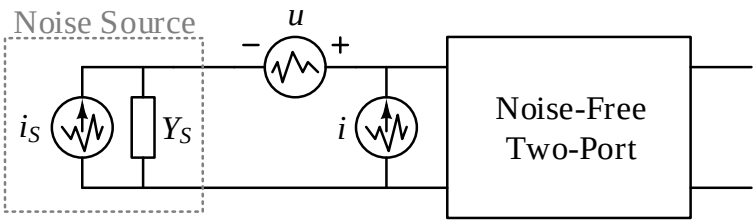


Fig. 2.2. A noisy two-port network with an input noise source that has an admittance of Y_S and equivalent noise current i_S [31]. It is used to calculate the noise factor, as well as to derive the noise parameters.

appear at its input, the noise factor can be expressed as the ratio of the total noise power to the noise power of the input source only, i.e.,

$$\begin{aligned} F &= \frac{\overline{i_S^2} + \overline{|\, i + Y_S u \,|^2}}{\overline{i_S^2}} \\ &= 1 + \frac{\overline{|\, i_{un} + (Y_{cor} + Y_S) u \,|^2}}{\overline{i_S^2}} \\ &= 1 + \frac{\overline{i_{un}^2}}{\overline{i_S^2}} + \frac{|\, Y_{cor} + Y_S \,|^2 \, \overline{u^2}}{\overline{i_S^2}} \\ &= 1 + \frac{G_{un}}{G_S} + \frac{|\, Y_{cor} + Y_S \,|^2 \, R_u}{G_S} \end{aligned} \tag{2.11}$$

where G_S is the real part of Y_S. Substituting (2.8) into (2.11), we can simplify the latter as

$$\begin{aligned} F &= 1 + \frac{G_i - |\, Y_{cor} \,|^2 \, R_u}{G_S} + \frac{|\, Y_{cor} + Y_S \,|^2 \, R_u}{G_S} \\ &= 1 + \frac{G_i + R_u \left[(G_S + G_{cor})^2 + (B_S + B_{cor})^2 - ({G_{cor}}^2 + {B_{cor}}^2) \right]}{G_S}. \end{aligned} \tag{2.12}$$

To achieve the minimum noise factor F_{min}, we can use the derivatives of F with respect to B_S and G_S and make them equal zero. The resulting optimum values for B_S and G_S are called B_{opt} and G_{opt} [31], which are given by

$$B_{opt} = -B_{cor} \tag{2.13}$$

and

$$G_{opt} = \sqrt{\frac{G_i}{R_u} - B_{cor}{}^2}. \tag{2.14}$$

The minimum noise factor F_{min} happens when $B_S = B_{opt}$ and $G_S = G_{opt}$, which is given by

$$F_{\min} = 1 + 2R_u(G_{cor} + G_{opt}). \tag{2.15}$$

Based on F_{min}, the noise factor F can then be expressed as a function of the source admittance Y_S, i.e.,

$$F = F_{\min} + \frac{R_n}{G_S}\left|Y_S - Y_{opt}\right|^2 \tag{2.16}$$

where R_n is the same as R_u and is called the equivalent noise resistance. Here the noise factor F is contributed by two terms with the first being the minimum achievable noise factor $F_{\min}$ when $Y_S = Y_{opt}$ and the second being the increase of noise factor due to mismatch between the two admittances. In (2.16), there are four coefficients by counting the complex admittance Y_{opt} as two, namely the minimum noise factor F_{min}, the equivalent resistance R_n, and the real and imaginary parts of the optimum source admittance Y_{opt}. The four noise parameters can fully describe the noise properties of the two-port network based on (2.16) which gives the noise factor at any given source admittance.

Another widely used noise factor expression is in terms of the reflection coefficient Γ,

$$F = F_{\min} + \frac{4R_n}{Z_0} \frac{\left|\Gamma_{opt} - \Gamma_S\right|^2}{\left|1+\Gamma_{opt}\right|^2 \left(1-\left|\Gamma_S\right|^2\right)} \tag{2.17}$$

where Γ_S is the source reflection coefficient and Γ_{opt} is the optimum reflection coefficient. The reflection coefficients are more intuitive with the help of Smith Chart and can be converted to admittance values using

$$Y = Y_0 \cdot \frac{1-\Gamma}{1+\Gamma} \tag{2.18}$$

and reversely using

$$\Gamma = \frac{Y_0 - Y}{Y_0 + Y} \tag{2.19}$$

where Y_0 is the characteristic admittance.

2.2.2 Computation of Noise Parameters

In practice, the noise parameters are not directly measured but computed from a number of measured noise factors. From (2.16), we can view the four noise parameters as the coefficients and the source admittance Y_S as the variable. Theoretically, with four noise factor measurements at different Y_S values, we will be able to solve for the four coefficients. However, due to the large uncertainties in noise measurements, usually 20 to 30 noise factors are measured to over-determine the coefficients. To achieve better results for the noise parameters, various optimization algorithms have been developed for their computation [32]-[42], most of which are based on the least-squares fitting technique. Due to the statistical nature of the noise parameter computation procedure, it is also broadly called the optimization procedure.

One of the earliest noise parameter calculation method comes from Lane [32] and remains effective. In Lane's method, the noise factor expression (2.16) is cast in a form that is linear with respect to four new parameters *A*, *B*, *C*, and *D*, i.e.,

$$F = A + BG_S + \frac{C + BG_S^2 + DB_S}{G_S} \tag{2.20}$$

where the four new parameters are related to the noise parameters by

$$F_{\min} = A + \sqrt{4BC - D^2}\,, \tag{2.21}$$

$$R_n = B\,, \tag{2.22}$$

$$G_{opt} = \frac{\sqrt{4BC - D^2}}{2B}\,, \tag{2.23}$$

and

$$B_{opt} = \frac{-D}{2B}\,. \tag{2.24}$$

By applying least-squares fitting algorithm on the linear equation (2.20) with at least four noise factors measured at their corresponding source admittances, we can find the optimum values for the parameters *A*, *B*, *C*, and *D*, and thus the four noise parameters. Thereafter, many research efforts [33]-[42] have been contributed to improve the accuracy and robustness of the noise parameter computation procedure, yet it remains a hard problem when the measurement errors are large. In this work, Lane's method is used for the noise parameter computation for its simplicity and effectiveness on the given data sets.

2.2.3 Noise Correlation Matrices

The noise property of a noisy two-port network can be fully described by the previously introduced four noise parameters. Alternatively, for the same purpose, we can also utilize

the noise correlation matrices which provide a matrix representation of the noise properties of the network. Their definitions and related derivations are presented as follows.

Based on the two representations of noisy linear two-port in Fig. 2.1, we can define the normalized noise correlation matrix in admittance form $\boldsymbol{C}_Y$ and chain form $\boldsymbol{C}_A$ as [30],[51]

$$\boldsymbol{C}_Y = \frac{1}{4kT\Delta f} \cdot \begin{bmatrix} \overline{i_1^2} & \overline{i_1 i_2^*} \\ \overline{i_1^* i_2} & \overline{i_2^2} \end{bmatrix} \tag{2.25}$$

and

$$\boldsymbol{C}_A = \frac{1}{4kT\Delta f} \cdot \begin{bmatrix} \overline{u^2} & \overline{ui^*} \\ \overline{u^* i} & \overline{i^2} \end{bmatrix}. \tag{2.26}$$

The two matrices can be converted to each other with the help of the admittance parameters $\boldsymbol{Y}$ of the two-port. We can recall the conversion between the noise sources (2.1) and rewrite it in the matrix form

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} -Y_{11} & 1 \\ -Y_{21} & 0 \end{bmatrix} \begin{bmatrix} u \\ i \end{bmatrix} \tag{2.27}$$

and

$$\begin{bmatrix} u \\ i \end{bmatrix} = \begin{bmatrix} 0 & -1/Y_{21} \\ 1 & -Y_{11}/Y_{21} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix}. \tag{2.28}$$

Thus, based on the definition of the correlation matrices, we have

$$\boldsymbol{C}_Y = \boldsymbol{T}_A \boldsymbol{C}_A \boldsymbol{T}_A^\dagger \tag{2.29}$$

and

$$\boldsymbol{C}_A = \boldsymbol{T}_Y \boldsymbol{C}_Y \boldsymbol{T}_Y^\dagger \tag{2.30}$$

where † means conjugate transpose, and $\boldsymbol{T}_A$ and $\boldsymbol{T}_Y$ are the two transformational matrices given by

$$\boldsymbol{T}_A = \begin{bmatrix} -Y_{11} & 1 \\ -Y_{21} & 0 \end{bmatrix} \tag{2.31}$$

and

$$\boldsymbol{T}_Y = \begin{bmatrix} 0 & -1/Y_{21} \\ 1 & -Y_{11}/Y_{21} \end{bmatrix}. \tag{2.32}$$

Here $\boldsymbol{T}_A$ and $\boldsymbol{T}_Y$ are reciprocal, i.e.,

$$\boldsymbol{T}_Y \boldsymbol{T}_A = \boldsymbol{I}_2. \tag{2.33}$$

The noise correlation matrices are equivalent with the four noise parameters in the sense that they can both fully describe the noise properties of the linear two-port. Therefore, it is possible to convert between them. Using the definition in (2.25) and recalling (2.5) ~ (2.7), we can express $\boldsymbol{C}_A$ as

$$\boldsymbol{C}_A = \begin{bmatrix} R_n & Y_{cor}^* R_n \\ Y_{cor} R_n & G_i \end{bmatrix} = \begin{bmatrix} R_n & \dfrac{NF_{\min} - 1}{2} - R_n Y_{opt}^* \\ \dfrac{NF_{\min} - 1}{2} - R_n Y_{opt} & R_n \left| Y_{opt} \right|^2 \end{bmatrix}. \tag{2.34}$$

For $\boldsymbol{C}_Y$, we can directly use (2.29) to convert from $\boldsymbol{C}_A$.

2.2.4 Noise Factor in Terms of Noise Correlation Matrices

In the previous derivations, the noise factor is expressed as a function of the noise parameters and meanwhile the noise parameters are equivalent with the noise correlation matrices. Therefore, it is possible to express the noise factor in terms of the noise correlation matrices. Based on the noise factor expression in (2.11), we can rewrite the noise factor in terms of the noise sources i and u, and the source admittance Y_S and its real part G_S by

$$F = 1 + \frac{1}{4kT\Delta f \cdot G_S}\overline{\left|i + Y_S u\right|^2}. \tag{2.35}$$

Then we can rearrange the terms and express the $|\cdot|^2$ term in the form of matrix products, i.e.,

$$4kT\Delta f\,(F-1)\cdot G_S = \overline{\left|i + Y_S u\right|^2} = \begin{bmatrix} Y_S & 1 \end{bmatrix} \begin{bmatrix} \overline{u^2} & \overline{ui^*} \\ \overline{u^* i} & \overline{i^2} \end{bmatrix} \begin{bmatrix} Y_S \\ 1 \end{bmatrix}^{\dagger}. \tag{2.36}$$

With the definition of $\boldsymbol{C}_A$ in (2.26) and the two-element vector $\boldsymbol{t}_A$ defined as

$$\boldsymbol{t}_A = \begin{bmatrix} \dfrac{Y_S}{\sqrt{G_S}} & \dfrac{1}{\sqrt{G_S}} \end{bmatrix}, \tag{2.37}$$

we can reach the following compact form of noise factor as

$$F = 1 + \boldsymbol{t}_A \boldsymbol{C}_A \boldsymbol{t}_A^{\dagger}. \tag{2.38}$$

Similarly, by using the conversion equation (2.30), we can also write F in terms of $\boldsymbol{C}_Y$ by

$$F = 1 + \boldsymbol{t}_Y \boldsymbol{C}_Y \boldsymbol{t}_Y^{\dagger} \tag{2.39}$$

where the vector $\boldsymbol{t}_Y$ is

$$\boldsymbol{t}_Y = \boldsymbol{t}_A \boldsymbol{T}_Y = \frac{1}{\sqrt{G_S}} \cdot \begin{bmatrix} 1 & -\dfrac{Y_{11} + Y_S}{Y_{21}} \end{bmatrix}. \tag{2.40}$$

Here, we have successfully arrived at the matrix representation of the noise factor using (2.38) or (2.39). As will be seen in the following derivation, it serves as the essential bridge that connects noise factor to deembedding theories based on noise correlation matrices.

2.3 Noise Factor Deembedding

The last section is about the noise theory of the linear two-port network itself and the case where only an input noise source is present. In practice, the two-port network, or the DUT, works together with other components in the system. The configuration can vary, e.g., a cascade of two-ports, two-port surrounded by a four-port parasitics, etc. In these cases, the directly measured noise factor is not for the DUT itself, but for the combination of networks. Deembedding has to be performed in post-processing to achieve the noise factor of the target device.

2.3.1 Deembedding Noise Factors for Two- and Three-Ports

The well-known Friis' equation can handle the deembedding of noise factors for two-ports in a cascade configuration, which takes the following form [1]

$$F_{total} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots, \tag{2.41}$$

where F_{total} is the overall noise factor, F_i and G_i are the noise factor and gain, respectively, for the two-port networks connected in cascade, as shown in Fig. 2.3. If the noise factor of one of the networks is unknown but F_{total} is measured, it is possible to rearrange the equation and solve for it. This can be counted as a form of noise factor deembedding for a cascade of two-port networks.

Abidi and Leete [26] extended the Friis' equation to three-port devices, such as baluns that are used to convert between single-ended signal and differential signal. Fig. 2.4 shows

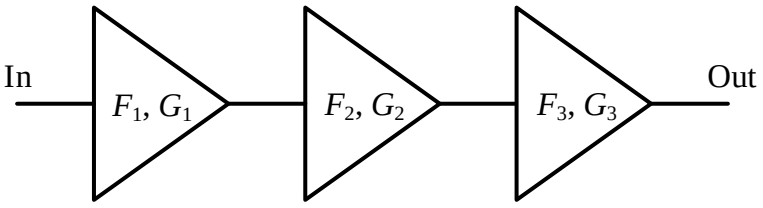


Fig. 2.3. A cascade of two-port networks with their noise factor as F_i and gain as G_i, where i = 1, 2, 3… The overall noise factor can be expressed as a function of each component's noise factor and gain using Friis' equation [1].

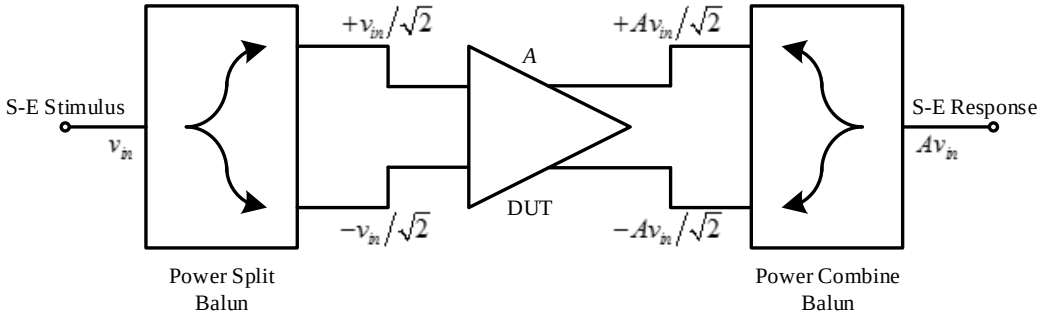


Fig. 2.4. Two-port DUT (differential amplifier) surrounded by two baluns [26]. The baluns are three-port devices that can split power from its input port equally to the two output ports, or combine power in the reverse way. (© 2015 IEEE.)

the two-port DUT, which is a differential amplifier, surrounded by two baluns. The baluns are three-port devices that can split power from its input port equally to the two output ports, or combine power if connected in the reverse way. The noise factor and gain of the two baluns can only be measured in single-ended mode. If we denote the noise factor and gain of the input balun as F_1 and G_1, those of the output balun as F_2 and G_2, and those of

the DUT as F and A, respectively, the overall noise factor of the system can be expressed as [26]

$$F_{total} = \frac{1}{2}F_1 + \frac{1}{2}\frac{F-1}{G_1} + \frac{1}{4}\frac{F_2-2}{AG_1}. \tag{2.42}$$

It can be easily applied to deembedding once the gain and noise factors of the baluns, as well as those of the system, are measured.

In both the two-port and three-port cases described above, the input and output ports are isolated and the latter cannot interfere with the former. However, for on-wafer DUTs where the parasitics can form a complicated network and introduce feedback paths from the output to the input ports, the Friis' equation and its extended three-port variant cease to work. Existing theories for noise deembedding for this kind of problems are based on noise correlation matrices, or equivalently the four noise parameters. Noise factor deembedding in this situation remains unsolved.

2.3.2 Deembedding Noise Parameters for Four-Port Parasitics

In this subsection, we review the existing theories for noise parameter deembedding with four-port parasitics, which also serve as the theoretical foundation for the subsequent derivation on noise factor deembedding.

The device-under-test (DUT) on a wafer, also known as the extrinsic device, can be modeled as a two-port intrinsic device enclosed by a four-port parasitic network. As defined in Fig. 2.5, the extrinsic device can be modeled by two-port admittance matrix $\boldsymbol{Y}_{\text{ex}}$

and corresponding noise correlation matrix $\boldsymbol{C}_{\text{ex}}$ whereas the intrinsic device by $\boldsymbol{Y}_{\text{in}}$ and $\boldsymbol{C}_{\text{in}}$. These matrices, namely $\boldsymbol{Y}_{\text{ex}}$, $\boldsymbol{C}_{\text{ex}}$, $\boldsymbol{Y}_{\text{in}}$, $\boldsymbol{C}_{\text{in}}$, are all 2x2 matrices. The four-port parasitic network can be modeled by a 4x4 matrix $\boldsymbol{Y}_p$ which can be further partitioned as [29],[30]

$$\boldsymbol{Y}_p = \begin{bmatrix} \boldsymbol{Y}_{ee} & \boldsymbol{Y}_{ei} \\ \boldsymbol{Y}_{ie} & \boldsymbol{Y}_{ii} \end{bmatrix} \tag{2.43}$$

where the subscript *e* refers to the extrinsic ports (1-1' and 2-2') and the subscript *i* refers to the intrinsic ports (3-3' and 4-4'). Based on the assumption that the four-port parasitic network is linear and passive, its noise correlation matrix can be written as [29],[50]

$$\boldsymbol{C}_p = \frac{T}{T_0} \cdot \text{Re}(\boldsymbol{Y}_p) \tag{2.44}$$

where T is the ambient temperature and T_0 is the standard reference temperature 290 K.

Based on the configuration of the networks shown in Fig. 2.5, it is possible to derive the mathematical relations among the admittance matrices and noise correlation matrices, which serve as the foundation for the deembedding of noise parameters. To express extrinsic quantities using intrinsic quantities, we have

$$\boldsymbol{Y}_{ex} = \boldsymbol{Y}_{ee} - \boldsymbol{Y}_{ei}\left(\boldsymbol{Y}_{ii} + \boldsymbol{Y}_{in}\right)^{-1}\boldsymbol{Y}_{ie} \tag{2.45}$$

and

$$\boldsymbol{C}_{ex} = \boldsymbol{D}_{in}\boldsymbol{C}_{in}\boldsymbol{D}_{in}^{\dagger} + \boldsymbol{P}_{in}\boldsymbol{C}_p\boldsymbol{P}_{in}^{\dagger} \tag{2.46}$$

where $\boldsymbol{D}_{\text{in}}$ and $\boldsymbol{P}_{in}$ are transformational matrices that are given by

$$\boldsymbol{D}_{in} = -\boldsymbol{Y}_{ei}(\boldsymbol{Y}_{ii} + \boldsymbol{Y}_{in})^{-1} \tag{2.47}$$

and

$$\boldsymbol{P}_{in} = \left[\boldsymbol{I}_2 \mid \boldsymbol{D}_{in}\right]. \tag{2.48}$$

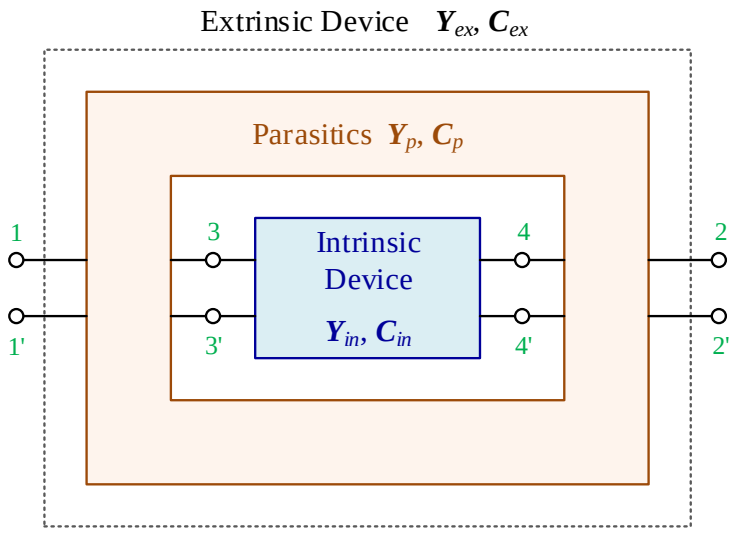


Fig. 2.5. Admittance matrices and noise correlation matrices of the 2-port extrinsic device ($\boldsymbol{Y}_{\text{ex}}$ and $\boldsymbol{C}_{\text{ex}}$), the 4-port parasitics ($\boldsymbol{Y}_p$ and $\boldsymbol{C}_p$) to model the probe pads and metal interconnections, and the 2-port intrinsic device ($\boldsymbol{Y}_{\text{in}}$ and $\boldsymbol{C}_{\text{in}}$). (© 2017 IEEE.)

On the other hand, the reverse direction of the above equations can also be derived as

$$\boldsymbol{Y}_{in} = \boldsymbol{Y}_{ie}\left(\boldsymbol{Y}_{ee} - \boldsymbol{Y}_{ex}\right)^{-1}\boldsymbol{Y}_{ei} - \boldsymbol{Y}_{ii} \tag{2.49}$$

and

$$\boldsymbol{C}_{in} = \boldsymbol{D}_{ex}\boldsymbol{C}_{ex}\boldsymbol{D}_{ex}^{\dagger} - \boldsymbol{P}_{ex}\boldsymbol{C}_{p}\boldsymbol{P}_{ex}^{\dagger} \tag{2.50}$$

where $\boldsymbol{D}_{\text{ex}}$ and $\boldsymbol{P}_{\text{ex}}$ are given by

$$\boldsymbol{D}_{ex} = -\boldsymbol{Y}_{ie}(\boldsymbol{Y}_{ee} - \boldsymbol{Y}_{ex})^{-1} \tag{2.51}$$

and

$$\boldsymbol{P}_{ex} = \left[\boldsymbol{D}_{ex} \mid \boldsymbol{I}_2\right]. \tag{2.52}$$

In addition, $\boldsymbol{D}_{in}$ and $\boldsymbol{D}_{ex}$ are reciprocal matrices, i.e., $\boldsymbol{D}_{in} = \boldsymbol{D}_{ex}^{-1}$. Based on the above noise deembedding theory using noise correlation matrices, we will derive the noise factor deembedding theory in the next subsection.

2.3.3 Deembedding Noise Factor for Four-Port Parasitics

By applying the noise factor equation (2.39) on the extrinsic device and using subscript *ex* to indicate extrinsic quantities, we can express its noise factor as

$$F_{ex} = 1 + \boldsymbol{t}_{ex}\boldsymbol{T}_{ex}\boldsymbol{C}_{ex}\left(\boldsymbol{t}_{ex}\boldsymbol{T}_{ex}\right)^{\dagger} \tag{2.53}$$

where

$$\boldsymbol{t}_{ex} = \frac{1}{\sqrt{G_S^{ex}}}\begin{bmatrix} Y_S^{ex} & 1 \end{bmatrix} \tag{2.54}$$

and

$$\boldsymbol{T}_{ex} = \begin{bmatrix} 0 & -1/Y_{21}^{ex} \\ 1 & -Y_{11}^{ex}/Y_{21}^{ex} \end{bmatrix}. \tag{2.55}$$

Similarly, by changing the subscript or superscript *ex* to *in*, we can express the noise factor of the intrinsic two-port network as

$$F_{in} = 1 + \boldsymbol{t}_{in}\boldsymbol{T}_{in}\boldsymbol{C}_{in}\left(\boldsymbol{t}_{in}\boldsymbol{T}_{in}\right)^{\dagger} \tag{2.56}$$

where

$$\boldsymbol{t}_{in} = \frac{1}{\sqrt{G_S^{in}}}\begin{bmatrix} Y_S^{in} & 1 \end{bmatrix} \tag{2.57}$$

and

$$\boldsymbol{T}_{in} = \begin{bmatrix} 0 & -1/Y_{21}^{in} \\ 1 & -Y_{11}^{in}/Y_{21}^{in} \end{bmatrix}. \tag{2.58}$$

By substituting $\boldsymbol{C}_{\text{ex}}$ in (2.53) using (2.46), we can write

$$F_{ex} = 1 + \Delta F_p + \Delta F_{in} \tag{2.59}$$

where ΔF_p and ΔF_{in} are the contribution of the package and the intrinsic device to the extrinsic noise factor, respectively, and they are defined as

$$\Delta F_p = \left(\boldsymbol{t}_{ex}\boldsymbol{T}_{ex}\boldsymbol{P}\right)\boldsymbol{C}_p\left(\boldsymbol{t}_{ex}\boldsymbol{T}_{ex}\boldsymbol{P}\right)^{\dagger} \tag{2.60}$$

and

$$\Delta F_{in} = \left(\boldsymbol{t}_{ex}\boldsymbol{T}_{ex}\boldsymbol{D}\right)\boldsymbol{C}_{in}(\boldsymbol{t}_{ex}\boldsymbol{T}_{ex}\boldsymbol{D})^{\dagger}. \tag{2.61}$$

To find the relationship between F_{in} and ΔF_{in}, we define

$$\boldsymbol{t}'_{in} \equiv \left(\boldsymbol{t}_{ex}\boldsymbol{T}_{ex}\boldsymbol{D}\right)\boldsymbol{T}_{in}^{-1} \tag{2.62}$$

which is a two-element complex row vector, and use complex scalar a and b to denote its two elements, i.e.,

$$\boldsymbol{t}'_{in} \equiv \begin{bmatrix} a & b \end{bmatrix} \equiv b \cdot \begin{bmatrix} a/b & 1 \end{bmatrix}. \tag{2.63}$$

Then we can rewrite the contribution of the intrinsic device to the noise factor (2.61) as

$$\Delta F_{in} = \boldsymbol{t}'_{in}\boldsymbol{T}_{in}\boldsymbol{C}_{in}\left(\boldsymbol{t}'_{in}\boldsymbol{T}_{in}\right)^{\dagger} \tag{2.64}$$

which has the same format as the 2nd term in (2.56) except $\boldsymbol{t_{in}}$ being replaced by $\boldsymbol{t}'_{in}$. This suggests that $\boldsymbol{t}'_{in}$ has the same characteristic as $\boldsymbol{t_{in}}$. Comparing their formats in (2.57) and (2.63), we can then define the source admittance seen by the intrinsic device Y_S^{in} as

$$Y_S^{in} = a/b. \tag{2.65}$$

In order to connect the contribution of the intrinsic device to the extrinsic noise factor ΔF_{in} given by (2.64) with the intrinsic noise factor F_{in} given by (2.56), we can derive the ratio between ΔF_{in} and $F_{in} - 1$ as

$$\begin{aligned}\frac{\Delta F_{in}}{F_{in}-1} &= \frac{\boldsymbol{t}'_{in}\boldsymbol{T}_{in}\boldsymbol{C}_{in}\left(\boldsymbol{t}'_{in}\boldsymbol{T}_{in}\right)^{\dagger}}{\boldsymbol{t}_{in}\boldsymbol{T}_{in}\boldsymbol{C}_{in}\left(\boldsymbol{t}_{in}\boldsymbol{T}_{in}\right)^{\dagger}} \\ &= \frac{\boldsymbol{t}'_{in}\cdot\boldsymbol{t}'^{\dagger}_{in}}{\boldsymbol{t}_{in}\cdot\boldsymbol{t}^{\dagger}_{in}} \\ &= bb^{*}\cdot G_{S}^{in} \\ &= bb^{*}\cdot \mathrm{Re}\left(a/b\right) \\ &= \mathrm{Re}\left(ab^{*}\right).\end{aligned} \tag{2.66}$$

Given that ΔF_{in} can be separated from extrinsic noise factor using (2.59), we can then reach the expression for the intrinsic noise factor of the device as

$$F_{in} = 1+\frac{\Delta F_{in}}{\mathrm{Re}(ab^{*})}. \tag{2.67}$$

Therefore, with a and b being the two elements of $\boldsymbol{t}'_{in}$ given by (2.62), and ΔF_{in} given by (2.64), the intrinsic noise factor F_{in} and source admittance Y_S^{in} are achieved using only the extrinsic noise factor and the admittance parameters of the DUT and the parasitic network based on (2.65) and (2.67).

Based on the equations derived, we can summarize the noise factor deembedding procedure as follows:

Part 1: Data Preparation

1. Measure the S parameters of the DUT and convert them to the admittance form $\boldsymbol{Y}_{\mathrm{ex}}$.
2. Measure the S parameters of the dummy structures and compute the four-port admittance matrix $\boldsymbol{Y}_p$ (depending on the modeling of the parasitics).
3. Measure the noise factor F_{ex} of the DUT together with the corresponding source admittance Y_{ex}^{S} .

Part 2: Calculation

4. Calculate the intrinsic admittance matrix $\boldsymbol{Y}_{\text{in}}$ using (2.49), and the transformational matrices $\boldsymbol{D}$ and $\boldsymbol{P}$ using (2.47).
5. Calculate the extrinsic matrices $\boldsymbol{t}_{ex}$ and $\boldsymbol{T}_{ex}$ using (2.54) and (2.55), and the intrinsic matrix $\boldsymbol{T}_{in}$ using (2.58), respectively.
6. Calculate $\boldsymbol{C}_p$ using (2.44) and ΔF_p using (2.60).
7. Calculate ΔF_{in} by $\Delta F_{in} = F_{ex} - 1 - \Delta F_p$ based on (2.59).
8. Calculate the matrix $\boldsymbol{t}'_{in}$ using (2.62), and thus its two elements a and b.
9. Finally, calculate the intrinsic source admittance Y_{in}^{S} using (2.65) and noise factor F_{in} using (2.67).

2.3.4 New Approach for Deembedding Noise Parameters

The theory described in the previous subsection complements the existing noise factor deembedding theories. It enables the de-embedding of a single noise factor for an active two-port surrounded by a passive four-port without requiring the four noise parameters. Nevertheless, the theory can be applied to noise parameter de-embedding.

Traditional approaches for noise parameters deembedding are built upon the theory introduced in Section 2.3.2 which relies on the complete noise information of the DUT, i.e., the four noise parameters or the noise correlation matrices. It indicates that they invariably require invoking the optimization procedure to obtain the noise parameters before performing the deembedding procedure. Fig. 2.6 (a) shows the workflow for traditional noise parameter deembedding approaches, in which the optimization is first performed to

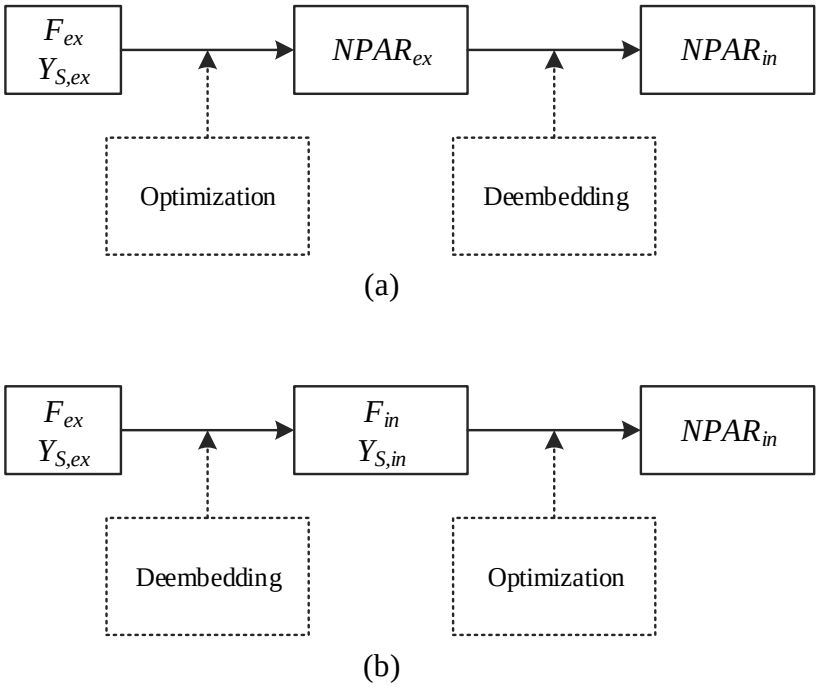


Fig. 2.6. Workflow for (a) traditional noise parameter deembedding approaches that perform optimization first and deembedding last; and (b) the proposed noise parameter deembedding approach that does deembedding first and optimization last.

compute the extrinsic noise parameters from extrinsic noise factors, and then followed by the deembedding to obtain the intrinsic noise parameters.

With the proposed noise factor deembedding algorithm, it is no longer required to have the complete noise parameters of the DUT before deembedding can be performed. This leads to a new approach for noise parameter deembedding in which we can first deembed the extrinsic noise factors and then run optimization to obtain the final intrinsic noise parameters. The workflow of the new approach is shown in Fig. 2.6 (b).

If we compare the two operations involved in those approaches, the deembedding is an arithmetic operation that works on measured *S*-parameters of the parasitics with

deterministic effects, while the optimization is a statistical operation that is relatively flexible. It happens in traditional approaches that the optimization produces extrinsic noise parameters that are physical, but the subsequent deembedding results in nonphysical values, such as $F_{\min} < 1$ or $|\Gamma_{opt}| = 1$ for the intrinsic noise parameters [42]. The proposed noise parameter deembedding approach can improve this situation by having the optimization as the last procedure, which can introduce the certain constraints to avoid the nonphysical values.

In addition, the noise factor deembedding algorithm can be used to examine the quality of the measured noise factors. For the extrinsic noise factors to be physical, we need to ensure $F_{ex} > 1$; after deembedding, we can recover another constraint that $F_{in} > 1$ for the intrinsic noise factors. It can be utilized by a data screening process to remove the nonphysical data points that may not appear to be by first look. In the proposed approach, this helps to improve the quality of input data for the optimization procedure and thus reduce the chance of the final intrinsic noise parameters being nonphysical.

2.3.5 Modeling Parasitics Using OPEN and SHORT Structures

Before we proceed to work with experimental data, another practical problem is the modeling of the parasitics. Physically, the parasitics of the DUTs in on-wafer noise measurements mainly come from the probe pads and metal interconnections. The proposed noise factor deembedding algorithm only requires the parasitics can be considered as a passive and linear four-port network whose admittance matrix $\boldsymbol{Y}_p$ can be extracted from

measurements. Due to the passive nature of the parasitics, the noise they add to the system is purely resistive noise and the corresponding noise correlation matrix is given by (2.44).

In practice, to extract the parasitic elements, it is necessary to have properly designed dummy structures on which the *S* parameter measurements can be performed. The dummy structures usually come with circuit models based on which the parasitic elements can be extracted. Existing models for the parasitics include parallel-series networks [43]-[45], cascade networks [46]-[48], as well as a general passive four-port network [28]. In the following, we will construct the four-port parasitic matrix $\boldsymbol{Y}_p$ based on one of the established parasitic models.

The parallel-series method using OPEN and SHORT dummy structures for modeling on-wafer parasitics is widely used in industry. It is relatively easy to implement and yet sufficiently accurate in the frequency range in this study as demonstrated in [44]. In practice, on-wafer OPEN and SHORT dummy structures are carefully designed to simulate the surroundings of the DUT. Their *S*-parameters are then measured and used in deembedding both *S*- and noise parameters of the DUT. A typical model of the OPEN and SHORT structures is shown by the equivalent circuits in Fig. 2.7 [44]. The OPEN structure can be modeled as a π network with three admittance elements Y_1, Y_2, and Y_3. Therefore, its admittance matrix can be written as

$$\boldsymbol{Y}_{open} = \begin{bmatrix} Y_1 + Y_2 & -Y_2 \\ -Y_2 & Y_2 + Y_3 \end{bmatrix}. \tag{2.68}$$

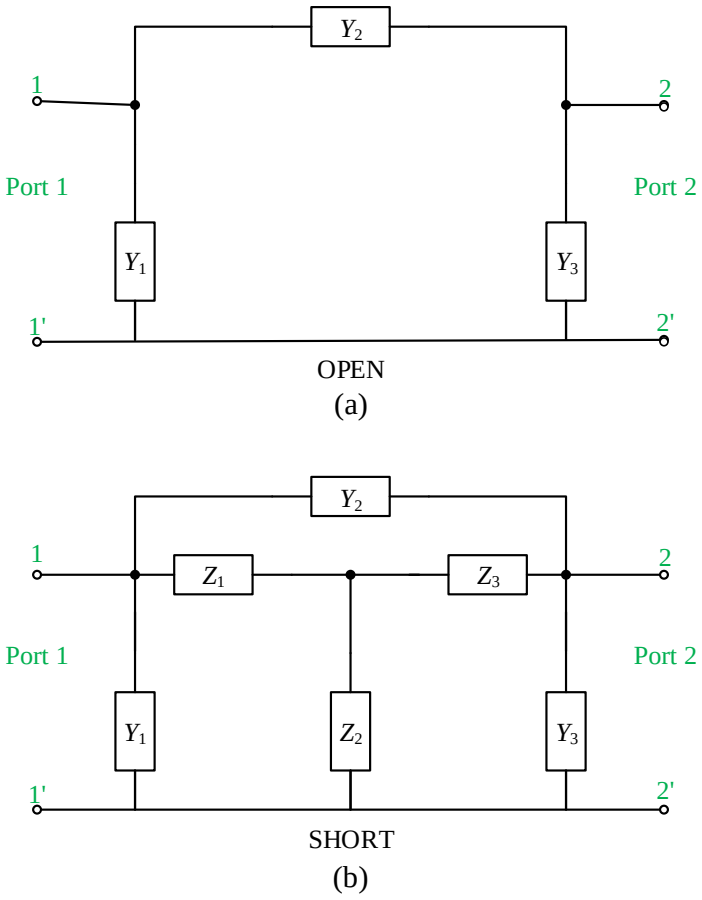


Fig. 2.7. Equivalent small signal circuits of (a) OPEN and (b) SHORT dummy structures [44]. Here Y_1, Y_2, Y_3 and Z_1, Z_2, Z_3 are complex admittances and impedances, respectively.

The SHORT structure can be modeled by adding a T network of three impedance elements Z_1, Z_2, and Z_3 in parallel with the OPEN network. As a result, its impedance matrix can be written as

$$\boldsymbol{Z}_{short} = \boldsymbol{Z}_{open} + \boldsymbol{Z}'_{short} = \boldsymbol{Y}_{open}^{-1} + \begin{bmatrix} Z_1 + Z_2 & Z_2 \\ Z_2 & Z_2 + Z_3 \end{bmatrix} \tag{2.69}$$

where $\boldsymbol{Z}_{open}$ is the impedance matrix of the OPEN structure $\boldsymbol{Z}_{open} = \boldsymbol{Y}_{open}^{-1}$ and $\boldsymbol{Z}'_{short}$ is the impedance matrix of the T network formed by Z_1, Z_2, and Z_3.

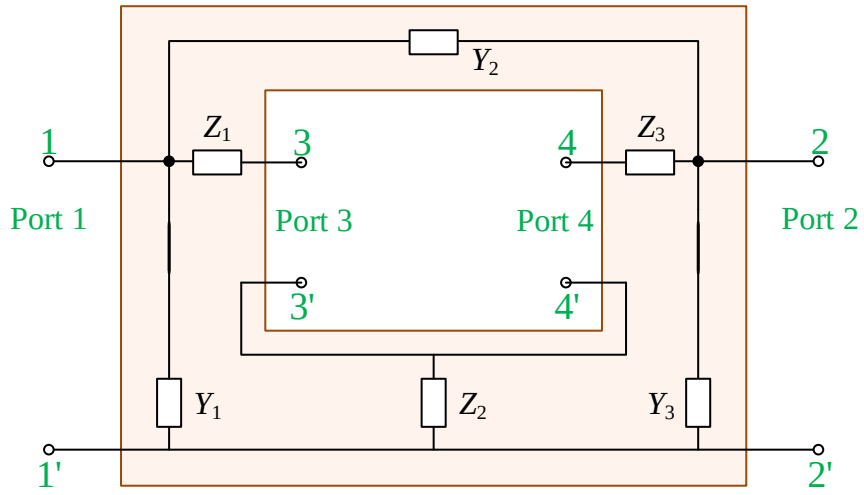


Fig. 2.8. Equivalent circuit of the four-port parasitic network. The admittance values of Y_1, Y_2, and Y_3, and the impedance values of Z_1, Z_2, and Z_3 are extracted from the measured S-parameters of the OPEN and SHORT dummy structures. The noise factor deembedding algorithm is independent of the modeling of the parasitics as long as they can be represented as a linear passive four-port network. (© 2017 IEEE.)

With all the admittances Y_1, Y_2, and Y_3 and impedances Z_1, Z_2, and Z_3, we can draw the four-port model of the parasitics as shown in Fig. 2.8. The overall extrinsic device can be seen as the intrinsic device in series with the T network, and then in parallel with the π network. Therefore, we can express the relation between extrinsic admittance matrix $\boldsymbol{Y}_{\text{ex}}$ and intrinsic admittance matrix $\boldsymbol{Y}_{\text{in}}$ as

$$\boldsymbol{Y}_{ex} = \boldsymbol{Y}_{open} + (\boldsymbol{Z}'_{short} + \boldsymbol{Y}_{in}^{-1})^{-1} \tag{2.70}$$

and

$$\boldsymbol{Y}_{in} = \left((\boldsymbol{Y}_{ex} - \boldsymbol{Y}_{open})^{-1} - \boldsymbol{Z}'_{short}\right)^{-1}. \tag{2.71}$$

This is also the basis of how deembedding is performed traditionally based on OPEN and SHORT structures.

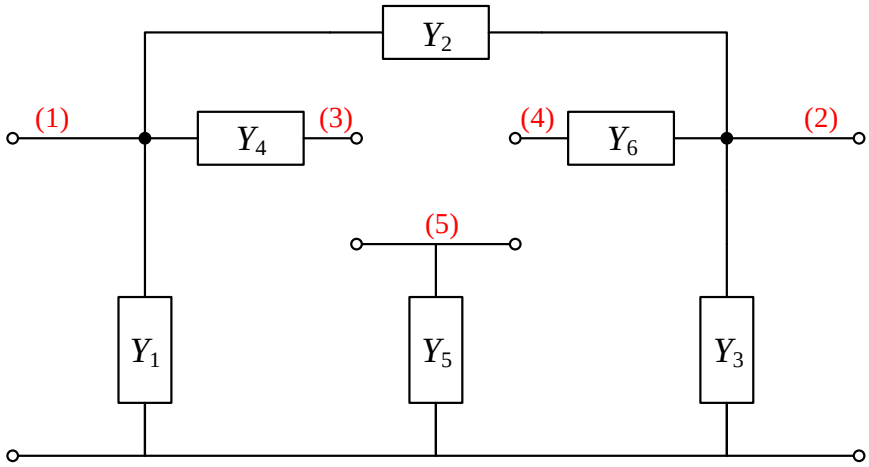


Fig. 2.9. Equivalent circuit of the parasitics with nodal assignments. Nodal elimination is required to reduce node (5) and achieve the 4x4 matrix $\boldsymbol{Y}_p$.

The previously introduced noise factor deembedding algorithm does not depend on the specific OPEN-SHORT model. It only requires the parasitics to be a linear passive four-port network. In this case, to implement the algorithm, we will need to construct the four-port admittance matrix based on the equivalent circuit in Fig. 2.8, which can be achieved by the method of nodal analysis. Based on the nodal assignments as shown in Fig. 2.9 and using the Kirchhoff's current law (KCL), we can describe the equivalent circuit using a 5x5 matrix by

$$\boldsymbol{Y}_{p,raw} = \begin{bmatrix} Y_1+Y_2+Y_4 & -Y_2 & -Y_4 & 0 & 0 \\ -Y_2 & Y_2+Y_3+Y_6 & 0 & -Y_6 & 0 \\ -Y_4 & 0 & Y_4 & 0 & 0 \\ 0 & -Y_6 & 0 & Y_6 & 0 \\ 0 & 0 & 0 & 0 & Y_5 \end{bmatrix} \tag{2.72}$$

where $Y_4 = 1 / Z_4$, $Y_5 = 1 / Z_5$, and $Y_6 = 1 / Z_6$. To reduce node (5) and achieve the 4x4 admittance matrix $\boldsymbol{Y}_p$, we can use the method of nodal elimination and each element in $\boldsymbol{Y}_p$ can be obtained by

$$Y_{ij}^{'} = Y_{ij} - \frac{Y_{5j} \times Y_{i5}}{Y_{55}}, \text{ where } 1 \le i < 5 \text{ and } 1 \le j < 5. \tag{2.73}$$

2.4 Experimental Results and Discussions

In this section, we will perform verification of the proposed deembedding algorithm using idealized data, as well as evaluation of the deembedding-first and optimization-last approach for noise parameter deembedding against the traditional approach.

2.4.1 Verification of the Noise Factor Deembedding Algorithm

Verification of the noise factor algorithm can be performed using idealized data which are free from measurement uncertainties. The procedure is shown by the workflow in Fig. 2.10. For the input, we prepare a sample set of extrinsic noise parameters as well as sample admittance matrices of the extrinsic device $\boldsymbol{Y}_{\text{ex}}$ and the parasitics $\boldsymbol{Y}_p$, all of which are chosen to resemble real experimental data. The parasitic matrix $\boldsymbol{Y}_p$ is constructed using (2.72) and (2.73) based on the OPEN-SHORT model. The element values in the OPEN-SHORT equivalent circuits are extracted from measured *S*-parameters of the dummy structures. The intrinsic device is modeled after the internal part of the MOSFET model with noise sources in Fig. 1 in [18]. The values of the elements and noise sources are extracted from measured *S*-parameters and noise parameters. Based on the models for the parasitics and intrinsic

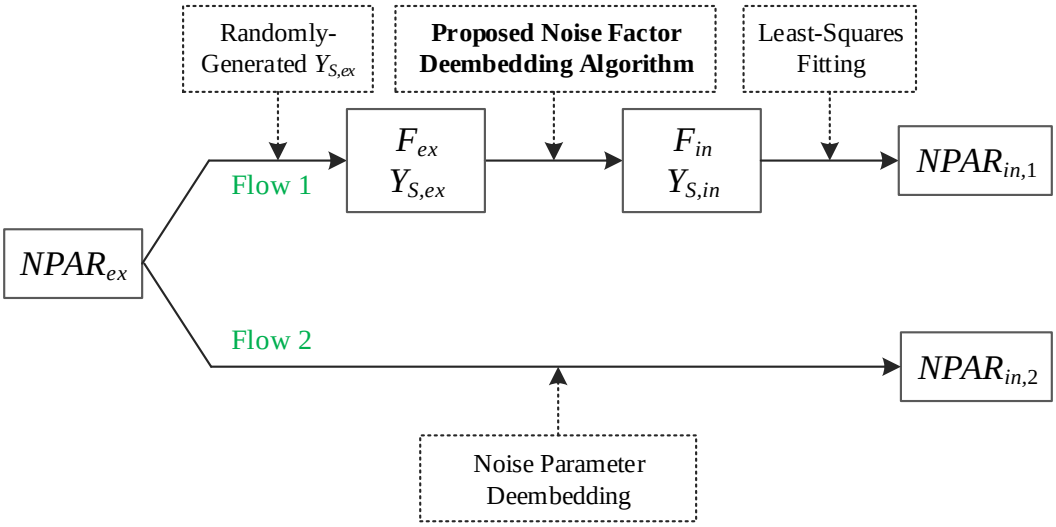


Fig. 2.10. Workflow of the verification procedure for the noise factor deembedding algorithm. *NPAR* stands for noise parameters, *F* for noise factors, and Y_S for the corresponding source admittances. The subscripts "*ex*" and "*in*" represent the quantity for the extrinsic or intrinsic device, respectively.

device, as well as the corresponding element values, the extrinsic admittance matrix $\boldsymbol{Y}_{\mathrm{ex}}$ and noise parameters $NPAR_{\mathrm{ex}}$ can be calculated at any given frequency, which serve as the input data for the verification procedure.

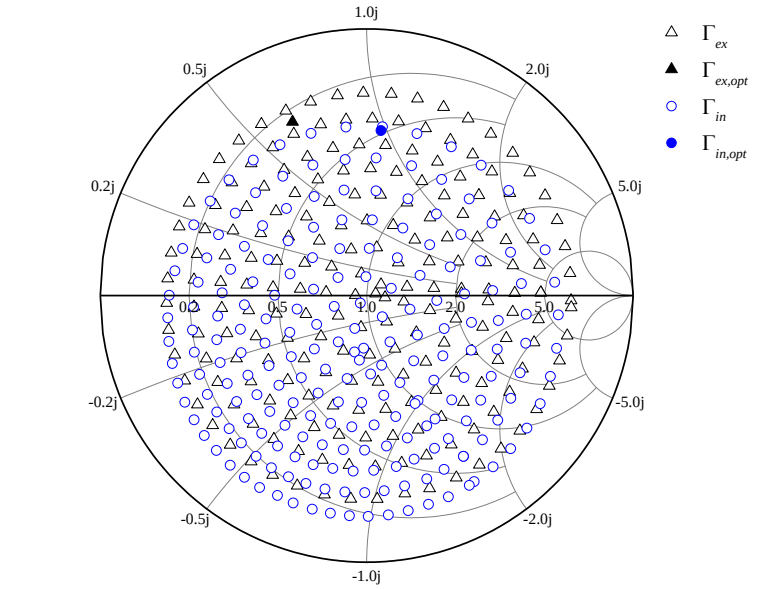
On one hand, shown by the bottom branch in the workflow, we can directly compute the intrinsic noise parameters using the noise parameter deembedding theory introduced in Section 3.3.2. On the other hand, shown by the top branch, we can also generate a number of (≥ 4, usually more than 20) source admittances $Y_{S\mathrm{ex}}$ and calculate the corresponding extrinsic noise factors using (2.16); we then feed them into the proposed noise parameter deembedding approach, i.e., first noise factor deembedding to get the intrinsic noise factors and then optimization to get the intrinsic noise parameters. For the optimization algorithm,

we choose to use Lane's least-squares fitting method for its simplicity. In fact, for idealized data without measurement errors, all the optimization algorithms should produce the same results. Finally, by performing both branches independently, it is found that the difference in the resulted intrinsic noise parameters is at a level dominated by machine precision. This proves the numerical accuracy of the proposed noise factor deembedding algorithm.

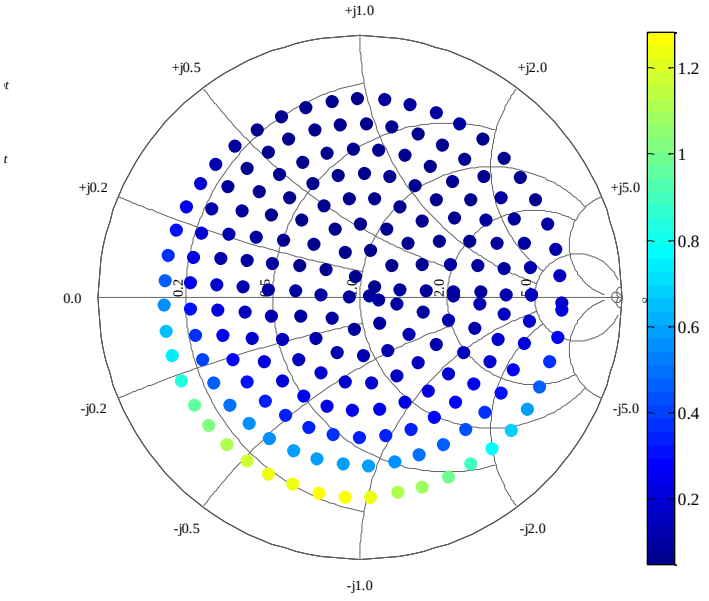
2.4.2 Visualization of the Deembedded Noise Factors

The proposed noise factor deembedding algorithm deducts the contribution of the parasitics from the extrinsic noise factor, as well as modifies the extrinsic source reflection coefficient to obtain the intrinsic value seen by the intrinsic device. To visualize the noise deembedding algorithm in action, we compare the noise factors and the corresponding source correlation matrices before and after deembedding using experimental data. The measured extrinsic noise factors come from an exhaustive measurement of around 200 source reflection coefficients that are evenly distributed on the Smith Chart at 20 GHz. The noise measurement system is from Focus Microwaves and consists of a noise source, a PNA-X microwave network analyzer with noise measurement capability, a microwave tuner, a low-noise amplifier, and other peripheral components like microwave switches and controllers, as shown in Fig. 3.4 (a). The DUT is an *n*-MOSFET with W/L = 128 × 1 μm / 90 nm fabricated in UMC's 90-nm CMOS technology. It is biased at V_{GS} = 1.0 V and V_{DS} = 0.8 V with DC current I_{DS} = 36.8 mA.

When applying the deembedding algorithm, we still use the OPEN-SHORT model in Fig. 2.8 to construct the parasitic matrix $\boldsymbol{Y}_p$. The extracted values for the parasitic elements



(a)



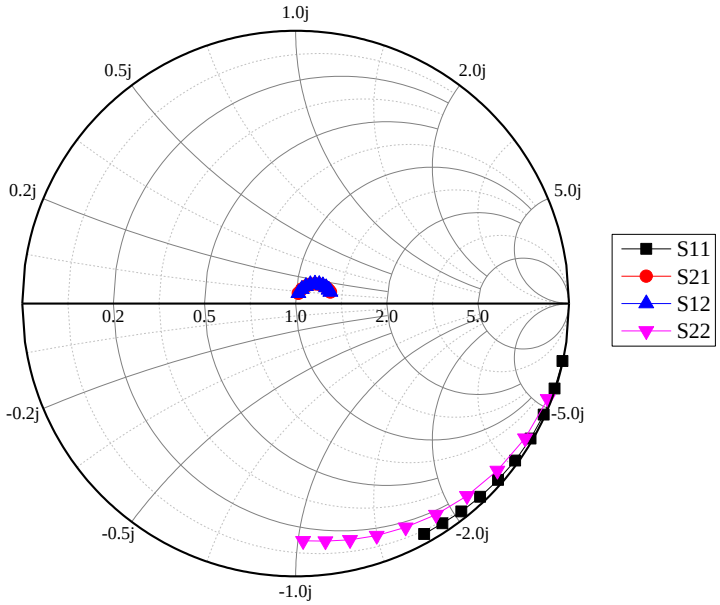
(b)

Fig. 2.11. (a) Measured extrinsic source reflection coefficients Γ_{ex}, the corresponding deembedded intrinsic source reflection coefficients Γ_{in}, and the computed optimum values $\Gamma_{ex,opt}$ and $\Gamma_{in,opt}$ at 20 GHz for the 90-nm *n*-MOSFET. (b) The difference between the extrinsic noise factors F_{ex} and the deembedded intrinsic values F_{in} in linear scale, i.e., $F_{ex} - F_{in}$, plotted against Γ_{ex} at 20 GHz for the same device. (© 2017 IEEE.)

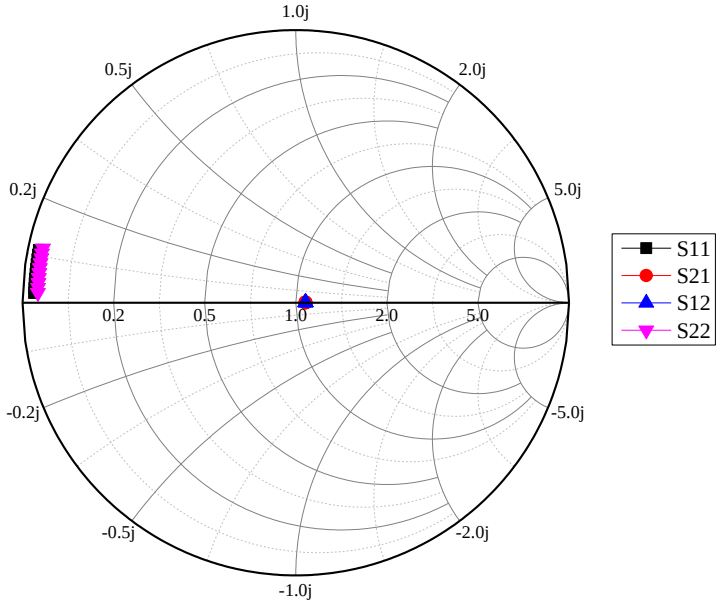
at 20 GHz are $Y_1 = 0.25 + j8.67$ mS, $Y_2 = 0.05 + j2.14$ mS, $Y_3 = 2.10 + j15.48$ mS, $Z_1 = 0.15 + j4.18$ Ω, $Z_2 = 0.82 + j0.26$ Ω, and $Z_3 = 0.36 + j4.26$ Ω, respectively. Fig. 2.11 (a) shows the measured extrinsic source reflection coefficients Γ_{ex}, the deembedded intrinsic ones Γ_{in} obtained using the proposed algorithm, as well as the computed optimum values, $\Gamma_{ex,opt}$ and $\Gamma_{in,opt}$, at 20 GHz. The shift in Γ_{opt} indicates that the parasitic admittance and impedance elements, which are largely imaginary, has a strong impact in changing the source reflection coefficient seen by the intrinsic device. Fig. 2.11 (b) shows the effect of deembedding on the noise factors by plotting the difference between the extrinsic noise factors F_{ex} and the intrinsic values F_{in} in linear scale, i.e., $F_{ex} - F_{in}$, against Γ_{ex} at 20 GHz. It shows that the reduction in extrinsic noise factor F_{ex} caused by deembedding is higher when the source reflection coefficient Γ_{ex} is farther away from the optimum value $\Gamma_{ex,opt}$.

2.4.3 Evaluation of the Noise Parameter Deembedding Procedure

Both *S*-parameter and noise parameter deembedding depends on the measured data of the parasitics. In the OPEN-SHORT method, the *S*-parameters of the OPEN and SHORT dummy structures are measured. Fig. 2.12 shows the *S*-parameters of the OPEN and SHORT structures designed for the 90-nm *n*-MOSFET for the frequency range of 4 to 22 GHz in Smith charts. Due to their passive nature, the magnitude of the S_{21} should be smaller than unity. Based on the measured *S*-parameters and the OPEN-SHORT model in Section 3.3.5, we can deembed the *S*-parameters of the DUT. The resulted intrinsic *S*-parameters, together with the corresponding extrinsic ones, are plotted in polar plots in Fig. 2.13 for the frequency range of 4 to 22 GHz.



(a)



(b)

Fig. 2.12. Measured *S*-parameters of the OPEN and SHORT dummy structures in Smith charts. The frequency range is from 4 to 22 GHz.

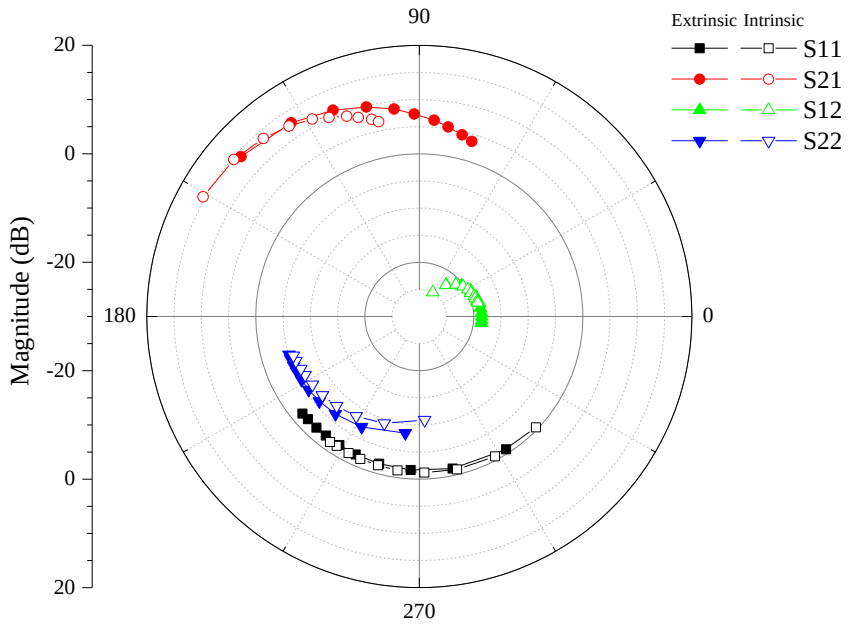
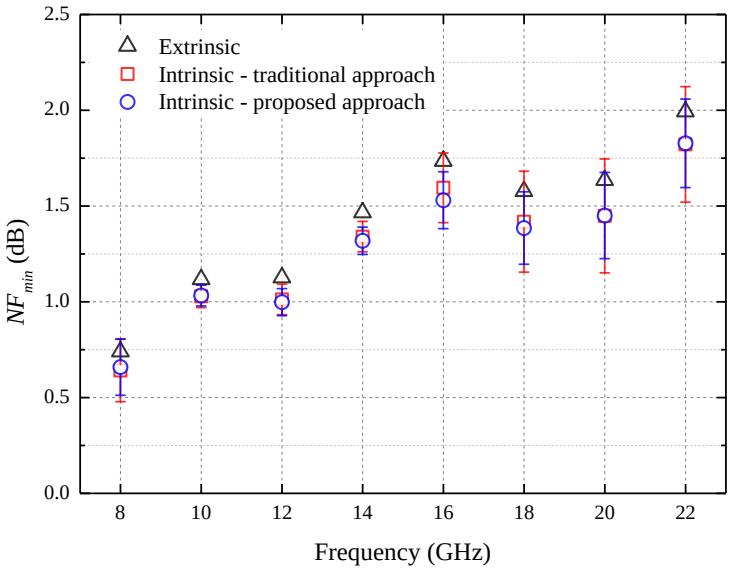
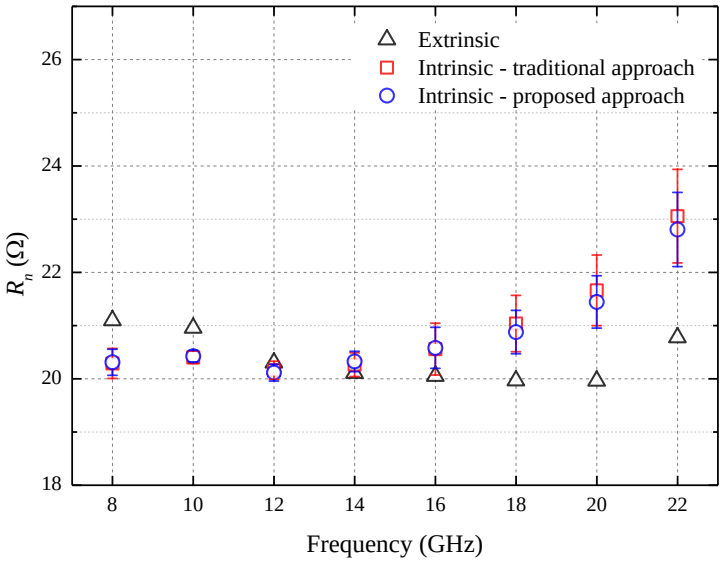


Fig. 2.13. Extrinsic (solid symbols) and intrinsic (empty symbols) *S*-parameters in a polar plot of an *n*-MOSFET with W/L = 128 × 1 μm / 90 nm biased V_{GS} = 1.0 V and V_{DS} = 0.8 V. The frequency range is from 4 to 22 GHz.

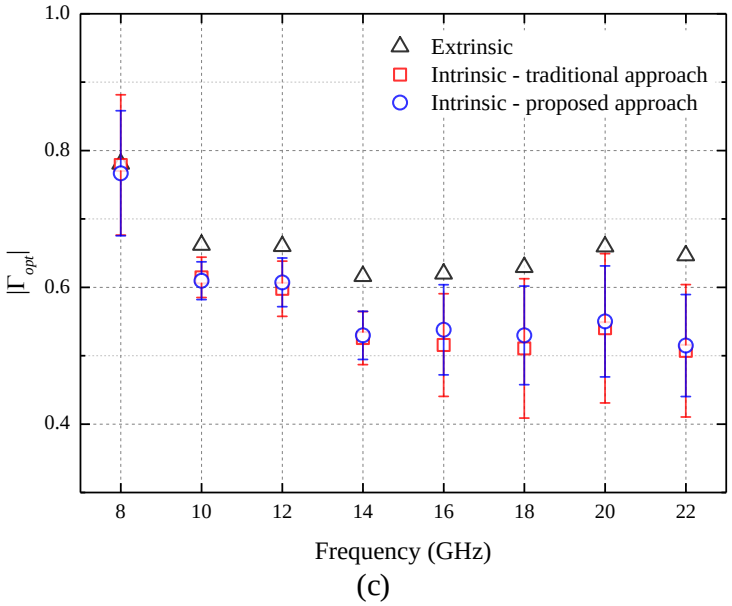
In the following, we perform noise parameter deembedding using both the traditional deembedding-last approach and the proposed optimization-last approach on the measured data of the 90-nm *n*-MOSFET. The procedures are described in Section 3.3.4. The resulted noise parameters, namely (a) NF_{min}, (b) R_n, (c) $|\Gamma_{opt}|$, and (d) $\angle\Gamma_{opt}$, are shown in Fig. 2.14. The symbols are calculated from around 200 noise factors at each frequency from 8 to 22 GHz. The error bars show the standard deviation [−σ, +σ] of the two sets of intrinsic noise parameters by randomly selecting 24 out of these 200 noise factors each time and repeating the calculation for 1000 times. While it is not possible to tell which set of symbols is more accurate, it can be observed that the proposed approach generally gives smaller error bars than the traditional approach.


Extrinsic
Intrinsic - traditional approach
Intrinsic - proposed approach
NF_{min} (dB)
Frequency (GHz)

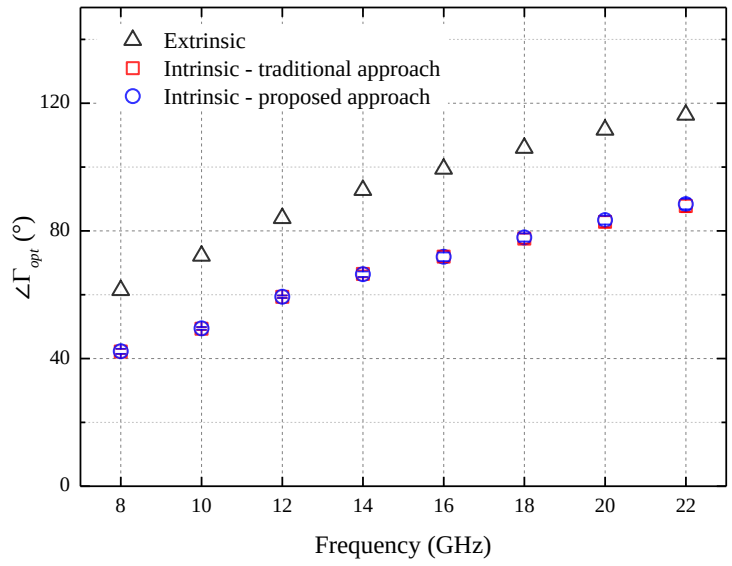
(a)


Extrinsic
Intrinsic - traditional approach
Intrinsic - proposed approach
R_n (Ω)
Frequency (GHz)

(b)

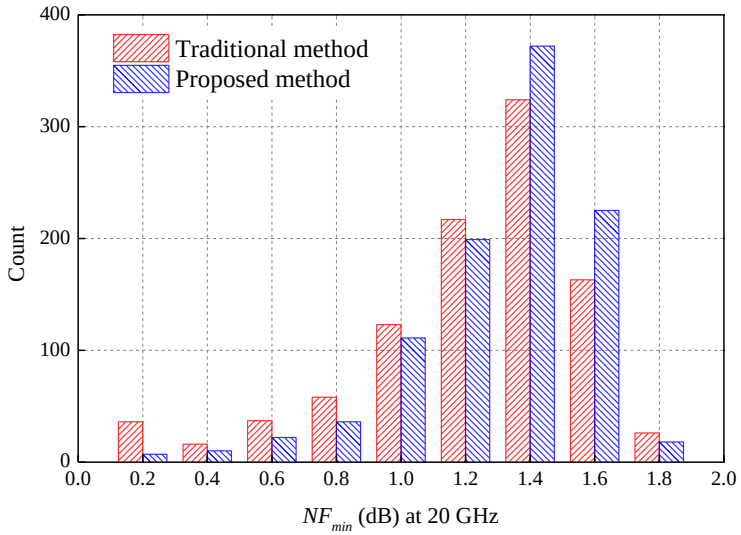


(c)

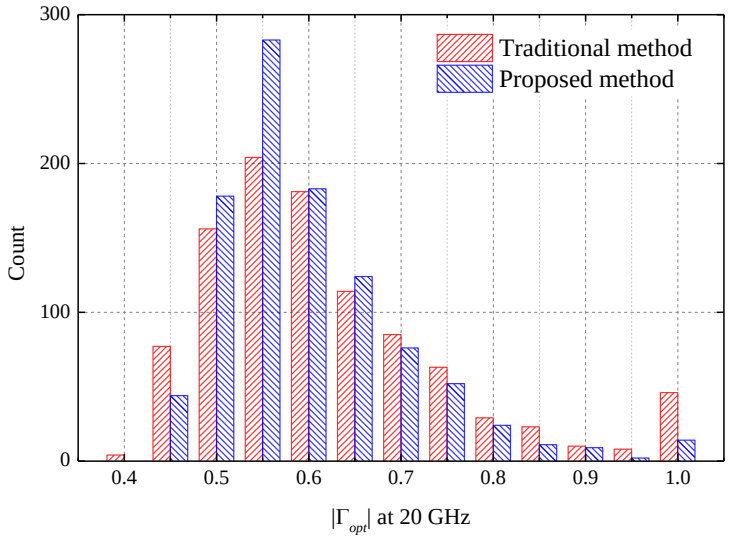


(d)

Fig. 2.14. The extrinsic and deembedded (or intrinsic) noise parameters, namely (a) NF_{min}, (b) R_n, (c) $|\Gamma_{opt}|$, and (d) $\angle\Gamma_{opt}$, versus frequency for an *n*-type 90-nm MOSFET calculated from around 200 measured extrinsic noise factors at each frequency. The error bars show the standard deviation [−σ, +σ] of the intrinsic noise parameters by randomly selecting 24 out of the ~200 noise factors each time and repeating the calculation 1000 times.



(a)

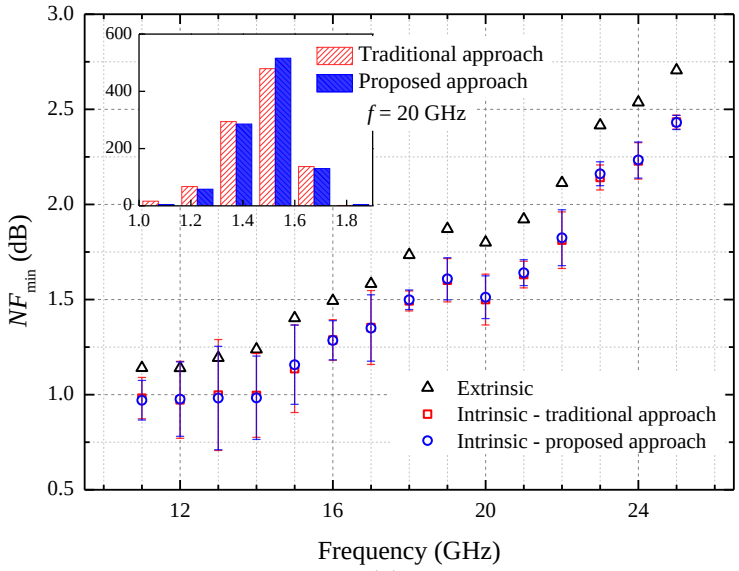


(b)

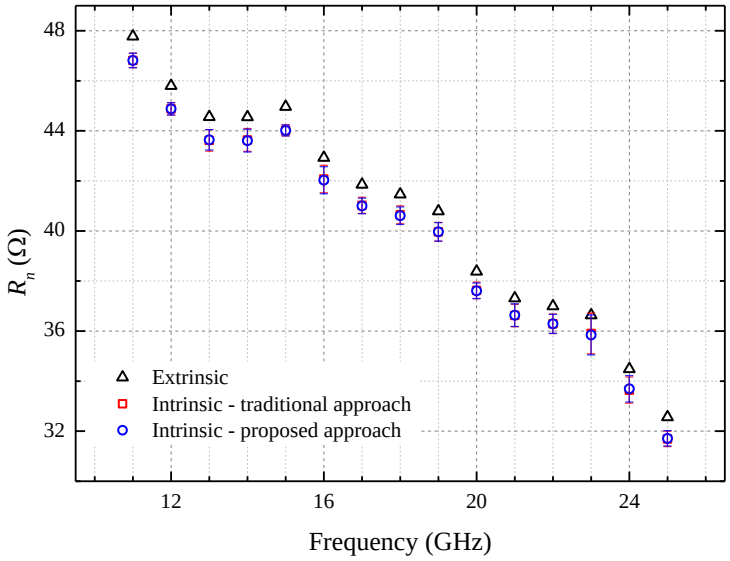
Fig. 2.15. Histograms of the 1000 results from previous random runs for (a) NF_{min} and (b) $|\Gamma_{opt}|$ at 20 GHz.

Fig. 2.15 shows a closer look at the meaning of the error bars using the histograms of (a) NF_{min} and (b) $|\Gamma_{opt}|$ for the 1000 results at 20 GHz. The proposed method converges better than the traditional method with smaller variations. In addition, it also produces less non-physical $|\Gamma_{opt}|$, i.e., $|\Gamma_{opt}| \approx 1$, than the traditional method, as shown in Fig. 2.15 (b). This suggests that the proposed optimization-last approach is statistically more robust and accurate for data with measurement errors or uncertainties.

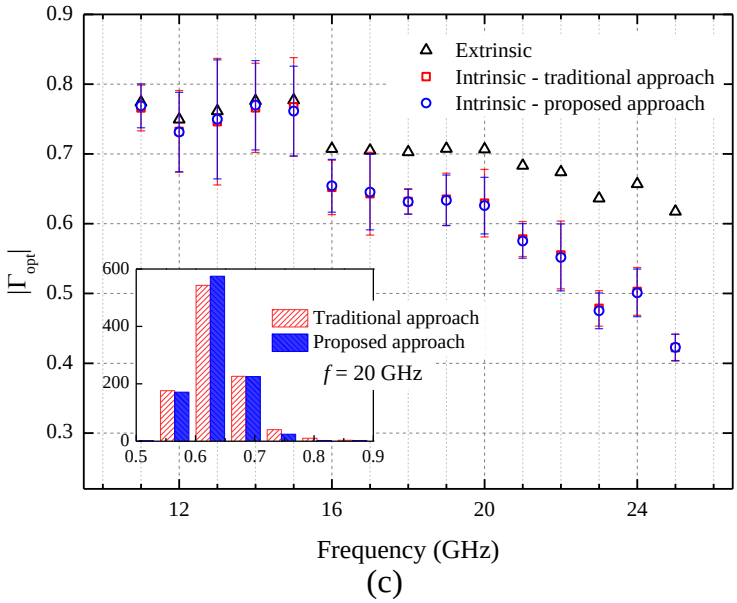
We evaluate the proposed optimization-last approach using another set of data which is measured on an *n*-MOSFET with W/L = 64×1 μm/28 nm fabricated in UMC's 28-nm CMOS technology. The DUT is biased at $V_{GS} = V_{DS}$ = 1.05 V with dc current I_{DS} = 34.5 mA. The noise factors are measured in the frequency range of 11 to 25 GHz and come from 61 calibrated source admittances at each frequency. By using the parasitics model in Fig. 2.8 and Lane's optimization method, we feed the measured data into the two approaches described above and plot the resulting noise parameters, i.e., (a) NF_{min}, (b) R_n, (c) $|\Gamma_{opt}|$, and (d) $\angle\Gamma_{opt}$, versus frequency in Fig. 2.16. The extrinsic and intrinsic noise parameters are obtained by the traditional approach using all these 61 measured extrinsic noise factors. The other set of intrinsic noise parameters is from our proposed optimization-last approach using the same data. By inspecting the results, we observe that for the final intrinsic noise parameters, the proposed approach gives very similar results compared to those obtained from the traditional approach. This demonstrates the effectiveness of our proposed noise factor deembedding algorithm and the optimization-last approach given large amount of measured data points.


3.0
2.5
2.0
1.5
1.0
0.5
NF_{min} (dB)
12
16
20
24
Frequency (GHz)
600
400
200
0
1.0
1.2
1.4
1.6
1.8
Traditional approach
Proposed approach
f = 20 GHz
Extrinsic
Intrinsic - traditional approach
Intrinsic - proposed approach

(a)


48
44
40
36
32
R_n (Ω)
12
16
20
24
Frequency (GHz)
Extrinsic
Intrinsic - traditional approach
Intrinsic - proposed approach

(b)



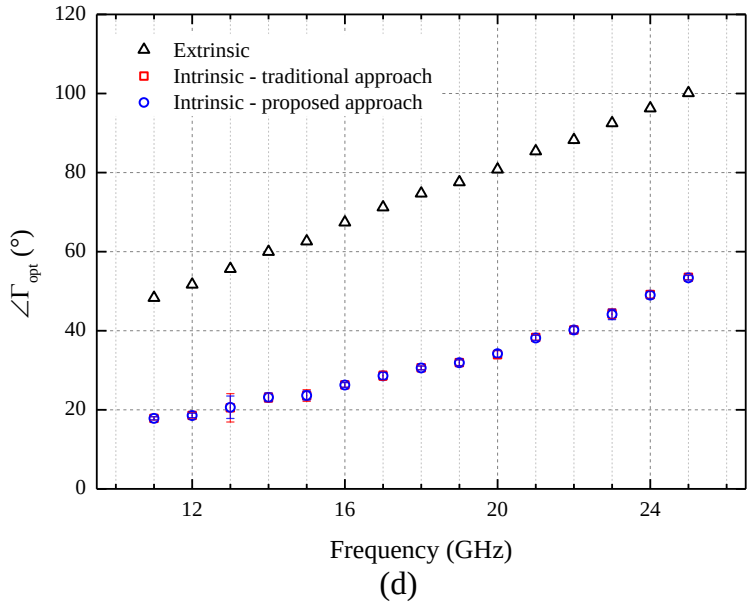


Fig. 2.16. The extrinsic and deembedded (or intrinsic) noise parameters, namely (a) NF_{min}, (b) R_n, (c) $|\Gamma_{opt}|$, and (d) $\angle\Gamma_{opt}$ for a 28-nm *n*-type MOSFET calculated from 61 measured extrinsic noise factors at each frequency. The error bars show the standard deviation [−σ, +σ] of the intrinsic noise parameters by randomly selecting 24 out of the 61 noise factors each time and repeating the calculation for 1000 times. The insets show the histograms of these 1000 results for (a) NF_{min} and (c) $|\Gamma_{opt}|$ at 20 GHz. (© 2017 IEEE.)

In practical situations, it is very time-consuming to measure large amount of data points at different tuner positions (in this case, 61), and therefore, only 20 ~ 30 source admittances are commonly used in practice. To simulate this real-world scenario, we have run a statistical evaluation of the two approaches by randomly selecting 24 out of these 61 measured data points each time and repeating the calculation for 1000 times. The results are also shown in Fig. 2.16 with the error bars showing the standard deviation [−σ, +σ] of the 1000 results at all the frequencies. The insets in Fig. 2.16 (a) and (c) are the histograms of the 1000 results for the intrinsic (a) NF_{min} and (c) $|\Gamma_{opt}|$ at 20 GHz. We observe that the optimization-last approach, in general, provides slightly smaller variations across these frequencies compared to the traditional one. The relatively small improvement in this particular design is expected because the deembedding, which is an arithmetic operation based on measured *S*-parameters, is not the decisive factor on the error bar sizes. Instead, they mainly depend on the design of the DUT, the quality of the measured noise factors, and the choice of the optimization algorithm which, in this experiment, are the same for both approaches.

To illustrate the impact of the proposed approach on different designs and see its ability of suppressing non-physical results, we perform another experiment by controlling the number of data points used in optimization and counting the non-physical occurrences in $|\Gamma_{opt}|$ out of 1000 random runs. We run these two approaches using both the 90-nm data (216 points in total) and the 28-nm data (61 points in total) at 20 GHz. We choose $|\Gamma_{opt}| > 0.95$ as the criteria for it to be considered as non-physical, comparing with the expected value of around 0.6 at this particular frequency. The resulted non-physical counts in $|\Gamma_{opt}|$

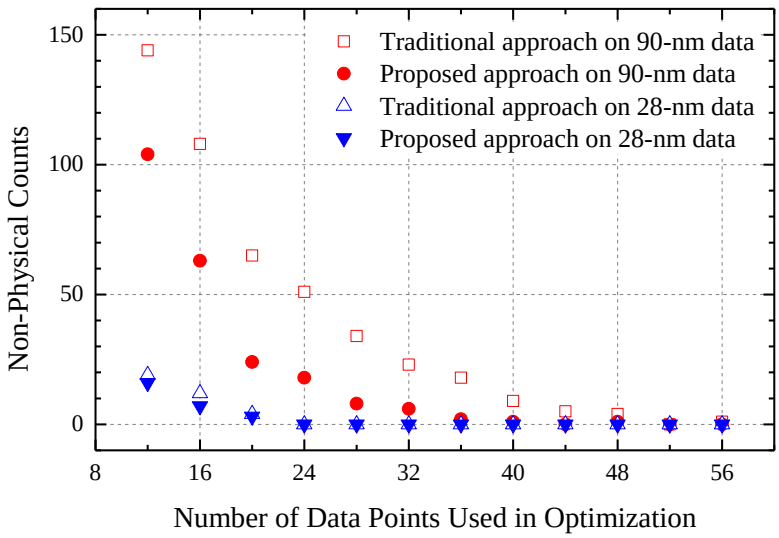


Fig. 2.17. Non-physical counts in $|\Gamma_{opt}|$ out of 1000 random runs versus the number of data points used in optimization at 20 GHz. We choose $|\Gamma_{opt}| > 0.95$ as the criteria for it to be considered as non-physical. We measured 216 data points in total for the 90-nm device and 61 for the 28-nm device. (© 2017 IEEE.)

versus the number of points used in optimization are plotted in Fig. 2.17. An interesting trend observed is that the non-physical counts decrease exponentially as we increase the number of data points used in optimization for both approaches. In addition, the proposed approach outperforms the traditional approach by resulting less non-physical $|\Gamma_{opt}|$ in both data sets when using less data points in optimization. The suppression of non-physical results is stronger for the 90-nm data than that of the 28-nm data because of bigger probe pads (larger parasitics) and smaller g_m (lower gain) in the 90-nm design. As a result, the parasitics have a stronger impact on the noise deembedding of the 90-nm data than that of the 28-nm data. Consequently, the proposed optimization-last approach, by its design, works more effectively for the low-gain DUT with large parasitics.

Finally, the improvement above in the proposed approach over the traditional approach is established only by exchanging the order of the optimization and the deembedding, which is not possible to achieve without the newly developed noise factor deembedding algorithm. Theoretical proof of the improvement and its relation with different optimization algorithms are still highly desired, but they are beyond the focus of this work.

2.5 Conclusion

In this chapter, we present a noise factor deembedding algorithm for an active device surrounded by a four-port parasitic network to complement the existing theories. For on-wafer noise measurements, it enables a new approach to obtain the intrinsic noise parameters by performing deembedding first and optimization last. We also use the optimization-last approach to prove the accuracy and effectiveness of the proposed noise factor deembedding algorithm. Statistical evaluation of the optimization-last approach gives slightly smaller standard deviation and less non-physical situations than the traditional optimization-first, deembedding-last approach when applied on experimental data of an *n*-type MOSFET fabricated in a 28-nm CMOS technology from UMC. The optimization-last approach, however, works more effectively for a DUT with low gain and large parasitics as shown in the 90-nm design. Further research on theoretical proof for the accuracy enhancement in the optimization-last approach is still highly desired.

Chapter 3.

Challenges in Device Noise Performance and Characterization

3.1 Introduction

The future of the semiconductor industry requires low-power and high-frequency solutions for wireless and mobile applications. One of the reasons is that economical demands are driving the industry to provide solutions with lower cost per function, more functions per chip area and better chip performance. Due to the high-level of integration and high unity-gain frequency in the order of hundreds of GHz, CMOS technology is still the most competitive candidate for future mobile applications. However, one of the major challenges faced by the CMOS technology is the high power density and its resulting heat dissipation problem from the chip. The solution to the problem is to reduce the power supply voltage V_{dd}, which has been done together with the scaling of the devices.

Due to the reduction of the power supply voltage in advanced nanoscale technologies, the noise generated by the MOSFET channel becomes not only the major noise source for high-frequency circuits working in the order of tens or hundreds of GHz, but also one of the most fundamental limitations in analog circuits (e.g., operational amplifiers [52],[53]),

mixed signal circuit (e.g., analog-to-digital converters [54],[55]), and even for digital circuits [56]. The 1/*f* noise is dominant at low frequency, but can be reduced or eliminated by offset-cancellation techniques such as auto-zero configuration [57] and amplifier chopping [58],[59]. It leaves the channel thermal noise, which is white noise with constant power spectral density across the frequencies of interest, as an important and unavoidable issue. Therefore, the characterization and modeling of channel thermal noise in nanoscale transistors become critically important.

3.2 Noise Sheet Resistance as a Figure of Merit

In recent years, characterization and modeling of the channel thermal noise of nanoscale MOSFETs have attracted a lot of attentions, particularly after the much higher-than-expected noise level reported by Jindal in 1985 [60]. Various short-channel effects caused by scaling, which have been included in the DC models of MOSFETs earlier, have been used to explain the excess noise above the long-channel prediction, such as the hot carrier effect [61] and the channel length modulation (CLM) effect [62]. On the other hand, the mobility degradation of the carriers in the channel, which is caused by the high electric field, can reduce the channel thermal noise as reported in [64].

For circuit applications, circuit designers not only concern about the absolute noise level, but also the signal-to-noise ratio (SNR). The previously mentioned short channel effects have an impact on the noise performance and the DC and AC performance of MOSFETs at the same time. Therefore, when comparing the noise performance of different technology nodes, it does not show the full picture if we only focus on the noise level of

the technologies. In light of this, a figure-of-merit (FoM), namely the noise sheet resistance R_{nsh}, has been proposed for proper evaluation of noise performance of different technologies, which is defined as [65],

$$R_{nsh} = \frac{S_{id}}{4kT_0 g_m^2} \cdot \left(\frac{W}{L} \right) \tag{3.1}$$

where S_{id} is the power spectral density (PSD) of the noise that appears in the drain current, W/L is the aspect ratio of the device, k is Boltzmann's constant, T_0 is the standard temperature 290 K, and g_m is the transconductance of the transistor.

The advantage of R_{nsh} is that the term $S_{id} / 4kT_0 g_m^2$ (in Ω) captures the characteristics of the noise S_{id} and the transconductance g_m simultaneously. In addition, it is equal to the portion of the equivalent noise resistance R_n, which is one of the four noise parameters of the intrinsic device, caused by S_{id} [66]. As mentioned in [65], because S_{id} is proportional to W/L and g_m^2 is proportional to $(W/L)^2$, we need to normalize this parameter by (W/L) in order to have a meaningful comparison for devices with different geometries. After the geometry normalization, R_{nsh} becomes a process-related parameter with the unit Ω/□, and is therefore called the noise sheet resistance. R_{nsh} depends on the physical parameters such as the effective mobility μ_{eff}, the critical electric field along the channel E_C, the threshold voltage V_T, and the gate-oxide capacitance C'_{ox} per unit area. When comparing two different technologies, higher R_{nsh} usually suggests higher R_n if the aspect ratio W/L of the device is kept the same. In other words, higher R_{nsh} suggests higher DC power consumption is required to have the same level of noise performance. With R_{nsh}, we can easily compare the

noise performance of devices fabricated in different processes, e.g., processes with strained Si or high-k dielectric and metal gate technologies. It should be noted that R_{nsh} does not necessarily determine the noise factor of the device. R_{nsh} is closely related to R_n which is one of the four noise parameters. Based on the noise factor expression (2.16), R_n indicates how much the noise factor would increase compared to the minimum noise factor $F_{\min}$ if the source admittance Y_S is different from the optimum source admittance Y_{opt}. Therefore, it is important to decrease R_n to achieve smaller noise factor when noise mismatch exists at the input of the device. In addition, different from R_{nsh}, R_n does not consider the geometry normalization and thus not suitable for comparison between technologies.

3.3 Experimental Results and Discussion

In this section, we use the noise sheet resistance R_{nsh} to evaluate the noise performance for MOSFETs of various lengths. The impact of different technologies that are used to advance the scaling of MOSFETs on noise performance are then analyzed.

3.3.1 R_{nsh} in the Long-Channel Limit

Before studying the impacts of various short-channel effect on channel thermal noise, it is simple and informative to study the R_{nsh} under the noise theory for long-channel MOSFETs. For MOSFETs working in strong inversion and saturation, the long-channel theory for the channel thermal noise gives [67]

$$S_{id} = \frac{8kT}{3} g_m = \frac{8kT}{3} \cdot \frac{W}{L} \mu_{eff} C'_{ox} (V_{GS} - V_{TH}) \tag{3.2}$$

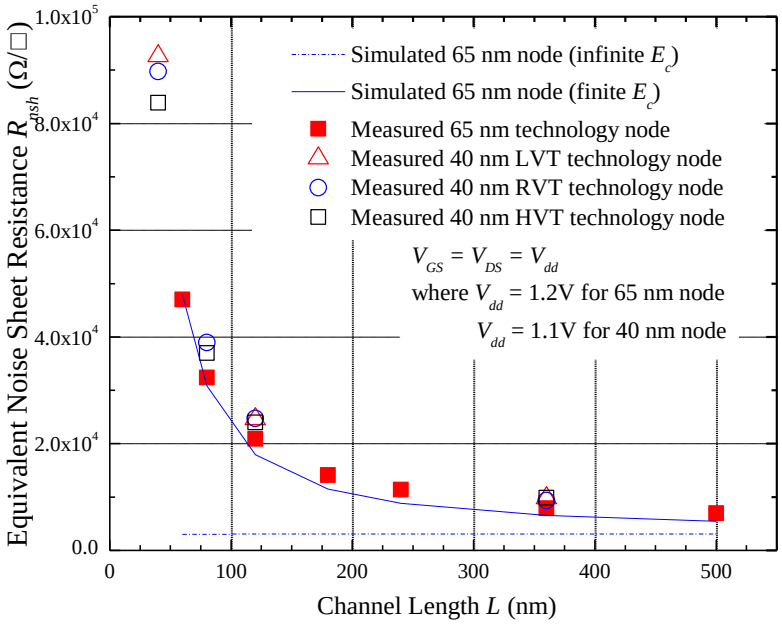


Fig. 3.1. Measured (symbols) and calculated (lines) R_{nsh} versus channel length L for n-type MOSFETs fabricated in 65-nm and 40-nm CMOS technology nodes biased at $V_{DS} = V_{GS} = V_{dd}$. The dashed line represents simulated results for the 65-nm node using (3.3) by assuming no mobility degradation. The solid line represents results following [65] by considering mobility degradation due to finite E_C. (© 2015 IEEE.)

where the body effect is ignored. Substituting (3.2) into (3.1), R_{nsh} can be expressed by

$$R_{nsh} = \frac{2}{3}\frac{T}{T_0 \mu_{eff} C'_{ox}(V_{GS} - V_{TH})}. \tag{3.3}$$

It suggests that for low-noise applications, R_{nsh} can be improved with materials having larger μ_{eff} or thinner oxide thickness (i.e., larger C'_{ox}). In addition, for the same technology, (3.3) shows no dependence on the channel length under long-channel theory.

3.3.2 Experimental R_{nsh} for Sub-100-nm MOSFETs

To compare these predictions with experimental data, we have conducted noise measurements on devices fabricated using UMC's 65 nm and 40 nm technology nodes, respectively. Both technology nodes have introduced strain engineering for mobility enhancement. All of the device-under-tests (DUTs) are in multi-finger structure with their total widths being $M \times N_f \times W_f$ μm. Here M is the number of multi-finger structures connected in parallel, N_f is the number of fingers in a multi-finger structure, and W_f is the finger width of each transistor. After proper deembedding of the pad parasitics [44], we followed the procedures in [68] to extract the experimental channel noise S_{id} and transconductance g_m to calculate the experimental R_{nsh}, which are shown as symbols in Fig. 3.1. Long-channel prediction by (3.3) is shown as the dashed line in Fig. 3.1. It can be seen that the experimental results which are directly calculated from definition (3.1) increase with shorter channel length, whereas the long-channel prediction fails to capture the trend. This discrepancy between the long-channel prediction and the experimental data can be explained by various short-channel effects that have an impact on noise performance.

3.3.3 Impact of Different Technologies on Noise Performance

It is important to understand the physics that causes the increasing trend of R_{nsh} in the short-channel transistors, particularly for the process engineers to improve the technologies and for the circuit designers to select the proper technology for low-noise applications.

We first study the mobility degradation effect, which is caused by the increasing lateral electric field in the MOSFET channel as it becomes smaller while the saturation velocity stays almost constant [71]. A practical model for this effect is given by the following equation for the drift velocity of the carriers [69]

$$v_{drift} = \begin{cases} \dfrac{\mu_{eff}}{1 + E / E_C} E & \text{for } E \le E_C \\ v_{sat} & \text{for } E > E_C \end{cases} \tag{3.4}$$

where μ_{eff} is the low-field effective mobility without considering the degradation effect, E_C is the critical electric field, and v_{sat} is the saturation velocity. The latter two are related by $E_C = 2v_{sat} / \mu_{eff}$ [70]. When using the strained Si in devices, the effective mobility μ_{eff} can be enhanced by about 2 times, but the saturation velocity does not change. In this case, E_C is actually reduced by a half. As has been studied in [65], for short-channel devices, the reduction in the critical field E_C could explain the increase in R_{nsh} when the channel length decreases, which is also shown in Fig. 3.1. Consequently, engineering techniques to improve the effective mobility without increasing the saturation velocity at the same time actually degrades the noise performance of nanoscale MOSFETs. We can foresee that the devices made by materials with higher mobility but lower saturation velocity (such as the strained Ge in [72] and [73]) could result in worse noise performance.

Another commonly used technique to boost the device mobility utilizes different channel plane orientations, e.g., (110) for *p*-channel transistors [74]. As shown in [65], to maintain the same noise performance for the doubled effective mobility, we need to improve the saturation velocity by about 10 times. In reality, the doubled hole mobility

using stressors in the (110) direction only improves the saturation velocity by about 2 times [75]. Materials with higher saturation velocity such as InSb or $In_{0.7}Ga_{0.3}As$ in [76] might be the potential solution for the low-noise technologies in the future.

On the other hand, based on (3.3), the noise sheet resistance can also be reduced by increasing C'_{ox} through the reduction of the oxide thickness or the use of high-k materials. This is preferable comparing to enhancing the saturation velocity since the latter is much harder to realize. It should be noted, the intrinsic minimum noise factor does not depend on C'_{ox} [77] and therefore the noise factor of the device may not be improved much by reducing C'_{ox} if the noise matching is good.

Finally, the impact of the power supply reduction can be seen from the comparison of R_{nsh} for the 40-nm-technology devices which are biased at $V_{DS} = V_{GS} = V_{dd} = 1.1$ V and the 65-nm-technology devices which are biased at $V_{DS} = V_{GS} = V_{dd} = 1.2$ V. With lower power supply voltage, the 40-nm technology node presents larger R_{nsh} that suggests worse noise performance. This makes future low-noise design even more challenging with reduced power supply and channel length when using nanoscale MOSFETs for low-power applications.

3.4 Recent Technology Trend

To verify the predictions described in the previous section, we can calculate the equivalent sheet resistance R_{nsh} using published data on current CMOS technologies and technologies that are in the research and development stage. The devices include 20 nm

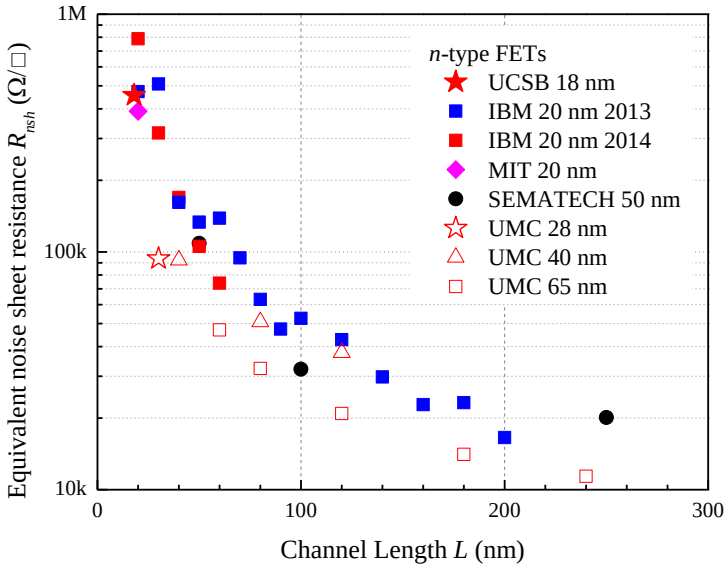


Fig. 3.2. R_{nsh} versus channel length for *n*-type FETs using measured R_{on}, R_{ext}, and peak g_m. The data sources are from UCSB 18 nm [82], IBM 20 nm in year 2013 [78] and 2014 [79], MIT 20 nm [80], SEMATECH 50 nm [81], and UMC's 28 nm [85], 40nm [84], and 65 nm [65] technology nodes. (© 2015 IEEE.)

$In_{0.53}Ga_{0.47}As$ MOSFETs from IBM [78],[79], 20 nm $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ quantum-well (QW) MOSFETs from MIT [80], 50 nm Extremely-Thin-Body(ETB) InAs QW MOSFETs from SEMATECH [81], 18nm InAs QW MOSFETs from UCSB [82], and 40 nm Ge *p*-type MOSFETs from TSMC [83]. UMC's 65 nm CMOS [65], 40 nm CMOS [84], and 28nm CMOS technology nodes [85] are the benchmarks in these comparisons.

To evaluate R_{nsh} with limited experimental data from the literature, we approximate the channel noise as the thermal noise generated by the channel resistance at $V_{DS} = 0$ V, i.e., $S_{id} = 4kT/R_{ch0}$, which is the highest level predicted by the long-channel theory. The actual channel noise, which differs from the approximated value by a factor of the noise

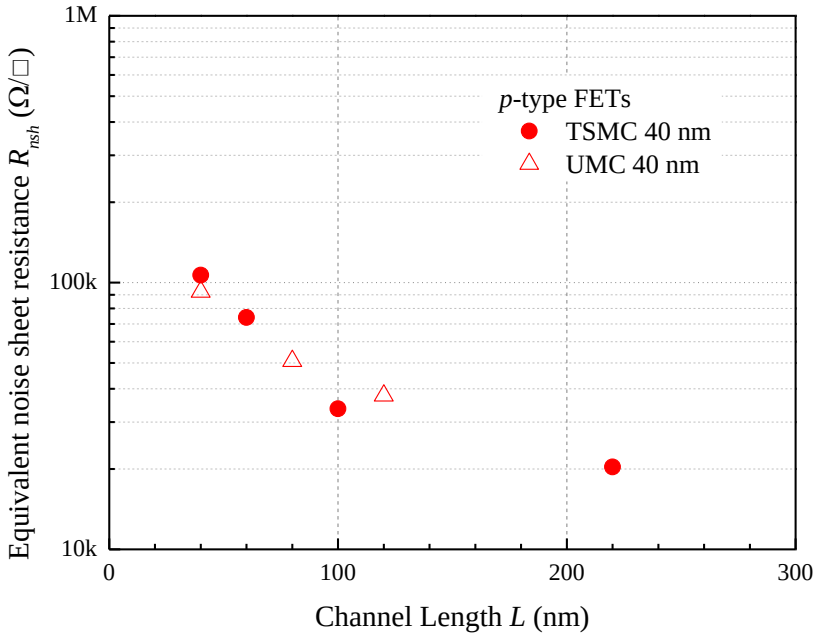


Fig. 3.3. R_{nsh} versus channel length for *p*-type FETs using measured R_{on}, R_{ext}, and peak g_m. The data sources are from TSMC 40 nm Ge FinFET [83] and UMC 40 nm CMOS [84] technology nodes. (© 2015 IEEE.)

coefficient γ [1], is usually not measured and reported in the above mentioned papers. Moreover, since the channel resistance R_{ch0} is usually not presented directly, we evaluate it as $R_{ch0} = R_{on} - R_{ext}$, where R_{on} is the on-resistance at $V_{DS} = 50$ mV and R_{ext} is the extracted external resistance. In addition, we approximate the transconductance g_m using its peak value at the same V_{GS} under which the channel resistance R_{ch0} is extracted. Based on these approximations, we are able to conduct a relatively fair comparison of the noise performance across different modern technologies using R_{nsh}. Therefore, we can rewrite (3.1) as

$$R_{nsh} = \frac{T}{T_0 R_{ch0} g_{m,peak}^2} \cdot \left(\frac{W}{L}\right) \tag{3.5}$$

where $g_{m,peak}$ is the highest transconductance reported for the technology and T is the ambient temperature in Kelvin. The evaluated results for *n*-type FETs and *p*-type FETs are plotted in Fig. 3.2 and Fig. 3.3, respectively. They confirm the prediction that channel engineering using III-V and Ge materials in the channel does degrade the noise performance of many future technology nodes. In addition, as shown in Fig. 3.2, although the g_m for the process in [79] is enhanced, its R_{nsh} still degrades comparing to the old process in [78] for devices with channel length of 20 nm. This agrees with the prediction that to maintain the noise performance of nanoscale devices, the saturation velocity of the carriers needs to be enhanced simultaneously with the effective mobility.

In conclusion, the noise sheet resistance can serve as a figure of merit for the process engineers to predict the noise performance of a technology. Channel engineering techniques need to be able to improve the saturation velocity more than the effective mobility to improve the device's noise performance. In addition, increasing C'_{ox} by reducing the effective gate oxide thickness could reduce the noise sheet resistance of the device and thus could be a promising direction for future low-noise technologies.

3.5 Calibration of Noise Receiver with Gain Variations

For transistors working at high frequencies with reduced power consumption, the impact of the channel thermal noise on the circuit performance plays an increasingly important role as described in the previous sections. Due to the enhancement of the carrier mobility using strained Si or III-V materials in the device channel without increasing the carrier saturation velocity, the reduction in the critical electrical field E_C increases the

thermal noise and results in higher minimum noise figure $NF_{\min}$ in 32 nm node [87], and 3 times higher noise-to-signal ratio in recent devices with feature size smaller than 30 nm [85],[88]. Therefore, accurate characterization of the thermal noise at high frequencies attracts much attention in recent years.

In high-frequency noise measurement, the calibration of the noise system plays a critical role in measurement accuracy. To improve the calibration accuracy, there are two main issues to be considered, namely the difference in the output reflection coefficients $\Delta\Gamma_{ns}$ of the noise source between the "hot" and "cold" states, and the impedance matching between the noise source and its load [89]. Tiemeijer *et al.* proposed an improved *Y*-factor method by calculating the effective *Y*-factor and the effective Excess Noise Ratio (ENR) [90]. Chen *et al*. resolved this issue by proposing a power-equation based method [91].

In this section, we investigate the impact of the difference in noise source reflection coefficients $\Delta\Gamma_{ns}$ on the accuracy of the extracted gain-bandwidth constant *kBG* and the measured noise factors and noise parameters. In addition, the issue of the *kBG* dependence on the source impedance will be explored and a method to mitigate this issue be suggested.

3.5.1 Fundamentals of Noise Receiver Calibration

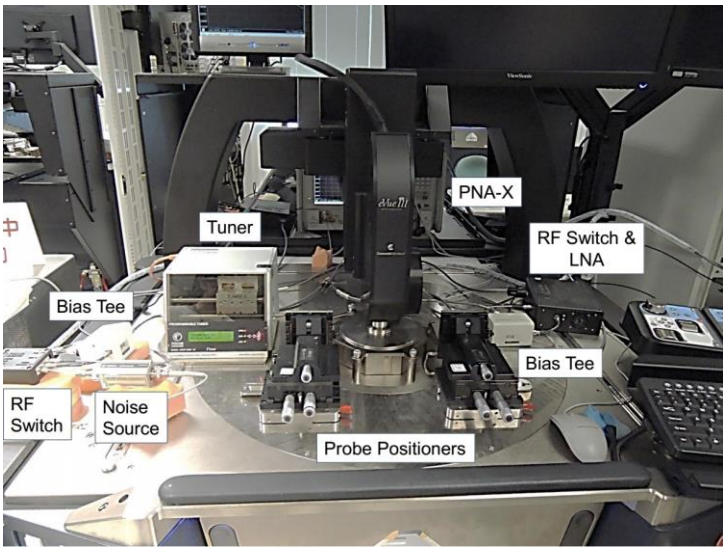
A noise receiver has to be calibrated before it can be used to measure the noise factors of the device-under-test (DUT). Based on the "cold-only" method [92], the noise factor measured by the noise receiver at any given source admittance is given by

$$F_R = \frac{P_C}{T_0 kBG}\frac{|1-\Gamma_R\Gamma_S|^2}{(1-|\Gamma_S|^2)} - \frac{T_C}{T_0} + 1 \tag{3.6}$$

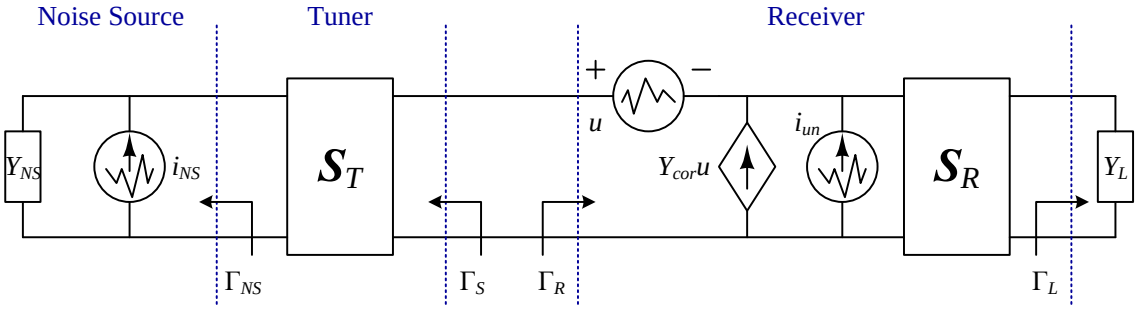
where P_C is the measured noise power at the cold state (i.e., the noise source is turned off), Γ_S is the source reflection coefficient seen by the noise receiver, Γ_R is the input reflection coefficient of the noise receiver, T_C is the ambient temperature, T_0 is the standard temperature (290 K), k is Boltzmann's constant, B is the bandwidth, and G is the equivalent gain of the receiver. Here the product kBG is known as the gain-bandwidth constant of the receiver and has the unit of $W \cdot K^{-1}$.

An example experimental setup of a noise measurement system from Focus Microwaves is shown in Fig. 3.4 (a). It consists of a noise source, a PNA-X microwave network analyzer with noise measurement capability, a microwave tuner, a low-noise amplifier (LNA), and other peripheral components like microwave switches and controllers. The noise measurement system can be modeled using the schematic shown in Fig. 3.4 (b) in which a noise receiver is connected to a noise source through a tuner [91]. In the schematic, the noisy receiver is simplified as a noiseless network with equivalent noise sources, i.e., voltage source u, correlated current source $Y_{cor}u$ and uncorrelated current source i_{un}. To use the "cold-only" method, the kBG constant of the receiver needs to be found out through a calibration process utilizing both the "cold" and "hot" states of the noise source.

Now we will derive fundamental equations on which the calibration process is based. The noise power detected by the receiver can be written as [91]



(a)



(b)

Fig. 3.4. (a) Experiment setup of a noise measurement system which consists of a noise source, a tuner, a network analyzer, an LNA, and other peripheral components like microwave switches and controllers. (© 2017 IEEE.) (b) Schematic of a noise receiver connected to a noise source via a tuner [91]. The noisy receiver is simplified as a noiseless network with equivalent noise sources, i.e., voltage source u, correlated current source $Y_{cor}u$ and uncorrelated current source i_{un}.

$$P_n = G_{TR} P_{AVS} \tag{3.7}$$

where G_{TR} is the transducer gain and P_{AVS} is the available noise power from the source. The available source power can be calculated by assuming the source is connected to a load with an impedance that is complex conjugate of the source impedance. The available noise power P_{AVS} consists of the noise power from the source which has an equivalent noise temperature T_S, and the equivalent noise power generated by the noisy receiver itself which is contributed by the noise voltage source u, the correlated noise current source $Y_{cor}u$, and the uncorrelated noise current source i_{un} shown in Fig. 3.4 (b). Thus the available source power can be expressed as [91]

$$\begin{aligned} P_{AVS} &= \frac{1}{4G_S}\left(\left|Y_S + Y_{cor}\right|^2 \overline{u^2} + \overline{i_{un}^2} + 4kT_S G_S\right) B \\ &= \frac{1}{4G_S}\left(\left|Y_S + Y_{cor}\right|^2 4kT_0 R_n + 4kT_0 G_{un} + 4kT_S G_S\right) B \\ &= kBT_0\left(\left|Y_S + Y_{cor}\right|^2 \frac{R_n}{G_S} + \frac{G_{un}}{G_S} + \frac{T_S}{T_0}\right) \end{aligned} \tag{3.8}$$

where Y_S is the source admittance, G_S is the real part of Y_S, T_0 is the standard temperature 290 K, and R_n and G_{un} are the equivalent noise resistance and conductance of noise sources u and i_{un}, respectively. On the other hand, the transducer gain of the receiver can be written as [91]

$$G_{TR} = \frac{1-\left|\Gamma_S\right|^2}{\left|1-\Gamma_R\Gamma_S\right|^2}\left|S_{21}\right|^2 \frac{1-\left|\Gamma_L\right|^2}{\left|1-S_{22}\Gamma_L\right|^2} = \frac{1-\left|\Gamma_S\right|^2}{\left|1-\Gamma_R\Gamma_S\right|^2} G \tag{3.9}$$

where G is a parameter that is not a function of the source admittance and is given by

$$G = \left|S_{R,21}\right|^2 \frac{1-\left|\Gamma_L\right|^2}{\left|1-S_{R,22}\Gamma_L\right|^2}. \tag{3.10}$$

Therefore, by combining (3.8) and (3.9), the noise power received by the receiver can be expressed as

$$P_n = kBGT_0 \frac{1-\left|\Gamma_S\right|^2}{\left|1-\Gamma_R\Gamma_S\right|^2}\left(\left|Y_S + Y_{cor}\right|^2 \frac{R_n}{G_S} + \frac{G_{un}}{G_S} + \frac{T_S}{T_0}\right) \tag{3.11}$$

in which the *kBG* is known as the gain-bandwidth constant of the noise receiver.

In order to find out *kBG*, we need to utilize the two states of the noise source, i.e., "hot" state with admittance Y_{SH} and equivalent noise temperature T_{SH}, and "cold" state with Y_{SC} and T_{SC}. Based on (3.11), the noise power under the two states can be written as

$$P_H = kBGT_0 \frac{1-\left|\Gamma_{SH}\right|^2}{\left|1-\Gamma_R\Gamma_{SH}\right|^2}\left(\left|Y_{SH} + Y_{cor}\right|^2 \frac{R_n}{G_{SH}} + \frac{G_{un}}{G_{SH}} + \frac{T_{SH}}{T_0}\right) \tag{3.12}$$

and

$$P_C = kBGT \frac{1-\left|\Gamma_{SC}\right|^2}{\left|1-\Gamma_R\Gamma_{SC}\right|^2}\left(\left|Y_{SC} + Y_{cor}\right|^2 \frac{R_n}{G_{SC}} + \frac{G_{un}}{G_{SC}} + \frac{T_{SC}}{T_0}\right). \tag{3.13}$$

Traditionally, the source admittances of the "hot" and "cold" states are assumed to be the same, i.e., $\Gamma_{SC} \approx \Gamma_{SH} \approx \Gamma_S$, thus the gain-bandwidth constant can be simplified as [92]

$$kBG \approx \frac{P_H - P_C}{T_H - T_C} \frac{\left|1-\Gamma_R\Gamma_S\right|^2}{1-\left|\Gamma_S\right|^2}. \tag{3.14}$$

However, as shown in the experimental results in the next subsection, this assumption is not usually correct. In this case, the determination of *kBG* becomes complicated but is

resolved in [91] using an iteration method. It is then found that the following equation gives quite accurate result of the *kBG* compared to the iterated value, i.e.,

$$kBG \approx \frac{1}{T_{SH}G_{SH} - T_{SC}G_{SC}}\left(P_H G_{SH} \frac{\left|1-\Gamma_R\Gamma_{SH}\right|^2}{1-\left|\Gamma_{SH}\right|^2} - P_C G_{SC} \frac{\left|1-\Gamma_R\Gamma_{SC}\right|^2}{1-\left|\Gamma_{SC}\right|^2} \right). \quad (3.15)$$

To differentiate the two cases described in (3.14) and (3.15), we refer to the difference in the noise source admittance as $\Delta\Gamma_{ns}$, *kBG* given by (3.14) as the uncompensated *kBG* and the one given by (3.15) as the compensated *kBG*.

3.5.2 Impact of Receiver Gain Variations in Noise Measurements

First, we study the impact of $\Delta\Gamma_{ns}$ on the extracted *kBG* values using experimental data from two commercial noise sources, namely Agilent 346C and Noisecom NC346KA. The noise system includes a noise source, a mechanical tuner CCMT-2607 (0.7 – 26 GHz) from Focus Microwaves Inc., a low-noise amplifier (LNA) from Miteq (AMF-4D-00101800-24-10P), and an Agilent PNA-X N5242A. The uncompensated *kBG* values (kBG_{uncom}) are extracted using (3.14) and the compensated *kBG* values (kBG_{com}) using (3.15). The normalized *kBG*, (i.e., kBG_{uncom} / kBG_{com} × 100%) values are plotted in Fig. 3.5 for the two noise sources. It can be observed that the normalized *kBG* values from Agilent 346C vary from 95.6% to 103.7% and those from Noisecom NC346KA have a larger variation from 94.9% to 108.1%.

We then investigate the impact of the *kBG* variation on noise factors and noise parameters. Using the tuner, we can vary the source admittance seen by the receiver, which

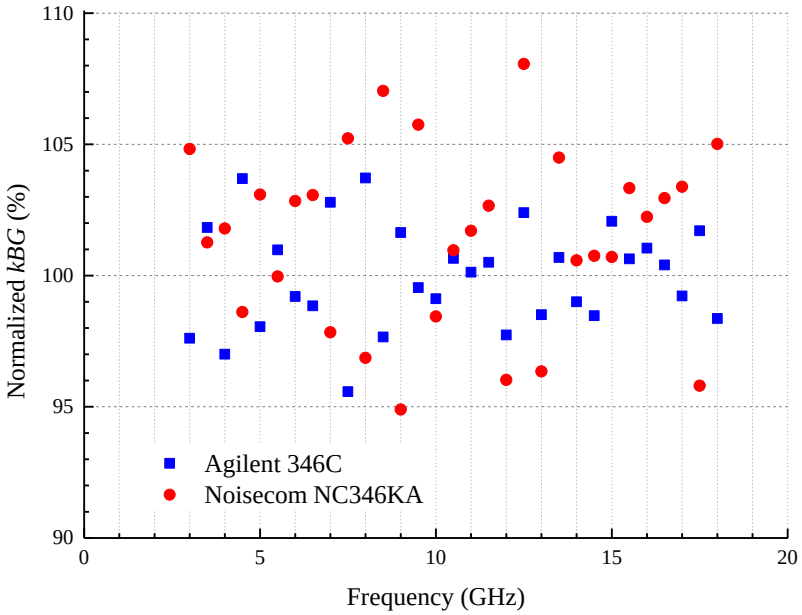
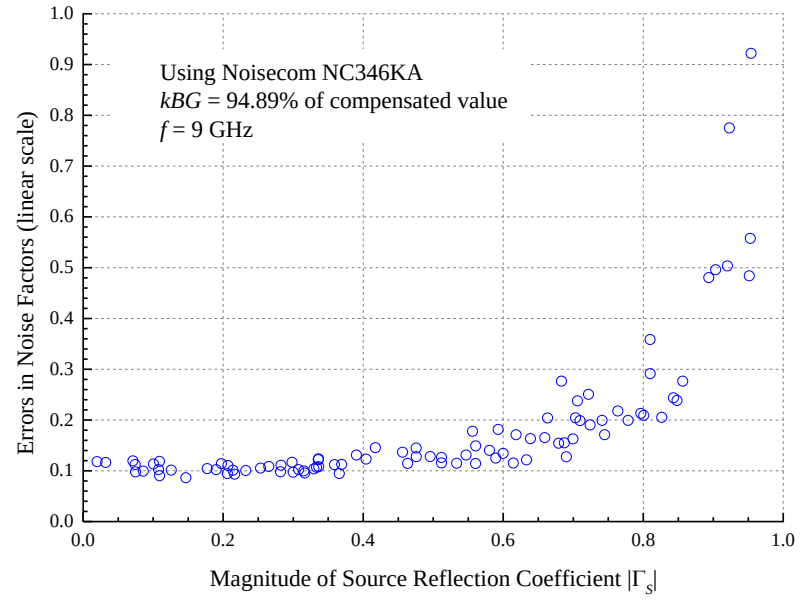
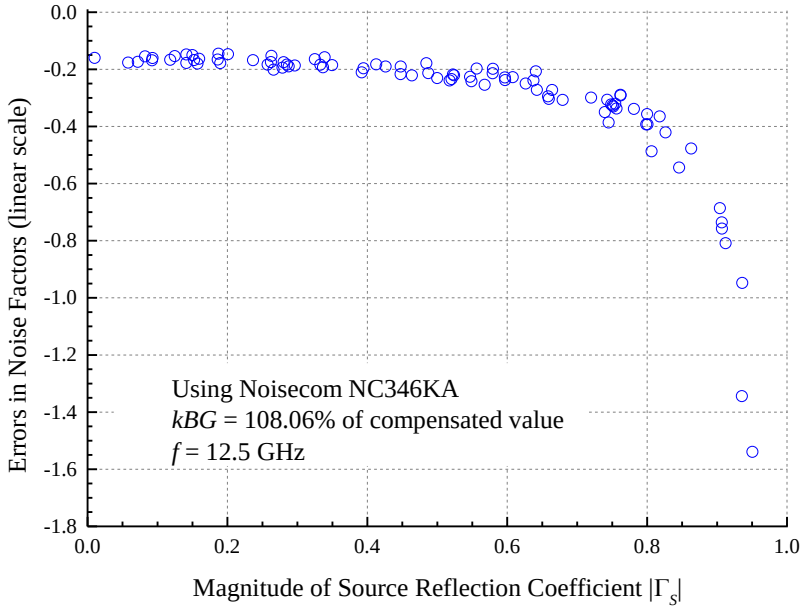


Fig. 3.5. Normalized *kBG* (to its compensated values obtained at the initial tuner positions) of a noise receiver versus frequency characteristics using Agilent 346C and Noisecom NC346KA. (© 2015 IEEE.)

is an essential step in noise measurements to achieve multiple noise factors and thus the noise parameters. In Fig. 3.6, the errors in noise factors using uncompensated *kBG* compared to those using compensated *kBG*, i.e., $F_{uncom} - F_{com}$, at a number of different tuner positions are plotted. Based on the noise factor equation (3.6) from cold-only method, the calculated noise factor is inversely proportional to *kBG*. In addition, since Γ_R is usually close to zero, the error in noise factors caused by *kBG* variation will be higher when the magnitude of the source reflection coefficients $|\Gamma_S|$ is higher. This trend is confirmed in Fig. 3.6 with errors in noise factors when (a) kBG_{uncom} = 94.9% of the compensated value and (b) kBG_{uncom} = 108.1% of the compensated value using the Noisecom device. A smaller uncompensated *kBG* results in larger calculated noise factor as shown in Fig. 3.6 (a) and a bigger uncompensated *kBG* does the opposite in Fig. 3.6 (b).

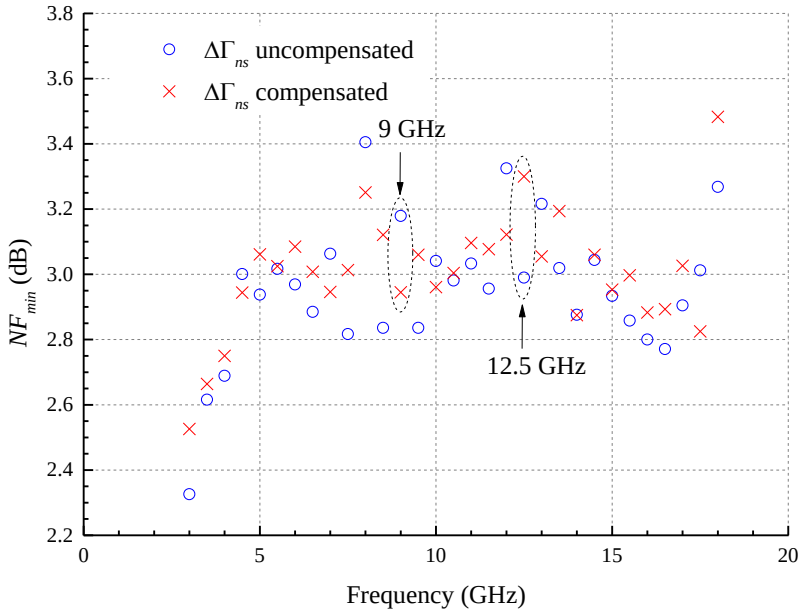


(a)

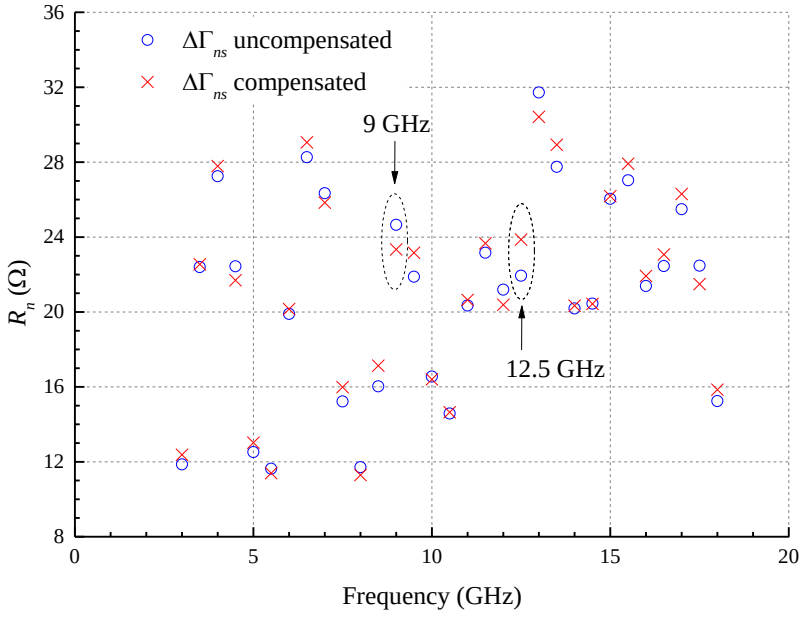


(b)

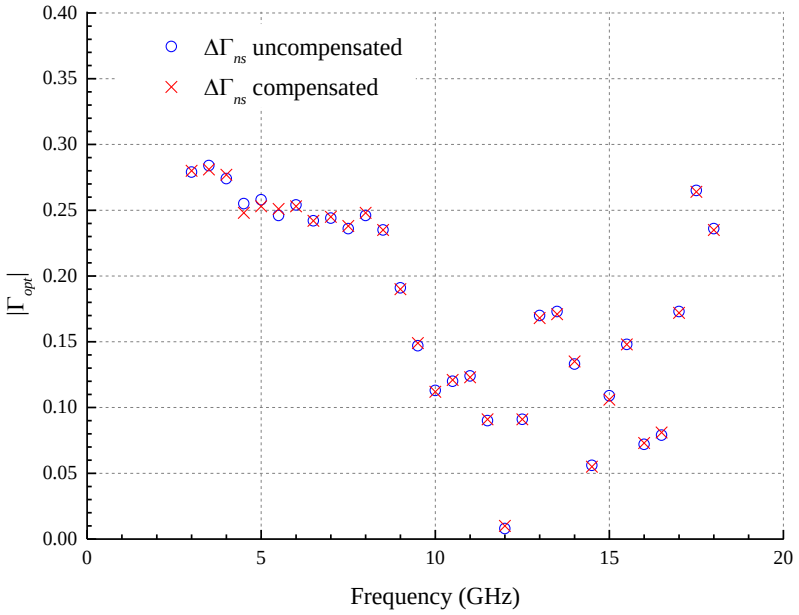
Fig. 3.6. Errors in the calculated noise factors (linear scale) of the noise receiver at different source reflection coefficients Γ_S resulted from (a) *kBG* = 94.9% of the compensated value at 9 GHz and (b) *kBG* = 108.6% of the compensated value. (© 2015 IEEE.)


ΔΓns uncompensated
ΔΓns compensated
9 GHz
12.5 GHz
NFmin (dB)
Frequency (GHz)
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
0
5
10
15
20

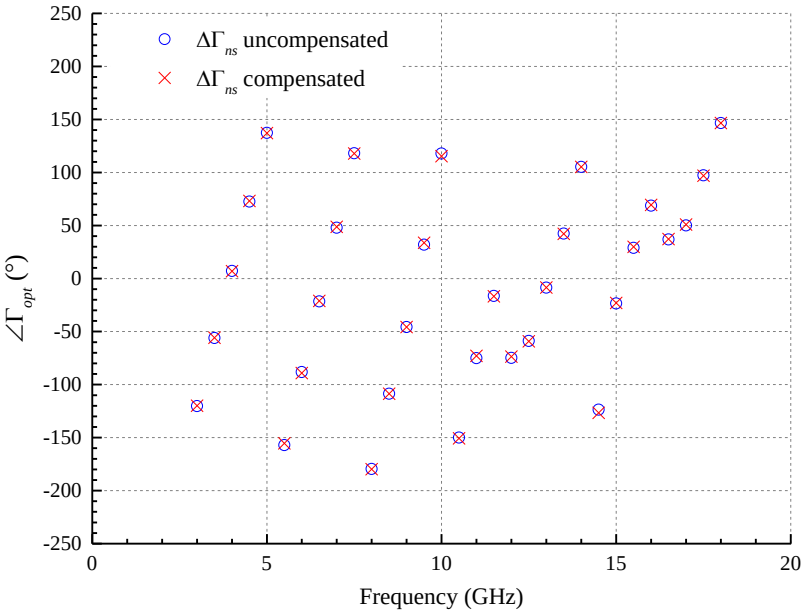
(a)


ΔΓns uncompensated
ΔΓns compensated
9 GHz
12.5 GHz
Rn (Ω)
Frequency (GHz)
36
32
28
24
20
16
12
8
0
5
10
15
20

(b)



(c)



(d)

Fig. 3.7. (a) Minimum noise figure $NF_{\min}$, (b) equivalent noise resistance R_n, (c) magnitude of the optimized source reflection coefficient $|\Gamma_{opt}|$, and (d) angle of the optimized source reflection coefficient $\angle\Gamma_{opt}$ of the noise receiver versus frequency characteristics with and without $\Delta\Gamma_{ns}$ compensation. (© 2015 IEEE.)

The four noise parameters, namely (a) minimum noise figure $NF_{\min}$, (b) equivalent noise resistance R_n, (c) magnitude of the optimized source reflection coefficient $|\Gamma_{opt}|$, and (d) angle of the optimized source reflection coefficient $\angle\Gamma_{opt}$ are also calculated based on both the compensated and uncompensated *kBG*. The resulting two sets of noise parameters versus frequency are plotted in Fig. 3.7 in the range of 3 to 18 GHz. Both the magnitude and angle of Γ_{opt} have very weak dependence on the variations in *kBG* whereas NF_{min} and R_n are much more sensitive to them. At 9 GHz when kBG_{uncom} = 94.9% of the compensated value, the uncompensated NF_{min} and R_n are both larger than the compensated values. On the other hand, the opposite happens to NF_{min} and R_n at 12.5 GHz when kBG_{uncom} = 108.1% of the compensated values.

3.5.3 Dependence of Receiver Gain on Source Impedance

By changing the tuner position, the source admittance Γ_S seen by the receiver can be changed. The gain-bandwidth constant *kBG* can be calculated at different tuner positions to see whether or not dependence exists. Fig. 3.8 (a) shows relative *kBG* values obtained at different Γ_S normalized to the value at initial tuner position at the frequency of 8 GHz using the noise soruce Agilent 346C. The normalized *kBG* values vary in a range from 84.2% to 110.5% at $|\Gamma_S| = 0.9$. The range reduces to from 88.6% to 102.7% when the $|\Gamma_S|$ decreases to 0.5. The variations also cause the variations in the noise factors NF_{50} for a 50-Ω source impedance as shown in Fig. 3.8 (b). As described in [91], in theory, the receiver gain should not depend on the source impedance at which the receiver gain was extracted. However, it is not clear from [91] what exactly causes the *kBG* dependence on the source impedance.

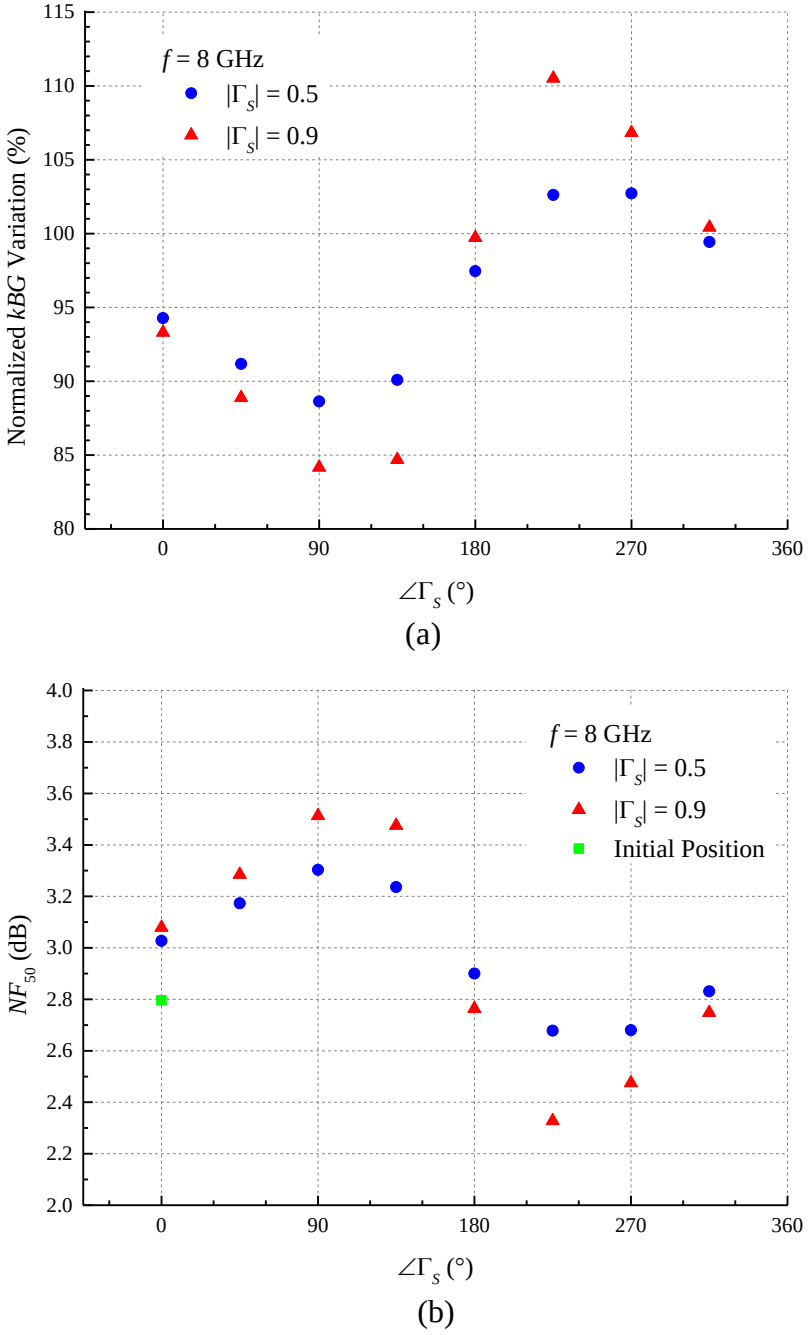
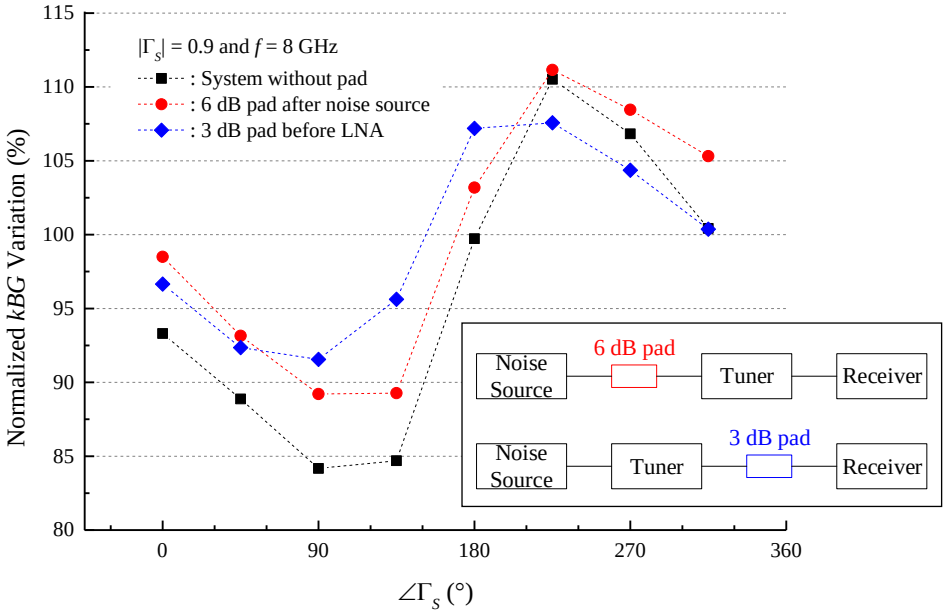


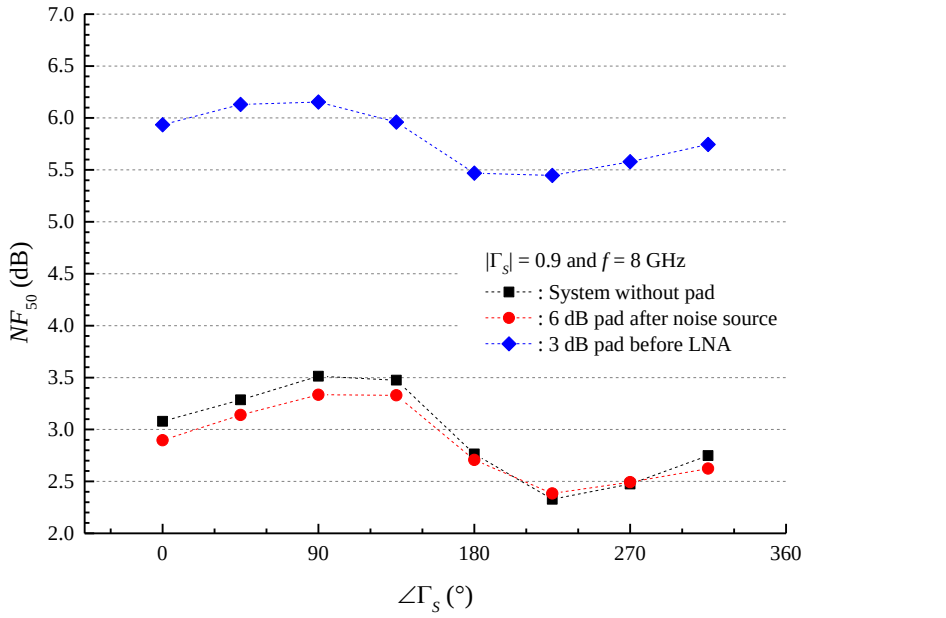
Fig. 3.8. (a) Normalized *kBG* variations obtained at different Γ_S (f = 8 GHz) and (b) its resulting noise figure NF_{50} for a 50 Ω source impedance. (© 2015 IEEE.)

To investigate the cause of the *kBG* dependence on Γ_S, we experimented with two different system configurations by inserting (a) a 6 dB attenuator (or pad) between the noise source and the input of the source tuner, and (b) a 3 dB attenuator between the output of the source tuner and the LNA. Fig. 3.9 (a) shows the normalized *kBG* values at 8 GHz for the noise system calibration without attenuator, with 3 dB, and 6 dB attenuators, respectively. We notice that the configuration with 3 dB pad added before the LNA has the smallest *kBG* variation. However, as shown in Fig. 3.9 (b), this configuration results in the highest NF_{50} and therefore is not suitable for low noise measurement. On the other hand, the configuration with a 6 dB pad also has smaller *kBG* variation compared with the original system and results in similar NF_{50}, which is therefore suggested for future system configuration.

To understand the improvement of *kBG* variation with pads added into the system, we need to go back to the definition of the Excess Noise Ratio (ENR). As mentioned in [93], the hot noise temperature T_H, calculated from the ENR table, is obtained when the noise source is connected to a 50-Ω termination. In addition, the impedance mismatch between the input impedance seen by the noise source and 50 Ω during the calibration results in power reflection and different effective hot noise temperature [89]. It also suggests placing an attenuator (or an isolator) between the noise source and the DUT, which can prevent the noise power reflections from reaching the noise source where they reflect again and combine with the incident signal [89]. Nevertheless, a 6 dB attenuator increases the return loss by 12 dB and reduces ENR by 6 dB at large bandwidth, whereas an isolator increases the return loss by 20 dB, reduces ENR due to its insertion loss, and limits the bandwidth to



(a)



(b)

Fig. 3.9. (a) Normalized *kBG* at 8 GHz for a noise system without attenuator (or pad), with a 3 dB attenuator, and a 6 dB attenuator, respectively. The inset shows the schematics of the latter two configurations. (b) The corresponding NF_{50} at 8 GHz for these three cases. (© 2015 IEEE.)

one octave or less. Therefore, we suggest that the gain-bandwidth constant *kBG* of a noise receiver should be extracted at the tuner position where the noise source sees a 50-Ω termination, which is the condition at which the noise source's ENR has been specified.

3.6 Conclusion

In this chapter, we use the noise sheet resistance as a figure of merit for the process engineers to predict the noise performance of a technology. Generally, future technologies, including planar MOSFETs and III-V quantum well FETs, present worse noise performance than the conventional longer channel CMOS technology. It is found that channel engineering techniques that can improve the effective mobility of the carriers but cannot improve the saturation velocity at the same time will degrade the device's noise performance. On the other hand, a promising solution for future low-noise technologies is to increase C_{ox} by reducing the effective gate oxide thickness.

In addition, two accuracy issues are explored in the calibration of the noise receiver for high-frequency noise measurements. The difference in the reflection coefficients of the noise source at "hot" and "cold" states, if not accounted for, has a large impact on the extracted gain-bandwidth constant *kBG*, and thus the measured noise factors and noise parameters. Mismatch between the noise source and the load impedance seen at the input of the tuner causes the *kBG* dependence on the Γ_S. Adding an attenuator/isolator between the noise source and the input of the tuner reduces the *kBG* variation because of a better matching to the 50 Ω at which the ENR values were calibrated.

Chapter 4.

Extraction of Gate and LDD Resistance

4.1 Introduction

Nanoscale CMOS technology has already become a potential candidate for sub-THz applications due to the aggressive scaling in the feature size and the resulting very high cutoff frequency (f_T) and maximum oscillation frequency (f_{max}) in the order of hundreds of GHz [94]. As predicted in the 2012 International Technology Roadmap for Semiconductors (ITRS) [95], multi-gate MOSFETs with 15.3 nm gate length can achieve 710 GHz and 622 GHz for f_T and f_{max}, respectively. However, when working at such high frequencies and aiming at low-power applications, the noise from the transistor itself becomes critically important and therefore accurate noise modeling is required.

Apart from the channel noise, induced gate noise and their correlations, the noise from the gate resistance is a major noise contributor because its effect is amplified by the transistor itself to the next stage. Therefore, accurate extraction of the gate resistance from experiments becomes very important for low-noise, low-power applications. In addition, the accuracy of the extracted gate resistance has a great impact on the extraction of the channel noise, induced gate noise and their correlations [68], the evaluation of noise

performance between technologies using the equivalent noise sheet resistance [96], the investigation of device reliability [97]-[99], and the noise modeling of devices [100]-[102].

Due to the gate-to-source and the gate-to-drain capacitances, the gate resistance cannot be directly characterized from the *I-V* measurements. As a result, *Z*- or *Y*-parameters obtained at high frequencies have been used to extract the gate resistance [103]-[114]. For the existing *Z*-parameter based approaches [103],[110],[114], in order to remove the impact from the frequency-dependent term (e.g., $A_g/(\omega^2+B)$ in [103]), *Z*-parameters measured at very high frequencies (e.g., 35 – 40 GHz) are needed [114], which puts a very high demand on the equipment and the accuracy of the equivalent circuit model.

On the other hand, at high frequencies, the total gate resistance R_g seen at the gate terminal not only consists of the gate contact resistance R_{gcon} and the distributed poly-silicon gate resistance R_{gpoly}, but also the distributed channel resistance R_{ch} [104],[108]. When extracting the gate resistance at such high frequencies, the non-quasi-static (NQS) effect becomes very pronounced and causes the distributed channel resistance R_{ch} to be the dominant contributor in the extracted gate resistance [108]. The separation of R_{ch} from R_{gcon} and R_{gpoly} is not only important in accurately describing the RF behavior of MOSFETs [113], but also crucial in the thermal noise modeling. Since the thermal noise in the channel is already taken care of by another noise current source [100], R_{ch} should be modeled as a noise free resistor. If R_{ch} is included in R_g as a noisy resistor and used in the channel noise extraction, it would result in an underestimation in the channel noise characterization.

In this chapter, we will introduce a new *Z*-parameter based approach to extract the gate resistance at low frequencies. Since the low-frequency extrapolation is applied in the

algorithm, we can use a simple small-signal equivalent circuit for medium-high frequencies in [115], and the high-frequency components such as C_m, C_{mb}, C_{mx}, and C_{sd} in [114],[115] can all be ignored. This low-frequency approach can not only reduce the complexity of the equivalent circuit without losing its generality and accuracy, but also simplify the corresponding analytical *Z*-parameter expressions. In addition, it prevents the impact from the substrate coupled into the experimental results. Finally, we evaluate the R_g extracted from experimental data on devices fabricated in different processes and explore its dependence on the channel length.

4.2 Existing R_g Extraction Methods

Due to the gate-to-source and gate-to-drain capacitances, it is not possible to directly characterize the gate resistance R_g from DC *I-V* measurements. Therefore, admittance (*Y*) or impedance (*Z*) parameters at high frequencies are usually required to characterize R_g. A gate resistance extraction algorithm depends on the small-signal model of the DUT at the chosen bias condition and frequency range of the measurements. Variation of those conditions can lead to different small-signal models and R_g equations. The literature has already provided a range of R_g extraction methods, which are either based on *Y*-parameters or *Z*-parameters of the DUT.

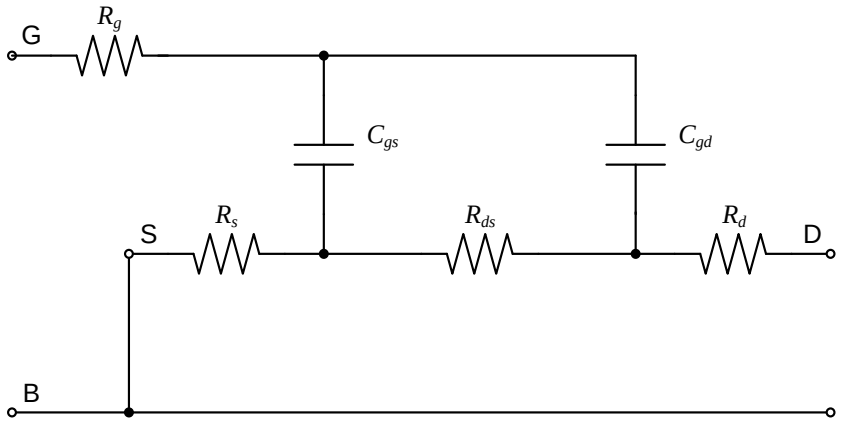


Fig. 4.1. Small-signal equivalent circuit for the MOSFET operating in linear region with V_{DS} = 0 V based on Jen *et al*. [105]. At V_{DS} = 0 V, the device is symmetric so that C_{gs} is approximately equal to C_{gd}. Y_{11} and Y_{21} are then calculated to find the expression for R_g. (© 1999 IEEE.)

4.2.1 *Y*-parameter-based Approaches for R_g Extraction

Jen *et al*. has provided an equation for R_g extraction based on the Y_{11} and Y_{12} of the MOSFET operating in linear region with V_{DS} = 0 V [105]. The small signal circuit provided by them for calculating Y_{11} and Y_{12} is redrawn in Fig. 4.1. Several assumptions are made so that the small-signal circuit and the derivations are valid, including 1) the substrate network can be ignored under the given bias, and 2) the relevant operation frequency is up to around 10 GHz, resulting the following simplifications:

$$(\omega C_{gs} R_s)^2 \ll 1, \tag{4.1}$$

$$(\omega C_{gd} R_d)^2 \ll 1, \tag{4.2}$$

and

$$\omega C_{gg} R_g \ll 1 \tag{4.3}$$

where C_{gg} is the total gate capacitance and $C_{gg} \cong C_{gs} + C_{gd} + C_{gb}$. Using the simplifications, the Y_{11} and Y_{12} of the small-signal circuit in Fig. 4.1 can be derived as

$$Y_{11} \cong j\omega C_{gg} + \omega^2 (C_{gg}^2 R_g + C_{gs}^2 R_s + C_{gd}^2 R_d) \tag{4.4}$$

and

$$Y_{12} \cong -j\omega C_{gd} - \omega^2 C_{gd} C_{gg} R_g . \tag{4.5}$$

Gate resistance R_g can then be determined as

$$R_g = \frac{\mathrm{Re}(Y_{12})}{\mathrm{Im}(Y_{11}) \cdot \left|\mathrm{Im}(Y_{12})\right|}. \tag{4.6}$$

Comparing to the original equation in Jen *et al.*'s paper, the outer modulus sign is removed since it is intentional created to avoid negative results.

On the other hand, Enz and Cheng's method [106] for R_g extraction only depends on Y_{11}. By neglecting the terminal resistances and also assuming $\omega R_g C_{gg} \ll 1$, the Y_{11} can be expressed by the following form

$$Y_{11} \cong \omega^2 R_g C_{gg}^2 + j\omega C_{gg} \tag{4.7}$$

and therefore R_g can be expressed as

$$R_g = \frac{\mathrm{Re}(Y_{11})}{\mathrm{Im}(Y_{11})^2}. \tag{4.8}$$

In addition, the condition of V_{DS} = 0 V is not required in their work.

In another work by Kang *et al.* [113], the distributed gate resistance R_{elec} is directly extracted. Under low frequency assumption, they express R_{elec} as

$$R_{elec} = \frac{C_{jd}^2 \,\mathrm{Re}(Y_{11}) - C_{gb}^2 \,\mathrm{Re}(Y_{22})}{\omega^2 \left[(2C_{gs0} + C_{gb})^2 C_{jd}^2 - C_{gd0}^2 C_{gb}^2 \right]} \tag{4.9}$$

where C_{gb} is the gate-to-body capacitance, C_{jd} is the drain junction capacitance, C_{gs0} and C_{gd0} are the gate-to-source and gate-to-drain capacitances at zero bias, respectively. Due to the symmetry of MOSFET, C_{gs0} should be approximately equal to C_{gd0} if the lateral bias V_{DS} = 0 V. When V_{GS} equals to the power supply voltage $V_{dd,}$ we can observe from their experimental results that $C_{gb} \approx$ 0 fF, and the imaginary part of Y_{12} can be expressed as

$$\mathrm{Im}(Y_{12}) \cong -\omega C_{gd0}. \tag{4.10}$$

Using (4.10), we can rewrite (4.9) in the following simplified form as

$$R_g = R_{elec} = \frac{\mathrm{Re}(Y_{11})}{4 \cdot \mathrm{Im}(Y_{12})^2}. \tag{4.11}$$

4.2.2 *Z*-parameter-based Approaches for R_g Extraction

Lee *et al.* has proposed a R_g extraction method based on Z-parameters [109],[110]. The small-signal equivalent circuit is shown in Fig. 4.2, in which the dotted box represents the intrinsic part of the device. The *Y*-parameters of the intrinsic part is fairly straightforward, which are given by

$$Y_{11} = j\omega(C_{gs} + C_{gd}), \tag{4.12}$$

$$Y_{12} = -j\omega C_{gd}, \tag{4.13}$$

$$Y_{21} = g_m - j\omega C_{gd}, \tag{4.14}$$

and

$$Y_{22} = g_{ds} + j\omega(C_{ds} + C_{gd}). \tag{4.15}$$

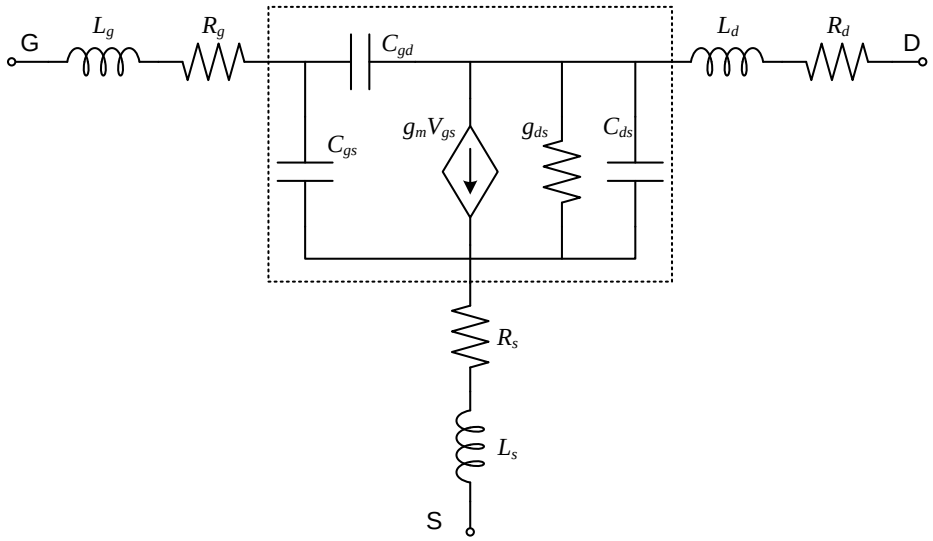


Fig. 4.2. A small-signal equivalent circuit for a Si MOSFET based on Lee *et al.* [109]. The dotted box shows the intrinsic part of the device. (© 1997 IEEE.)

Then *Z*-parameters of the intrinsic part can be directly converted from them. Since the external inductance and resistance are in series with the intrinsic part, they can be directly added to the *Z*-parameters. The resulted Z_{11} and Z_{12} are

$$Z_{11} = R_g + R_s + j\omega(L_g + L_s) + \frac{g_{ds} - j\omega(C_{gd} + C_{ds})}{D} \tag{4.16}$$

and

$$Z_{12} = R_s + j\omega L_s + \frac{j\omega C_{gd}}{D} \tag{4.17}$$

where $D = -\omega^2(C_{gs}C_{ds} + C_{gs}C_{gd} + C_{gd}C_{ds}) + j\omega[g_m C_{gd} + g_{ds}(C_{gs} + C_{gd})]$. Therefore, the gate resistance R_g can be expressed as [109]

$$R_g = \mathrm{Re}(Z_{11} - Z_{12}) - \frac{A_g}{\omega^2 + B} \tag{4.18}$$

where A_g and B are two parameters with complicated expression

$$A_g = \frac{C_{ds}\left[g_{ds}(C_{gs}+C_{gd})+g_m C_{gd}\right]}{(C_{gs}C_{ds}+C_{gs}C_{gd}+C_{gd}C_{ds})^2} - \frac{g_{ds}}{(C_{gs}C_{ds}+C_{gs}C_{gd}+C_{gd}C_{ds})} \tag{4.19}$$

and

$$B = \left[\frac{g_{ds}(C_{gs}+C_{gd})+g_m C_{gd}}{(C_{gs}C_{ds}+C_{gs}C_{gd}+C_{gd}C_{ds})}\right]^2. \tag{4.20}$$

Fortunately, at very high frequency up to 35-40 GHz, the second term can be ignored and R_g can be approximated as

$$R_g \approx \mathrm{Re}(Z_{11}-Z_{12})\big|_{30-40\,\mathrm{GHz}}. \tag{4.21}$$

Another Z-parameter-based approach is from Torres-Torres *et al.* [111]. The small-signal equivalent circuit is shown in Fig. 4.3, which is for a MOSFET under zero V_{DS} with negligible transconductance. The real parts of the *Z*-parameters are given as [111]

$$\mathrm{Re}(Z_{11}) = R_g + \frac{R_i}{2} + \frac{R_{ch}}{4} + R_s, \tag{4.22}$$

$$\mathrm{Re}(Z_{12}) = \mathrm{Re}\left(Z_{21}\right) = \frac{R_{ch}}{2} + R_s, \tag{4.23}$$

and

$$\mathrm{Re}(Z_{22}) = R_{ch} + R_d + R_s \tag{4.24}$$

where R_s and R_d are the source and drain resistances, respectively, and R_i is the resistance that accounts for the distributed effect of R_{ch}. Torres-Torres *et al.* found that $R_i = R_{ch}/6$, which is inconsistent with its reference where $R_i = R_{ch}/3$ [116], and rewrote the real part of Z_{11} as

$$\mathrm{Re}(Z_{11}) = R_g + \frac{R_{ch}}{3} + R_s. \tag{4.25}$$

Considering that R_d and R_s is much smaller than R_{ch}, R_g was approximated as

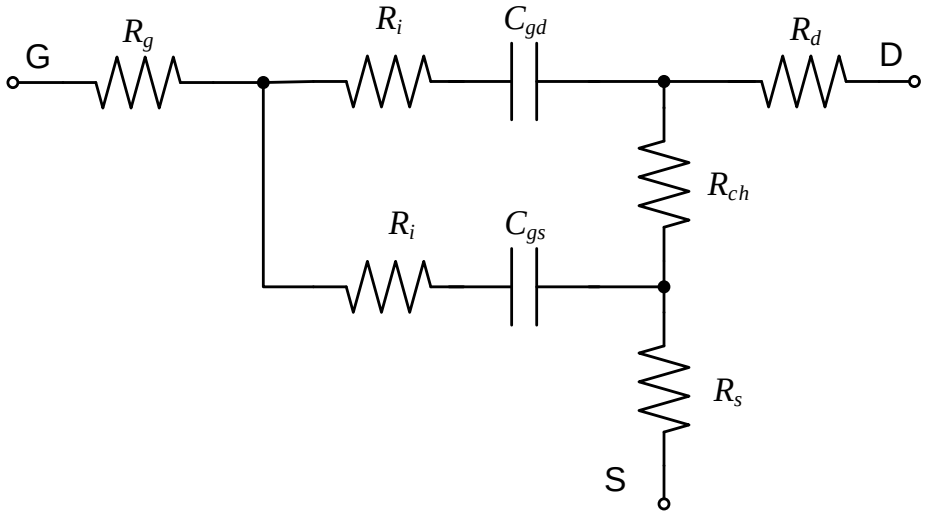


Fig. 4.3. A small-signal equivalent circuit of a MOSFET biased at V_{DS} = 0V based on Torres-Torres *et al.* [111]. R_i is the resistance that accounts for the distributed effect of the channel resistance R_{ch}. (© 2003 IEEE.)

$$R_g \cong \mathrm{Re}(Z_{11}) - \frac{\mathrm{Re}(Z_{22})}{3}. \tag{4.26}$$

It should be noted that, here the distributed effect is modeled by R_i and therefore R_g itself does not account for this effect. In addition, the accuracy of this method is affected by the assumption of the value of R_i.

4.3 Proposed *Z*-parameter-based R_g Extraction Method

In the proposed approach, the device-under-test (DUT) is biased at the $V_{GS} = V_{dd}$ (nominal supply voltage) and V_{DS} = 0 V. By setting $V_{GS} = V_{dd}$, the coupling between the gate and the substrate can be eliminated since the gate-to-substrate capacitance $C_{gb} \approx 0$ fF [113]. Under such bias condition, the small-signal equivalent circuit of the DUT at medium

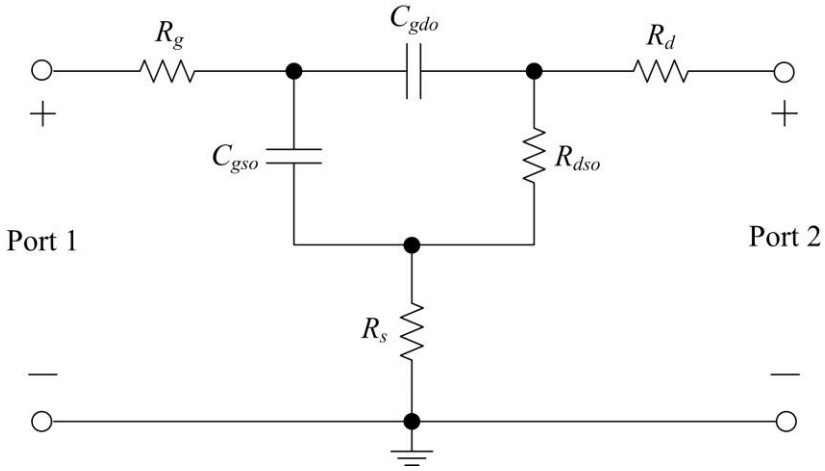


Fig. 4.4. Equivalent circuit for a MOSFET biased at $V_{GS} = V_{dd}$ (nominal supply voltage) and $V_{DS} = 0$ V for medium-high frequency applications. (© 2014 IEEE.)

high frequencies is shown in Fig. 4.4. For the resistances, R_g is the gate resistance to be extracted, R_d and R_s are the parasitic resistances at the source and drain terminals, respectively, R_{dso} consists of the channel resistance R_{ch} and the resistance of the lightly doped drain (LDD) regions. For the capacitances, C_{gso} and C_{gdo} are the gate-to-source and gate-to-drain capacitances, respectively, both of which consist of the intrinsic portion from the oxide capacitance (C_{ox}/2) [115] and the extrinsic portion from the overlap capacitance C_{ov}. Due to the symmetry of the DUT at $V_{DS} = 0$ V, the two capacitances are equal and here we define $C_{oxv} = C_{gso} = C_{gdo}$. Its value can be directly extracted from the measured imaginary part of the Y_{11}, i.e., Im(Y_{11}), versus frequency characteristics [106]. Similarly, the symmetry also leads to the equality of R_d and R_s, i.e., $R_d = R_s$.

Based on the equivalent circuit in Fig. 4.4, its Z-parameters can be expressed as

$$\begin{aligned} Z_{11} &= R_g + R_s + \frac{R_{dso}}{\omega^2 R_{dso}^2 C_{oxv}^2 + 4} - j\frac{\omega^2 R_{dso}^2 C_{ovx}^2 + 2}{\omega C_{oxv}(\omega^2 R_{dso}^2 C_{oxv}^2 + 4)} \\ &\cong \left(R_g + R_s + \frac{R_{dso}}{4} \right) - j\left(\frac{1}{2\omega C_{oxv}} \right), \end{aligned} \tag{4.27}$$

$$\begin{aligned} Z_{12} = Z_{21} &= R_s + \frac{2R_{dso}}{\omega^2 R_{dso}^2 C_{oxv}^2 + 4} - j\frac{\omega R_{dso}^2 C_{oxv}}{\omega^2 R_{dso}^2 C_{oxv}^2 + 4} \\ &\cong \left(R_s + \frac{R_{dso}}{2} \right) - j\left(\frac{\omega R_{dso}^2 C_{oxv}}{4} \right), \end{aligned} \tag{4.28}$$

and

$$\begin{aligned} Z_{22} &= R_d + R_s + \frac{4R_{dso}}{\omega^2 R_{dso}^2 C_{oxv}^2 + 4} - j\frac{2\omega R_{dso}^2 C_{oxv}}{\omega^2 R_{dso}^2 C_{oxv}^2 + 4} \\ &\cong \left(R_d + R_s + R_{dso} \right) - j\left(\frac{\omega R_{dso}^2 C_{oxv}}{2} \right). \end{aligned} \tag{4.29}$$

In the derivation, the only assumption made is that the frequency is low enough such that $\omega R_{dso} C_{oxv} \ll \sqrt{2}$ remains valid. If we define the critical frequency $\omega_c = 2\pi f_c$ so that $\omega_c R_{dso} C_{oxv} = \sqrt{2}/10$, then the critical frequency f_c, up to which (4.27) to (4.29) are valid, can be expressed as

$$f_c = (\frac{1}{10}) \cdot \frac{1}{\sqrt{2}\pi R_{dso} C_{oxv}}. \tag{4.30}$$

For the technology studied in this paper, f_c is in the range of tens of GHz if the channel length is shorter than 240 nm for *n*-type MOSFETs and 120 nm for *p*-type MOSFETs. As indicated in (4.30), f_c will become higher when the channel length decreases because of the reduction in R_{dso} and C_{oxv}.

For mature CMOS technology, it is always engineered to make the parasitic resistance of the drain R_d and source R_s much smaller than the intrinsic channel resistance R_{dso}, so that

the applied drain-source voltage V_{DS} mainly drops on the channel itself. In practice, the ratio between R_{dso} and R_d / R_s is about 100. Under the condition $R_{dso} \gg R_{d(s)}$, we can further simplify the expressions for Z_{11} and Z_{22} to

$$Z_{11} \cong \left(R_g + \frac{R_{dso}}{4} \right) - j\left(\frac{1}{2\omega C_{oxv}} \right) \tag{4.31}$$

and

$$Z_{22} \cong R_{dso} - j\left(\frac{\omega R_{dso}^2 C_{oxv}}{2} \right). \tag{4.32}$$

Combining the above two equations, we can arrive at the equation to extract R_g using only Z_{11} and Z_{22}, i.e.,

$$R_g = \mathrm{Re}(Z_{11}) - \frac{\mathrm{Re}(Z_{22})}{4} \tag{4.33}$$

where Re(·) denotes the real part of the parameters.

4.4 Experimental Results and Discussions

With the R_g expression (4.33), only Z_{11} and Z_{22} at medium-low frequency are required to extract R_g. In the following, we will show how it works on experimental data. The accuracy of the small-signal model that the method is based on will also be evaluated using Z-parameters. Finally, the proposed method will be compared with other published methods in a statistical evaluation using large amount of measured data.

4.4.1 Extraction of the Gate Resistance

The DUTs used in the work are in multi-finger structures, which means the total channel width W can be expressed as $M \times N_f \times W_f$ μm, where M is the number of transistors

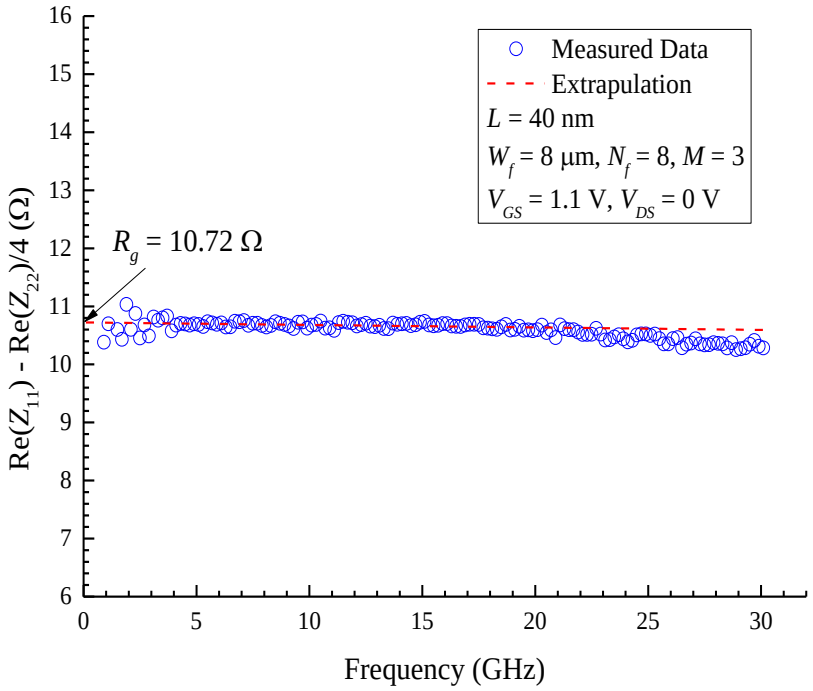


Fig. 4.5. $\mathrm{Re}(Z_{11}) - \mathrm{Re}(Z_{22})/4$ vs. frequency characteristics for an *n*-type FET with L=40 nm, W_f=8 µm, N_f=8, and M = 3 biased at V_{GS}= 1.1 V and V_{DS}= 0 V. (© 2014 IEEE.)

connected in parallel in a DUT, N_f is the number of transistor fingers, and W_f is the finger width of each transistor. In this measurement, the DUT is an *n*-type FET with L = 40 nm, W_f = 8 µm, N_f = 8, and M = 3 biased at V_{GS} = 1.1 V and V_{DS} = 0 V. Since the small-signal model in Fig. 4.4 only represents the intrinsic device, the measured Z-parameters have to undergo deembedding first to remove the effects of the pad parasitics.

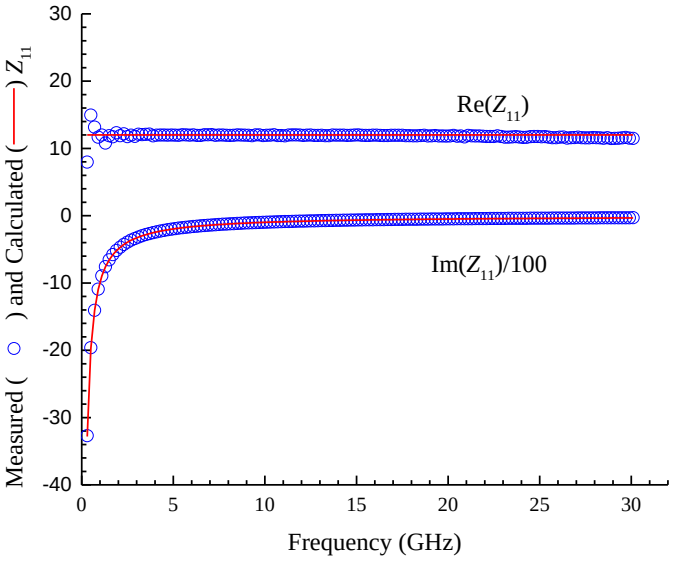
In Fig. 4.5, we plot the R_g expression $\mathrm{Re}(Z_{11}) - \mathrm{Re}(Z_{22})/4$ using deembedded Z-parameters vs. frequency up to 30.1 GHz. Based on (4.30), the critical frequency f_c is 53.8 GHz which is much higher than the highest frequency 30.1 GHz in the measurements. As shown in Fig. 4.5, the measured $\mathrm{Re}(Z_{11}) - \mathrm{Re}(Z_{22})/4$ is pretty constant up to 15 GHz.

Therefore, the gate resistance R_g can be obtained by extrapolating the measured Re(Z_{11}) – Re(Z_{22})/4 vs. frequency characteristics down to DC. In this case, the extrapolation is shown in red dashed line and it gives R_g = 10.72 Ω for this device.

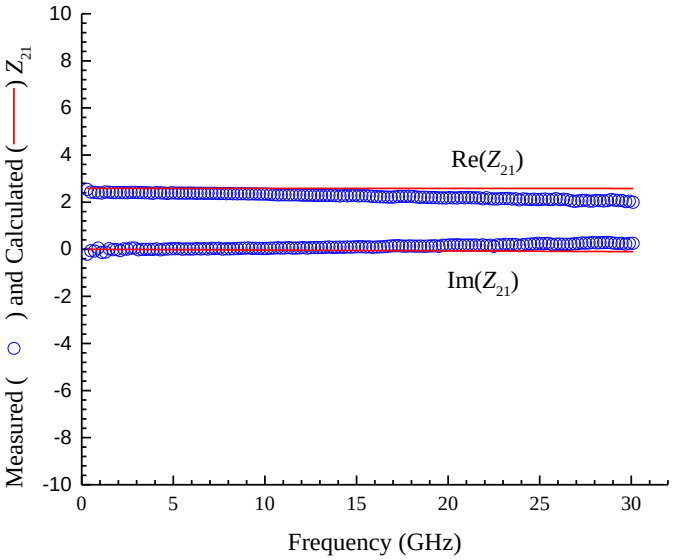
4.4.2 Accuracy of the Small-Signal Model

Since the derivation of the R_g expression is based on the equivalent circuit in Fig. 4.4, its accuracy depends on the accuracy of the equivalent circuit in the proposed frequency range. The next step is to verify the accuracy of the equivalent circuit used in this study.

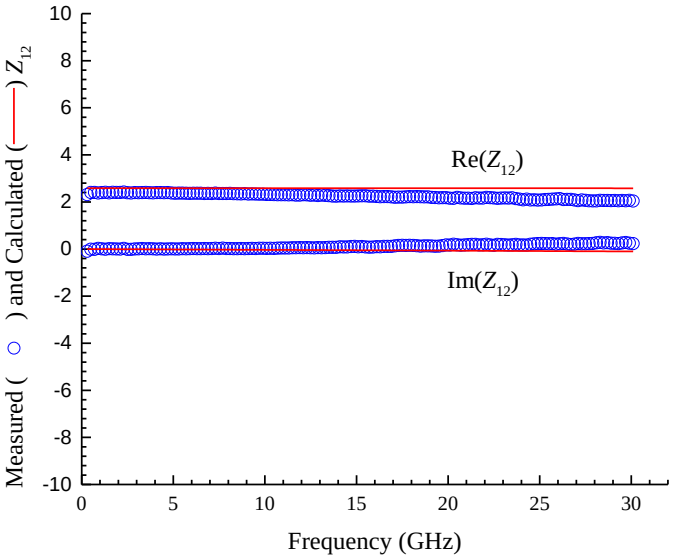
The same set of Z-parameters as in the previous section is used. Based on the equivalent circuit in Fig. 4.4, we are also able to extract the values of other elements apart from R_g using Z-parameters. In this case, their values are R_g = 10.72 Ω, C_{gso} = C_{gdo} = 80.8 fF, R_{dso} = 5.2 Ω, and R_s = R_d = 0 Ω, respectively. Fig. 4.6 plots the measured (symbols) and calculated (lines) intrinsic *Z*-parameters versus frequency characteristics. Very good agreement is obtained for all four parameters up to 15 GHz which is high enough for our extraction procedure. The excellent agreement in Z_{11} and Z_{22} confirms that the strategy using the bias condition at V_{GS} = 1.1 V and V_{DS} = 0 V successfully blocks the impact of substrate parasitics coupled into the measured *Z*-parameters and the simple equivalent circuit shown in Fig. 4.6 is applicable in the gate resistance extraction. The discrepancy between the measured Z_{12} and Z_{21} above 15 GHz could be caused by neglecting high-frequency components such as C_m, C_{mb}, C_{mx}, and C_{sd} in [114] and [115].


Re(Z_{11})
Im(Z_{11})/100
Measured (○) and Calculated (—) Z_{11}
Frequency (GHz)

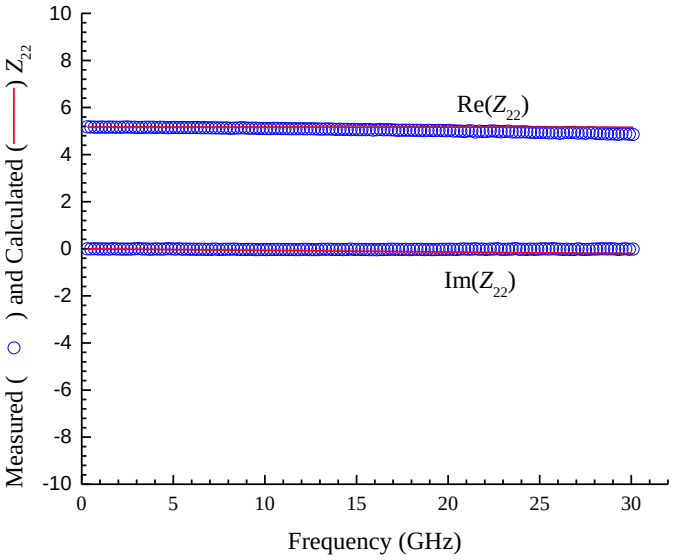
(a)


Re(Z_{21})
Im(Z_{21})
Measured (○) and Calculated (—) Z_{21}
Frequency (GHz)

(b)



(c)



(d)

Fig. 4.6. Measured and calculated intrinsic *Z*-parameters vs. frequency characteristics for an *n*-type FET with $L = 40$ nm, $W_f = 8$ μm, $N_f = 8$, and $M = 3$ biased at $V_{GS} = 1.1$ V and $V_{DS} = 0$ V. The calculation is based on the equivalent circuit in Fig. 4.4 with $R_g = 10.72$ Ω, $C_{gso} = C_{gdo} = 80.8$ fF, $R_{dso} = 5.2$ Ω, and $R_s = R_d = 0$ Ω, respectively. (© 2014 IEEE.)

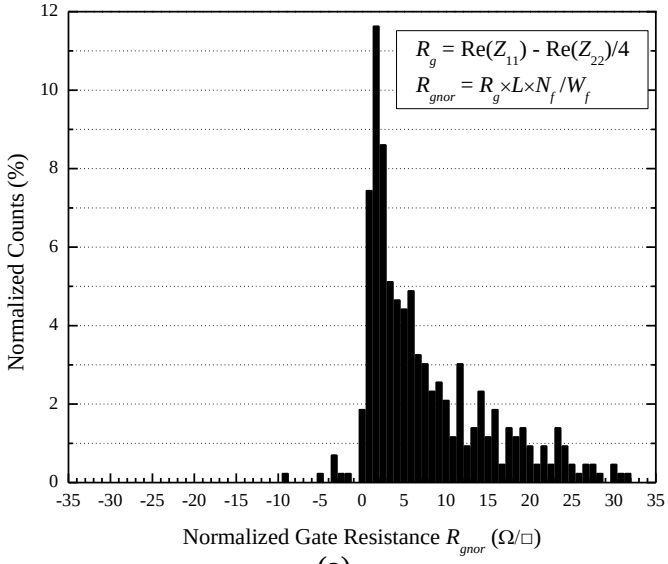
4.4.3 Statistical Verification of the R_g Extraction Procedure

For future sub-100-nm technology nodes, process variations, coming from both historical sources and emerging sources, become a very crucial aspect in the device fabrications [115]. Therefore, any proposed parameter extraction routine needs to be robust enough to demonstrate its statistical stability. To verify the robustness of this newly proposed extraction procedure, we apply our procedure to the experimental data from 430 devices fabricated in UMC's 40-nm, 55-nm, 90-nm, and 110-nm CMOS technology nodes. The 430 devices have different values of channel length (L), finger width (W_f), number of fingers (N_f), and multiplier (M). In comparison, we also apply the previously published Y-parameter based approaches described in Section 4.2 to the same set of data.

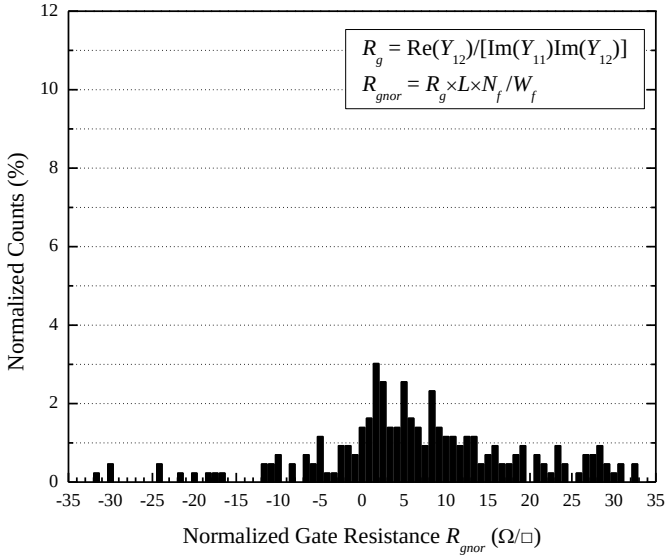
For the purpose of comparing the gate extraction methods, we define the normalized gate resistance as

$$R_{gnor} = R_g \cdot \left(\frac{L \cdot N_f \cdot M}{W_f} \right) \tag{4.34}$$

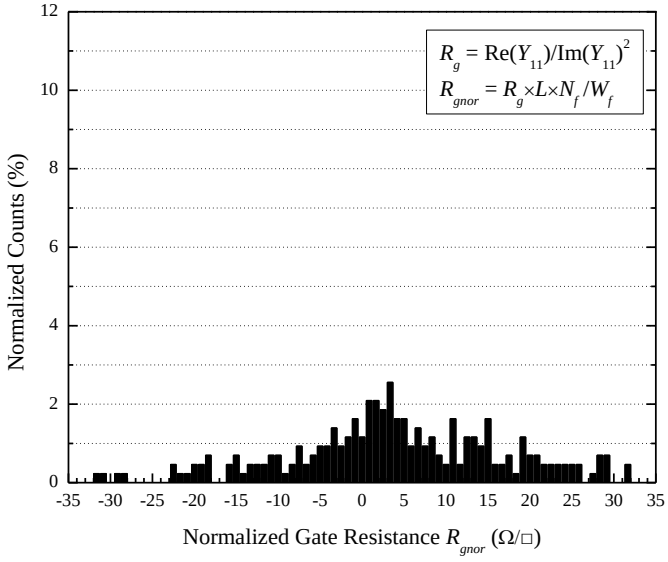
which is the gate resistance independent of device geometry. It should be noted that the normalized gate resistance R_{gnor} is not the same as the sheet resistance R_{gsh} of the distributed poly-silicon resistance R_{gpoly} because the extracted R_g still consists of the distributed channel resistance R_{ch} and the gate contact resistance R_{gcon}. Fig. 4.7 shows the normalized histograms (normalized to 430) of R_{gnor} obtained by the proposed method $R_g = \mathrm{Re}(Z_{11}) - \mathrm{Re}(Z_{22})/4$, (b) $R_g = \mathrm{Re}(Y_{12}) / [\mathrm{Im}(Y_{11}) \cdot \mathrm{Im}(Y_{12})]$ [105], (c) $R_g = \mathrm{Re}(Y_{11}) / \mathrm{Im}(Y_{11})^2$ [106],[114], and (d) $R_g = \mathrm{Re}(Y_{11}) / [4 \cdot \mathrm{Im}(Y_{12})^2]$ [113].We observed that for our proposed method, 86%


R_g = Re(Z_{11}) - Re(Z_{22})/4
$R_{gnor} = R_g \times L \times N_f / W_f$
Normalized Counts (%)
Normalized Gate Resistance R_{gnor} (Ω/□)

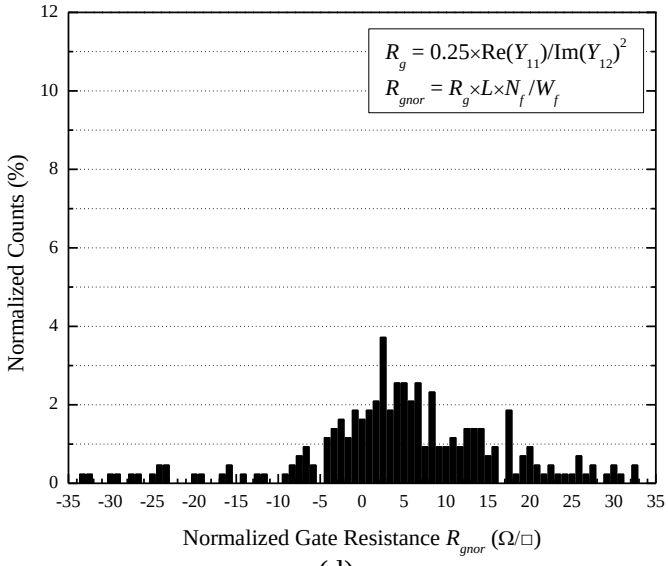
(a)


R_g = Re(Y_{12})/[Im(Y_{11})Im(Y_{12})]
$R_{gnor} = R_g \times L \times N_f / W_f$
Normalized Counts (%)
Normalized Gate Resistance R_{gnor} (Ω/□)

(b)



(c)



(d)

Fig. 4.7. Normalized histograms (normalized to 430) of the normalized gate resistance R_{gnor} (Ω/□) using R_g extracted from (a) the proposed method $R_g = \mathrm{Re}(Z_{11}) - \mathrm{Re}(Z_{22})/4$, (b) $R_g = \mathrm{Re}(Y_{12}) / [\mathrm{Im}(Y_{11}) \cdot \mathrm{Im}(Y_{12})]$ [105], (c) $R_g = \mathrm{Re}(Y_{11}) / \mathrm{Im}(Y_{11})^2$ [106],[114], and (d) $R_g = \mathrm{Re}(Y_{11}) / [4 \cdot \mathrm{Im}(Y_{12})^2]$ [113]. (© 2014 IEEE.)

of the extracted R_{gnor} (i.e., 370 samples) falls in between 0 and 35 Ω/□, while 38.8% (167 samples) for the method in [105], 34.9% (150 samples) for the approach in [106],[114], and 42.3% (182 samples) for the method in [113] fall in the same range. In addition, among all the methods, our proposed approach gives the fewest counts for the negative R_{gnor} which is physically incorrect. Therefore, our proposed extraction procedure is statistically most stable and robust among all of the published methods in the literature.

4.5 Extraction of Channel Resistance and LDD Resistance

In order to calculate the critical frequency f_c given by (4.30), the channel resistance R_{dso} at V_{DS} = 0 V needs to be extracted first. Based on the same small-signal model in Fig. 4.4 and (4.32), we can extract R_{dso} using the real part of Z_{22}. In this experiment, we measured *n*-type FETs (NFETs) and *p*-type FETs (PFETs) with W_f = 4 μm, N_f = 8, and M = 4 biased at V_{GS} = 1.1 V and V_{DS} = 0 V fabricated in a 55-nm technology node with low V_T (LVT), regular V_T (RVT), and high V_T (HVT) processes.

The extracted R_{dso} versus channel length characteristics are plotted in Fig. 4.8 for both NFETs and PFETs in all three processes. Compared to NFETs, PFETs demonstrate much higher channel resistance R_{dso}. In addition, for both NFETs and PFETs, different implants were used to change the threshold voltage V_{TH} and therefore resulted in different R_{dso}. For the three processes, LVT devices show the smallest channel resistance at the same bias condition, which are closely followed by RVT devices. HVT devices have the largest channel resistance compared to the other two processes given the same V_{GS} and V_{DS}.

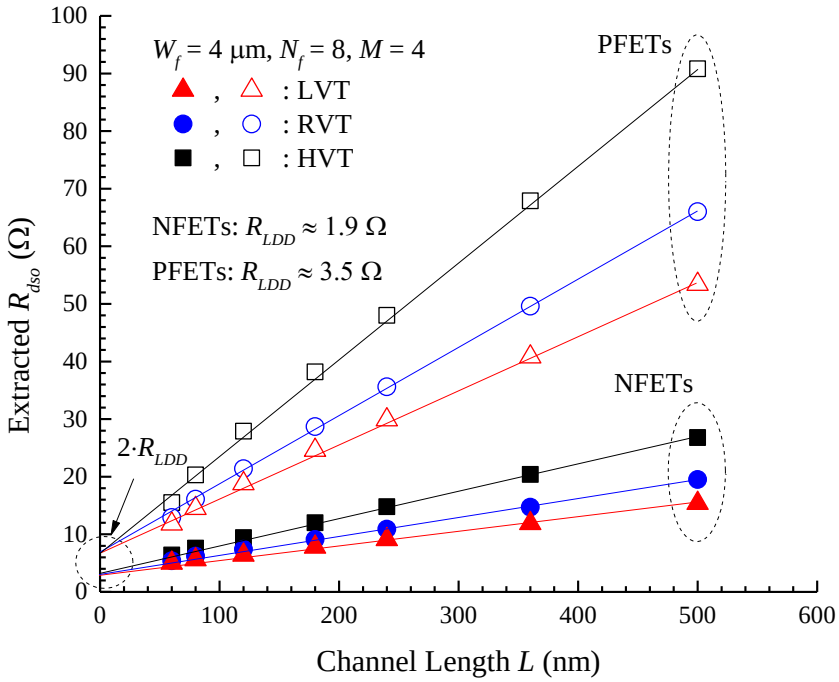


Fig. 4.8. Extracted channel resistance R_{dso} vs. channel length characteristics for devices with W_f = 4 μm, N_f = 8, and M = 4 biased at V_{GS} = 1.1 V and V_{DS} = 0 V fabricated in a 55-nm node with LVT, RVT, and HVT processes, respectively. (© 2014 IEEE.)

In Fig. 4.8, by extrapolating the extracted R_{dso} down to L = 0 nm, we can extract the resistance R_{LDD} in the lightly-doped drain (LDD) region at the source (or drain) side, which gives 1.9 Ω and 3.5 Ω for *n*- and *p*-type FETs, respectively. Moreover, R_{LDD} for all three processes are the same for the same type of devices. For traditional CMOS planar technology, the LDD resistance is rather small and only becomes more significant in shorter channel length. However, for emerging technologies like FinFETs, the LDD resistance plays a much bigger role and can cause gate voltage dependence and affect the extraction of mobility [118].

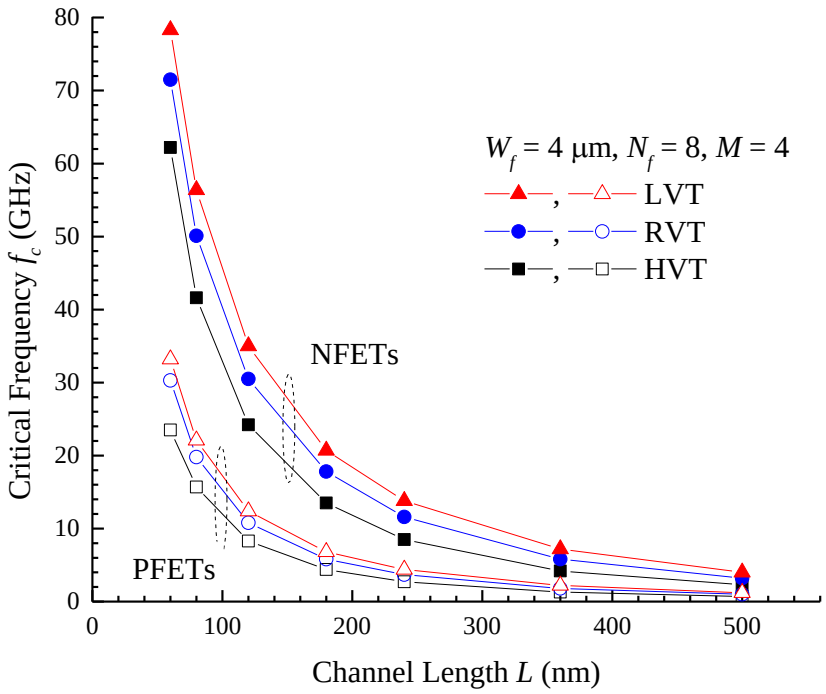


Fig. 4.9. Critical frequency f_c vs. channel length characteristics for devices with W_f = 4 μm, N_f = 8, and M = 4 biased at V_{GS} = 1.1 V and V_{DS} = 0 V fabricated in a 55-nm node with LVT, RVT, and HVT processes, respectively. (© 2014 IEEE.)

Based on the extracted R_{dso}, the critical frequency f_c for the NFETs and PFETs in the three processes are calculated and shown in Fig. 4.9. Based on (4.30), higher channel resistance R_{dso} results in lower f_c. This trend can be observed from Fig. 4.9 where PFETs show much lower critical frequencies than NFETs in the same process. HVT devices also have the lowest f_c in the three processes. As the channel length becomes shorter than 100 nm, the critical frequencies for all processes increase rapidly to above 10 GHz. Therefore, we can expect the proposed R_g extraction method has a wider frequency range to work with for short channel devices.

4.6 Dependence of Extracted R_g on Channel Length and Process

The gate resistance is dependent on the channel length in several ways. First, the distributed polysilicon gate resistance R_{gpoly} is proportional to W_f/L [104]. Second, the NQS effect starts to make an impact on the gate resistance when the frequency is high by bringing in the effect of the distributed channel resistance R_{ch} which is proportional to L/W_f [108]. However, since the proposed method works in a relatively low frequency range and uses extrapolated R_g values down to DC, the NQS effect is supposed to be absent. Third, as we will see from the experimental data, the gate contact resistance, which is the part of the gate resistance that is independent from L, also plays a significant role in the extracted gate resistance. In addition, as demonstrated in the previous section, both channel length and process have an impact on the channel resistance of the device, and therefore affect the critical frequency f_c. If f_c is too small and causes reduction of the valid frequency range of the measured data, it could reduce the accuracy of the extracted R_g.

Using the measured Z-parameters on the same devices described in the previous section, the gate resistance is extracted and plotted in Fig. 4.10. The gate resistance R_g of LVT and RVT NFETs decreases monotonically as the channel length increases, which agrees with the length dependence of the polysilicon gate resistance. However, for NFETs in HVT process and PFETs in all three processes, R_g first decreases as the channel length L becomes bigger and then increases after L is longer than a certain value. This can be explained by the critical frequency plot in Fig. 4.9 where those devices have relatively small f_c and could

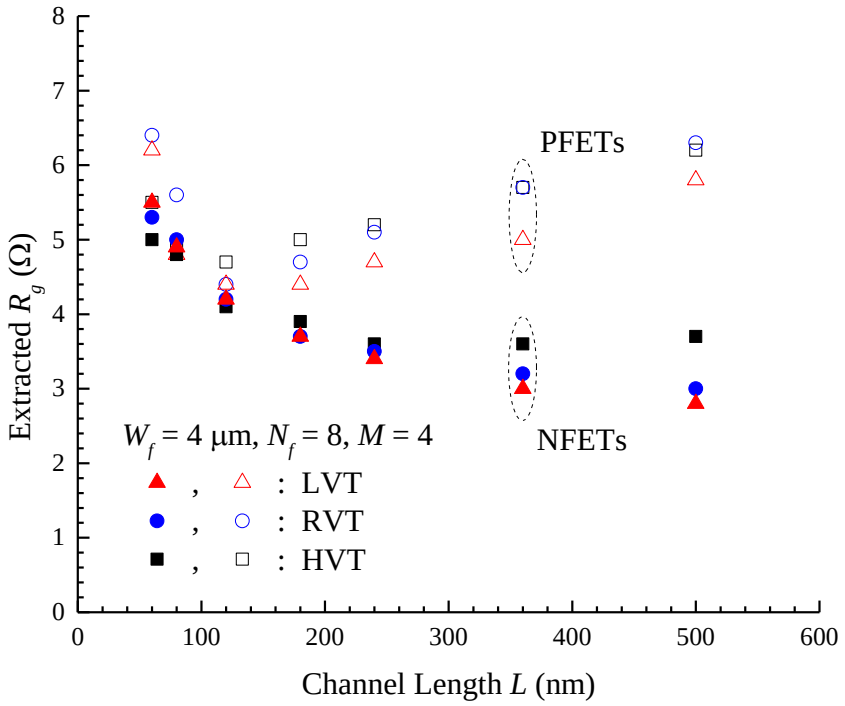


Fig. 4.10. Extracted R_g vs. channel length characteristics for devices with W_f = 4 μm, N_f = 8, and M = 4 biased at V_{GS} = 1.1 V and V_{DS} = 0 V fabricated in a 55-nm node with LVT, RVT, and HVT processes, respectively. (© 2014 IEEE.)

cause the malfunction of the proposed method. Fortunately, it is not an issue for sub-100-nm devices which are the focus of the future MOSFETs.

To further investigate the dependence of R_g on channel length quantitatively, we plot the extracted R_g versus 1 μm/L for NFETs in LVT, RVT, and HVT processes in Fig. 4.11, respectively. A linear relationship can be found through the dashed line with a slope of 0.17 Ω and an intercept of 2.65 Ω. The resistance of the polysilicon gate is a function of the geometry of the gate and should follow a 1/L dependence, as given by [112]

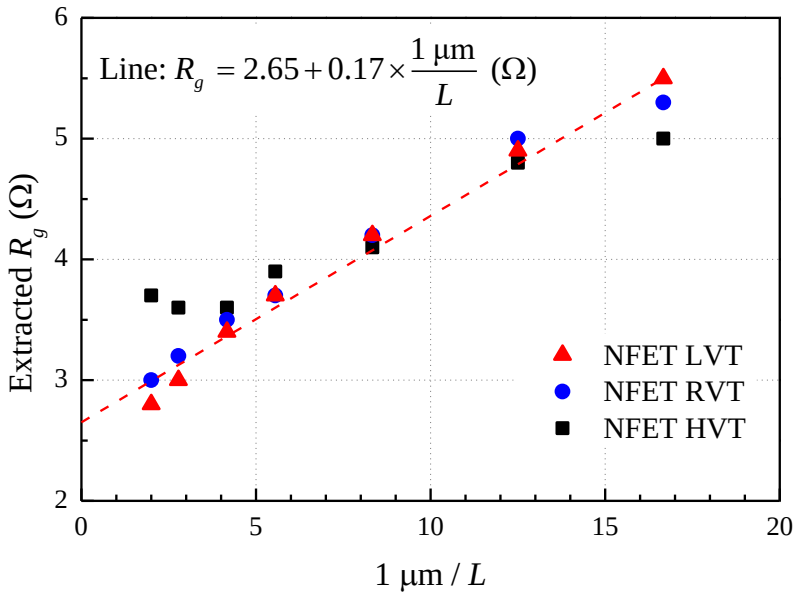


Fig. 4.11. Extracted R_g vs. the inverse of channel length for NFETs in a 55-nm node with LVT, RVT and HVT processes. A linear relationship between R_g and $1/L$ can be found through the dashed line.

$$R_{gpoly} = \frac{R_{gsh}}{N_f L}\left(W_{ext} + \frac{W_f}{\alpha}\right) \tag{4.35}$$

where R_{gsh} is the gate sheet resistance, W_{ext} is the extension of the polysilicon gate over the active region, and α is a factor for the distributed effect, which is 3 and 12 for single-sided and double-sided connections, respectively. However, as shown in Fig. 4.11, a finite intercept of the dashed line with the vertical axis, which essentially means an extrapolation to $L \rightarrow \infty$, indicates that there exists non-negligible contact resistance in the extracted R_g. Therefore, we can express R_g as

$$R_g = R_{gcon} + R_{gpoly} \tag{4.36}$$

where R_{gploy} is given by (4.35) and R_{gcon} is the gate contact resistance. From Fig. 4.11, R_{gcon} is extracted to be 2.65 Ω for this technology, which is about half of the total gate resistance for the 55-nm devices. In addition, if we ignore the W_{ext} in (4.35), the gate sheet resistance R_{gsh} can be extracted as 4.1 Ω/□.

4.7 Conclusion

In this chapter, a novel *Z*-parameter based approach to extract the gate resistance at low frequencies is presented. This newly proposed extraction method demonstrates its statistical stability and robustness compared to all of the published methods in the literature. This statistical robustness and stability are particularly important for the future technology nodes with smaller feature size and larger process variations. The channel resistance at zero V_{DS} and the LDD resistance are extracted using the same small-signal model. MOSFETs with smaller channel resistance demonstrate larger critical frequencies and thus wider frequency ranges for the proposed R_g extraction method. Finally, the dependence of extracted gate resistance on channel length and process is investigated and discussed. The length-independent gate contact resistance is also extracted and it represents a significant portion of the total gate resistance.

Chapter 5.

Noise Modeling for Future Nanoscale MOSFETs

5.1 Introduction

Research in high-frequency noise modeling of MOSFET has been a driving force to address the signal-to-noise dilemma in deep sub-micron MOSFETs for radio-frequency (RF) integrated circuits (ICs) [119],[120]. The largest contributor of MOSFET noise is that generated in the channel and appearing in the drain current [121]. Therefore, accurate modeling of the channel noise, also known as the drain current noise, has been a major focus of noise research for MOSFETs.

Previous research efforts have identified the channel noise as thermal noise generated by the resistive channel based on experimental results from long-channel MOSFETs [1]. Van der Ziel introduced the noise coefficient γ as the ratio of the actual channel noise with the thermal noise of the channel resistance when the lateral bias V_{DS} is zero [1]. Long-channel noise theory predicts that $\gamma = 1$ when $V_{DS} = 0$ and $\gamma = 2/3$ when the MOSFET reaches saturation. As Moore's Law guides the semiconductor industry and device downscaling continues, the long-channel theory cannot be used to explain experimental data that requires γ higher than 2/3 for shorter-channel transistors. As shown in Fig. 5.1,

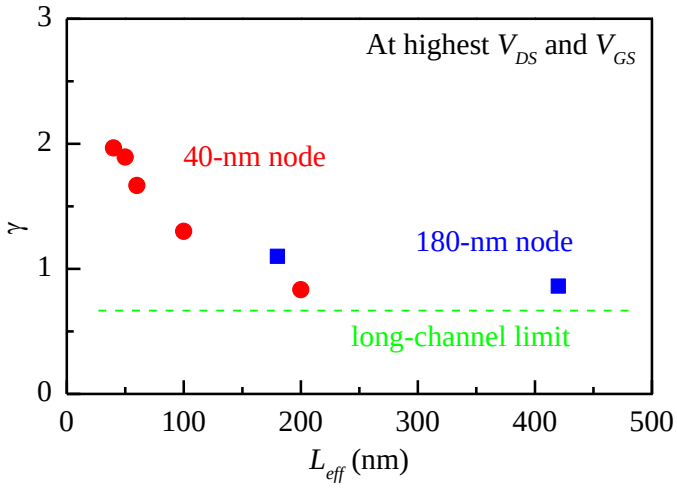


Fig. 5.1. Noise coefficient γ versus channel length for MOSFETs in 180-nm [122] and 40-nm [127] technology nodes. The green dashed line shows the long-channel limit of 2/3.

MOSFETs in 40-nm technology biased at highest V_{DS} and V_{GS} can have γ increase as channel length decreases and reach around 2 at 40 nm channel length. In order to extend the theory to work for those devices, researchers managed to keep using thermal noise as the fundamental noise source and incorporated various short-channel effects, e.g., the channel length modulation (CLM) effect in [122]-[125], and the non-uniform mobility in [126], etc. However, as the channel length shrinks even shorter to sub-100 nm, experimental excess noise, which signifies the onset of non-equilibrium transport in which the noise is no longer thermal, was reported in [127].

On the other hand, shot noise, which originates from the discreteness of charge transfer events, is inherently based on non-equilibrium. The well-known shot noise expression $S_{shot} = 2qI$ can be derived by viewing the charge transfer as a Poissonian process, i.e., the charge transfer events are identical and independent [1]. It has been well accepted as the main

noise mechanism when there is a barrier in the current flow path, e.g., noise in diodes and bipolar transistors. Contrary to the common belief that shot noise and thermal noise are two independent noise mechanisms, research in mesoscopic physics shows that they are in fact the specific forms of a more general noise theory [128]. Shot noise has been successful in explaining the channel noise for MOSFETs working in the sub-threshold region [115] due to the fact that charge transfer events are scarce and have minimal interaction [129]. However, as the MOSFET moves into strong inversion, a thin inversion layer with abundant free charges makes interaction inevitable. The interaction, also known as scattering, makes the actual noise to be smaller than the shot noise limit $2qI$. The physical sources of the interaction include elastic electron-electron scattering caused by Pauli exclusion and Coulomb repulsion [130], as well inelastic electron-phonon scattering. Under strong inelastic scattering, the carriers will be in equilibrium or near-equilibrium and the noise will reduce to thermal noise [131].

For very short-channel MOSFETs, recent modeling research has replaced drift-diffusion with ballistic transport [132] where carriers have no time to reach thermal equilibrium. For drain current noise under ballistic transport, it is natural to adopt a shot noise perspective [129]. However, due to the complexity of the scattering mechanisms, there is no analytical equation in terms of macroscopic device parameters to quantify the suppression of shot noise relative to the $2qI$ limit.

In this chapter, we propose a consistent perspective to interpret the channel noise as suppressed shot noise for both long- and short-channel MOSFETs. We will first derive an

easy-to-use analytical equation for the suppression factor that is applicable to MOSFETs working in the saturation region. It is followed by comparison between the predicted suppression factors using this equation and experimental data, as well as discussions on its advantages and disadvantages.

5.2 Thermal Noise Based Approaches for Channel Noise Modeling

Traditional modeling efforts on MOSFET channel noise have been primarily based on the thermal noise assumption, i.e., the channel is in thermal equilibrium and the noise sources are thermal, thus the popular name – channel thermal noise. In this section, we will review those efforts and also explain the challenges they are facing as the scaling of transistors goes on.

5.2.1 Origin of Thermal Noise

Thermal noise was first discovered experimentally from voltage fluctuations of conductors by Johnson in 1928 [133], hence the alternative name Johnson noise. It was explained theoretically by Nyquist in the same year [134], which leads to the well-known Nyquist's equation for current fluctuations of a conductor with conductance G

$$S_I = 4kTG \tag{5.1}$$

and for voltage fluctuations of a resistor with resistance R

$$S_V = 4kTR \tag{5.2}$$

where k is Boltzmann's constant and T is the absolute temperature. The thermal noise is white noise, which means it has constant power spectral density throughout the frequency spectrum. The Nyquist's equation relies on the assumption that the system must be in thermal equilibrium. In addition, for the equation to be valid, the electrical component needs to be purely passive and linear.

In semiconductor devices, the diffusion current, which is subject to frequent inelastic collisions between the carriers and the lattice, also produces thermal noise. Van der Ziel has proved that the microscopic noise source generated by diffusion current reduces to thermal noise if Einstein's relation holds [1], i.e.,

$$D = \frac{kT}{q}\mu \tag{5.3}$$

where D is the diffusion constant, μ is the mobility of the carriers, and q is the carrier charge. For the Einstein's relation to be valid, the semiconductor must be non-degenerate and reside in thermal equilibrium. However, in modern semiconductor devices which have very short channel lengths, this assumption is challenged frequently [126].

5.2.2 Thermal Noise Based Models for MOSFET Channel Noise

A MOSFET is a non-linear device in which current is not always linear with voltage, therefore the Nyquist's equation $4kTG$ cannot be applied directly. However, it is possible to divide the MOSFET channel into microscopic segments and each segment will produce thermal noise $4kTg(x)$ as long as the carriers can be considered in thermal equilibrium or near-equilibrium at channel position x. The overall noise generated by the device can then

be found by integrating the contribution of each segment along the channel. This method was first proposed as the Klaassen-Prins equation [135] which is given by

$$S_{id} = \frac{1}{L^2}\int_0^L 4kTg(x)\,\mathrm{d}x = \frac{4kT}{L^2 I_D}\int_0^{V_{DS}} g^2(V)\,\mathrm{d}V \tag{5.4}$$

where $g(x)$ is the local conductance per unit length in the channel position x and L is the channel length. The change of the integrating variable from x to V in (5.4) is based on the current continuity equation

$$I_d = g(V)\frac{\mathrm{d}V}{\mathrm{d}x}. \tag{5.5}$$

The Klaassen-Prins equation can be useful in noise modeling of MOSFETs. However, the integration is not very practical for hand calculations or compact models in engineering.

For this purpose, van der Ziel has introduced the noise coefficient γ to characterize the channel thermal noise [136] which is now widely used. It is defined as the ratio of the actual channel noise to the thermal noise caused the channel resistance when the drain-to-source voltage is zero and the channel can be regarded as a linear resistor, i.e.,

$$\gamma = \frac{S_{id}}{4kTg_{d0}} \tag{5.6}$$

where g_{d0} is the channel conductance at V_{DS} = 0 V. Naturally, when V_{DS} is 0 V or close to 0 V, the channel behaves like a linear resistor and Nyquist's equation can be applied, which leads to γ = 1. When V_{DS} is further increased, the MOSFET will gradually reach saturation and its current voltage characteristics becomes nonlinear. In this case, the drain current noise will diverge from the macroscopic resistive noise and its value relies on the

evaluation of (5.4). Based on the DC model of long-channel MOSFETs, γ is evaluated to be 2/3 after the MOSFET reaches saturation, and remains constant in saturation region. However, as the scaling of MOSFETs continues, various short-channel effects have appeared and modified the long-channel DC model, which also have undermined the theoretical foundation of the long-channel value of γ.

In noise modeling, it is beneficial to have a compact equation for the channel noise of MOSFETs that works for all regions of operation. The local conductance for a channel segment dx at the position x can be expressed as

$$g(x) = W\mu_{eff}\left(-Q_{inv}(x)\right) \tag{5.7}$$

where W is the transistor width, μ_{eff} is the effective mobility, and $Q_{inv}(x)$ is the inversion charge density at position x which is negative for n-type MOSFETs. Substituting (5.7) into the Klaassen-Prins equation (5.4), we can write

$$\begin{aligned} S_{id} &= \frac{1}{L^2}\int_0^L 4kT \cdot W\mu_{eff}\left(-Q_{inv}(x)\right)\mathrm{d}x \\ &= 4kT\,\frac{\mu_{eff}(-Q_{inv})WL}{L^2} \end{aligned} \tag{5.8}$$

where μ_{eff} is assumed to be position-independent. By using the total inversion charge Q_I to replace $Q_{inv}WL$ and the effective channel length L_{eff} to replace L, we have a more compact form as follows

$$S_{id} = 4kT\,\frac{\mu_{eff}(-Q_I)}{L_{eff}^2}. \tag{5.9}$$

The total-inversion-charge-based MOSFET channel noise equation is proposed by Tsividis [115] and can work in all regions of operation, including subthreshold region where it reduces to shot noise expression.

As the channel lengths of MOSFETs keep shrinking, various short channel effects have a strong impact on the noise modeling. One of the earliest report of enhancement of thermal noise with respect to long-channel theory comes from Abidi, in which the thermal noise was found to be enhanced by a factor up to 12 for a 0.7-μm *n*-channel device and the effect of hot electrons was proposed to explain the anomalous results [137]. Although the report and explanation was not widely accepted [138], it has successfully attracted a lot of research efforts into the issue of noise enhancement for short-channel MOSFETs. Many short-channel effects start to make their way into noise modeling, and most of them are proven to cause an enhancing effect of the channel noise.

Chen & Deen pointed out that the velocity saturation region of the MOSFET channel, as indicated in Fig. 5.2, does not generate noise [122]. They modified the total-inversion-charge-based equation (5.9) to account for this effect, i.e.,

$$S_{id} = \frac{4kT}{L_{elec}^2}\mu_{eff}(-Q_I) \tag{5.10}$$

where L_{elec} is the electric channel length resulted from excluding the length of the velocity saturation region ΔL from the effective channel length L_{eff}. They also experimentally demonstrated that the channel length modulation (CLM) effect plays a significant role in

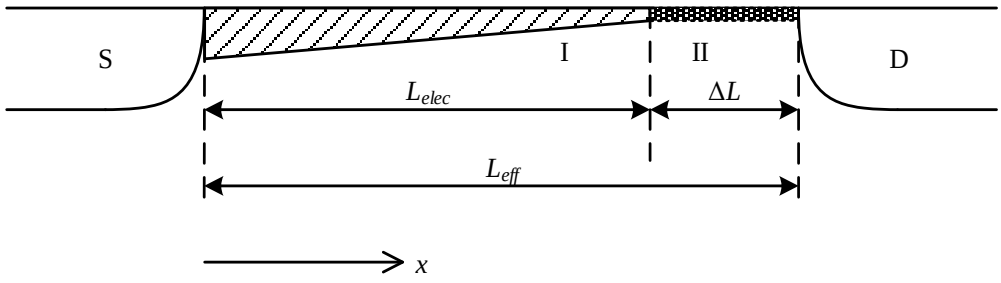


Fig. 5.2. Cross-section of a MOSFET in saturation. The channel is divided into two sections – the gradual channel region (I) and the velocity saturation region (II).

achieving accurate noise modeling using (5.10), especially for the channel length down to 180 nm.

Han *et al.* took into account the carrier heating and velocity saturation effect in the gradual channel region [139]. It leads to the modified Klassen-Prins equation which is given by

$$S_{id} = \frac{4kT}{I_d L_{elec}^2 \left(1+\frac{V_{DS}}{L_{elec}E_c}\right)^2} \int_0^{V_{DS}} g_0^2(V)\left(1+\frac{E}{E_c}\right) \mathrm{d}V \qquad (5.11)$$

where E_C is the critical electric field at which the carrier velocity becomes saturated. Interestingly, by making the approximation that $(1+E/E_c) = (1+V_{DS}/L_{elec}E_c)$, the complicated integration (5.11) simplifies to the total-inversion-charge-based (5.10).

In the derivation of the compact noise equations (5.9) and (5.10), the effective mobility μ_{eff} is assumed to be position independent along the channel. Due to velocity saturation effect in short-channel MOSFETs, the mobility is dependent on the lateral electric field.

Therefore, the assumption may not be true if the electric field in the channel is non-uniform. Roy and Enz took into account the non-uniformity of the lateral electric field in the channel and re-formulated the noise expression as [126]

$$S_{id} = 4kT \frac{W \cdot \int_0^{L_{elec}} (-Q_{inv}) \frac{\mu_0^2}{\mu_{eff} + \mu'_{eff} E} \mathrm{d}x}{L_{elec}^2 \left(1 + \frac{1}{L_{elec}} \int_{V_s}^{V_{deff}} \frac{\mu'_{eff}}{\mu_{eff} + \mu'_{eff} E} \mathrm{d}V \right)^2} \tag{5.12}$$

where μ_0 is the low-field mobility, μ_{eff} is the field-dependent mobility and μ'_{eff} is its derivative with respect to the lateral field. In deriving (5.12), the fundamental noise source is assumed to be $4kT_Cg$, where T_C is the carrier temperature, instead of the lattice temperature T_L. It turns out this has an opposite effect to the non-uniform mobility effect, which explains the close noise level predicted by (5.10) and (5.12) [126].

5.2.3 Excess Noise in the MOSFET Channel

Although being challenged by various short-channel effects, researchers have managed to model the channel noise quite accurately down to the channel length of 180 nm based on the assumption that the fundamental noise source remains thermal noise $4kTg$. However, for MOSFETs with channel length smaller than 100 nm, Smit *et al*. reported experimental data that exceeds the prediction of thermal-noise-based modeling [140], or in other words, "excess noise." Smit *et al*. were able to model this phenomenon by keeping existing derivation framework but modifying the fundamental microscopic noise source as [127]

$$s_i = 4kTg\left[1+3\left(\frac{V}{\phi_t}\frac{l}{L}\right)^2\right] \tag{5.13}$$

where l is the mean-free-path of the majority carriers, and ϕ_t is the thermal voltage kT/q. The second field-dependent term is to model the non-equilibrium transport of the carriers, which the authors argued to be the cause of the excess noise. Nonetheless, this renders the fundamental noise source no longer thermal. In the future, thermal noise may no longer be the adequate term to describe the channel noise of sub-100-nm MOSFETs.

5.3 Shot Noise as the Fundamental Noise Source

The discovery of shot noise is even earlier than thermal noise. While thermal noise is regarded as resistive noise, shot noise is generally perceived as the noise when there is a barrier in the current flowing path, e.g., in diodes and bipolar junction transistors (BJTs). As will be introduced later in this section, modern physics research shows that they are both specific forms of a more general noise theory and do not contradict each other.

5.3.1 Classical View of Shot Noise

Shot noise was first introduced by Schottky when studying the current fluctuations in vacuum tubes [141]. In classical theory, shot noise, which is also known as Poisson noise, occurs if the carrier transport can be considered as a Poissonian Process, i.e., the carrier transfer events are independent and identical [1]. Due to the discreteness of the charges, e.g., the charge of electrons q = 1.6x10^{-19} C, and the randomness in their arrivals, the instantaneous current will fluctuate around an average value. Schottky's equation for shot

noise, which is based on the assumption that the statistics of carrier transport is Poissonian, can be written as

$$S_{shot} = 2qI \tag{5.14}$$

where q is the charge of the carrier and I is the average current.

The shot noise expression $2qI$ has been well accepted as the main noise mechanism when there is a barrier in the current flow path, e.g., noise in diodes and BJTs. In addition, it is also successful in explaining the channel noise for MOSFETs working in the subthreshold region [142],[143] due to the fact that carriers are scarce and have minimal interaction when the channel is in weak inversion [129]. To make it more accurate, (5.14) needs to be modified to consider the noise generated by both the forward ($S \rightarrow D$) and the backward ($D \rightarrow S$) currents. From the subthreshold DC model for MOSFETs, the current is exponential with respect to the voltage between the terminals. If the forward current is denoted as I^+, the backward current should be $I^- = I^+ \exp(-qV_{DS}/kT)$ where V_{DS} is the voltage between the drain and source. The total drain current I_D is the difference between the two currents, i.e., $I_D = I^+ - I^-$, but the noise should be the summation of the contribution from them, i.e.,

$$\begin{aligned} S_{shot} &= 2q(I^+ + I^-) \\ &= 2qI_D \frac{1 + I^-/I^+}{1 - I^-/I^+} \\ &= 2qI_D \frac{1 + e^{-qV_{DS}/kT}}{1 - e^{-qV_{DS}/kT}}. \end{aligned} \tag{5.15}$$

We can further simplify it using the hyperbolic cotangent function coth(·), which leads to

$$S_{shot} = 2qI_D \coth\left(\frac{qV_{DS}}{2kT}\right). \tag{5.16}$$

The coth(·) function has the mathematical property that when $x \to 0$, $\coth(x) \to 1/x$, which will reduce (5.16) to $4kTG$ and achieves compatibility with thermal noise theory that models the MOSFET channel by a simple resistor as V_{DS} approaches 0 V. On the other hand, when $x \gg 1$, coth(x) reduces to 1 and (5.16) becomes the familiar unidirectional shot noise $2qI_D$. It should be noted that (5.16) does not take into account the various scattering mechanisms which have suppression effects on the drain current noise. Therefore, it can be called the unsuppressed shot noise.

Interestingly, the first two terms of the Taylor Series of (5.16) can be written as

$$S = 4kT\frac{I}{V_{DS}}\left[1+\frac{1}{12}\left(\frac{V_{DS}}{\phi_t}\right)^2\right] \tag{5.17}$$

where the first term is simply thermal noise $4kTG$ and second term shows its bias dependence with respect to V_{DS}. It shares resemblance to the microscopic noise source (5.13) proposed by Smit *et al*. that captures the onset of non-equilibrium.

5.3.2 Mesoscopic View of Shot Noise

Contrary to the common belief that shot noise and thermal noise are two independent noise mechanisms, research in mesoscopic physics shows that they are in fact the specific forms of a more general noise theory [128]. The mesoscopic physics research deals with devices that have intermediate dimensions between the microscopic and macroscopic range

where quantum theory plays a significant role. As the transistor scaling guided by Moore's law goes on, the transistor theory will eventually enter the mesoscopic field. In fact, there are already modeling efforts for nanoscale MOSFETs that are based on mesoscopic physics [132],[144].

In mesoscopic physics, following the Landauer approach [128], a two-terminal mesoscopic conductor can be regarded as having multiple conducting channels with different energy-dependent transmission probability, and the average current can be achieved by summing up the contribution from all the transmitting channels, i.e.,

$$\langle I_L \rangle = \frac{q}{2\pi\hbar} \sum_n \int T_n(E) \left[f_L(E) - f_R(E) \right] \mathrm{d}E \tag{5.18}$$

where f_L and f_R are the Fermi functions of the left and right terminals, and T_n is the transmission probability of each channel. It also leads to quantized conductance which is the summation of the quantum conductance of all the energy channels, i.e.,

$$G = \frac{q^2}{2\pi\hbar} \sum_n T_n. \tag{5.19}$$

The fluctuations of the current, or the noise, for a two-terminal mesoscopic conductor can be found from the statistical properties of the transmission channels, which can be written as [128]

$$S = \frac{q^2}{\pi\hbar} \sum_n \int \left\{ T_n \left[f_L(1 - f_L) + f_R(1 - f_R) \right] + T_n(1 - T_n)(f_L - f_R)^2 \right\} \mathrm{d}E \tag{5.20}$$

where the first two terms are equilibrium noise contribution and the third term is the non-equilibrium contribution. If the system is in thermal equilibrium, the left and right terminals then have the same Fermi function, i.e. $f_L = f_R$, and the noise will reduce to thermal noise $4kTG$. On the other hand, if the system is at the 0 K, the Fermi functions will become step functions and the noise will become suppressed shot noise. In particular, if the transmission probabilities are low, i.e., $T_n \ll 1$, the noise will become the unsuppressed shot noise $2q\langle I\rangle$.

5.3.3 Shot Noise Based Noise Modeling for Novel Devices

The scaling of CMOS planar technology is facing great challenges when the channel length decreases below 28 nm, where the alternative structures, e.g. FinFETs and FD-SOI FETs, emerge and become the technologies of choice. On the other hand, the research efforts on finding replacement for Silicon-based transistors have been going on for years. Notable candidates include the transistors based on 2D graphene-like material MoS_2 which features extremely high mobility [145],[146]; the transistors based on 1D carbon nanotubes [147]; and the tunneling-field-effect-transistor (TFET) which has steep subthreshold swing and aims for low power applications [148]-[150]. Research on the drain current noise of these novel devices have been primarily based on shot noise, e.g., for carbon nanotube transistors [152],[153] and for TFETs [154]. The suppression factor, also known as the Fano factor [155], is used in these research works to quantify the level of suppression with respect to the Poissonian level $2qI$. However, due to the complexity of the modeling for the various suppression mechanisms, a compact analytical equation for the suppression factor is yet to be found.

5.4 Shot Noise Suppression Factor

The channel noise of MOSFETs working in subthreshold region can be regarded as unsuppressed shot noise due to the low density of the carriers. But as we increase the gate-to-source voltage V_{GS}, the carriers in the channel become abundant and interaction between them are inevitable. In this case, the unsuppressed shot noise equation (5.16) becomes invalid and leads to an overestimation of the channel noise.

On the other hand, the noise coefficient γ, which is defined as the ratio between the actual noise and the resistive noise of the MOSFET when V_{DS} = 0 V, is significantly enhanced with respect to the long channel prediction in modern technology due various short channel effects. This makes it more of an empirical parameter but without a solid analytical expression. It undermines the role of γ as an easy-to-use indicator of the channel noise.

Unlike γ, which uses the resistive noise at V_{DS} = 0 V as the reference, we can define a new coefficient that references to the unsuppressed shot noise. It has the physical meaning of the level of suppression caused by various scattering mechanisms and therefore can be called suppression factor. Mathematically, we can define the suppression factor as the ratio of the actual noise in the drain current over the shot noise limit, i.e.,

$$F = \frac{\text{Actual Noise}}{\text{Shot Noise Limit}} = \frac{S_{id}}{S_{shot}}. \tag{5.21}$$

For MOSFETs working in the subthreshold region, we naturally have F = 1 which indicates unsuppressed shot noise. In addition, since the unsuppressed shot noise (5.16)

reduces to thermal noise when V_{DS} is around 0 V, it is also expected that $F = 1$ for V_{DS} close to 0 V even if V_{GS} is large. But when both V_{GS} and V_{DS} are large, F is expected to be a number between 0 and 1.

To simplify (5.21), we can substitute (5.10) and (5.16) into it. For practical applications, we assume that MOSFET is in the saturation region, i.e., the channel is in strong inversion and $V_{DS} > V_{DSat}$, where V_{DSat} is the saturation voltage. As long as V_{DS} is much larger than $2\phi_t$, the coth(·) function in (5.16) reduces to 1. The suppression factor can then be expressed in terms of the Q_I and I_D, i.e.,

$$F = 2\phi_t \frac{\mu_{eff}}{L_{elec}^2} \frac{(-Q_I)}{I_D}. \tag{5.22}$$

Now we need to further simplify (5.22) using expressions for the drain current I_D and total inversion charge Q_I.

For MOSFETs working in strong inversion and saturation, the drain current can be expressed as [157]

$$I_D = \frac{1}{2\alpha} \frac{W}{L_{elec}} \mu_{eff} C'_{ox} \left(V_{GS} - V_T\right)^2 \left(1 + \frac{V_{GS} - V_T}{\alpha E_C L_{eff}}\right)^{-1} \tag{5.23}$$

where α is the bulk charge factor, W is the transistor width, L_{eff} and L_{elec} are the effective channel length and the length for the gradual channel region as shown in Fig. 5.2, respectively, μ_{eff} is the effective mobility, C'_{ox} is the oxide capacitance per unit area, V_{GS} is the gate-to-source voltage, V_T is the threshold voltage of transistor, and E_C is the critical lateral field. Due to steep retrograde body doping profile in modern transistors, α becomes

basically a constant with a typical value of around 1.2 [157]. The electrical length L_{elec} is to model the CLM effect in the drain current and is dependent on the drain-to-source voltage V_{DS} [151]. The last $(\cdot)^{-1}$ factor is to specifically account for the velocity saturation effect that causes current degradation in short channel MOSFETs. The criteria for the MOSFET to be in saturation is that $V_{DS} \geqslant V_{Dsat}$, where V_{Dsat} is the saturation voltage that is given by

$$V_{Dsat} = \left(\frac{\alpha}{V_{GS} - V_T} + \frac{1}{E_C L} \right)^{-1}. \tag{5.24}$$

Using (5.24), the drain current can be simplified as

$$I_D = \frac{1}{2\alpha} \frac{W}{L_{elec}} \mu_{eff} C'_{ox} \left(V_{GS} - V_T \right) V_{Dsat}. \tag{5.25}$$

In deriving (5.25), a simple piece-wise model is used for the carrier velocity in the channel that can fit the experimental data rather accurately [156], which is given by

$$v = \begin{cases} \dfrac{\mu_{eff} E}{1 + E/E_C} & \text{for } E \leq E_C \\ v_{sat} & \text{for } E \geq E_C \end{cases} \tag{5.26}$$

where E_C is the critical lateral electric field at which carrier velocity becomes almost constant and does not increase further as the electrical field increases. By equating the two equations in (5.26), we can relate E_C to v_{sat} by

$$E_C = \frac{2v_{sat}}{\mu_{eff}}. \tag{5.27}$$

Empirical values for v_{sat} is 8×10^6 cm/s for electrons and 6×10^6 cm/s for holes [157]. The effective mobility μ_{eff} in MOSFETs are several times smaller than the bulk mobility due to the surface roughness scattering. It also depends on the type of the carriers, namely electrons or holes. Empirical equations for μ_{eff} have already been established as a function of the vertical field from various physical experiments. In this work, we use

$$\mu_{eff,n} = \frac{540\ \mathrm{cm}^2/\mathrm{Vs}}{1+\left(\frac{V_{GS}+V_T+0.2\mathrm{V}}{5.4t_{oxe}}\cdot\frac{\mathrm{cm}}{\mathrm{MV}}\right)^{1.85}} \tag{5.28}$$

for the effective mobility of NMOS [158] and

$$\mu_{eff,p} = \frac{185\ \mathrm{cm}^2/\mathrm{Vs}}{1-\left(\frac{V_{GS}+1.5V_T-0.25\mathrm{V}}{3.38t_{oxe}}\cdot\frac{\mathrm{cm}}{\mathrm{MV}}\right)}. \tag{5.29}$$

for that of PMOS [159], where t_{oxe} is the effective oxide thickness.

On the other hand, the total inversion charge Q_I in the channel of strongly inverted and saturated MOSFETs can be found as [157]

$$Q_I = -\frac{2}{3}WL_{elec}C'_{ox}(V_{GS}-V_T). \tag{5.30}$$

Here the electrical length L_{elec} is used instead of the effective channel length L_{eff}. It is based on the assumption that the length of the velocity saturation region is short and charge density in it is low, so we can ignore its contribution to the total inversion charge.

Using the drain current equation (5.25) and the total inversion charge equation (5.30), we can further simplify the suppression factor expression to

$$F = \frac{8\phi_t}{3}\frac{1}{V_{Dsat}} = \frac{8\phi_t}{3}\left(\frac{\alpha}{V_{GS}-V_T}+\frac{1}{E_C L_{eff}}\right). \tag{5.31}$$

In (5.31), the first term indicates dependence on the bias voltages and the second mainly on channel length. It also agrees with the experimental trend that if V_{GS} increases, the suppression will be stronger and F be smaller; and the same will happen if the channel length L_{eff} is longer.

Using the suppression factor, we can alternatively express the drain current noise for MOSFETs in saturation region as

$$S_{id} = 2qI_D \cdot F. \tag{5.32}$$

Since we have started with thermal noise based equation (5.10) in the derivation of F, it means that (5.32) is mathematically equivalent with (5.10) given the same charge and current equations used in the derivation. On one hand, this indicates that (5.32) can be as accurate as (5.10) in predicting channel noise for MOSFETs with $L \geq 180$ nm. On the other hand, equations (5.32) and (5.31) differ from (5.10) in that it makes the estimation of the channel noise more accessible by eliminating the electrical length L_{elec} which models the CLM effect and is complicated to calculate. It should be noted, this is achieved not by ignoring the CLM effect, but to cancel the effect by dividing the noise by the current that encounters the same effect. In the next section, we will demonstrate how well the suppression factor expression works with experimental data.

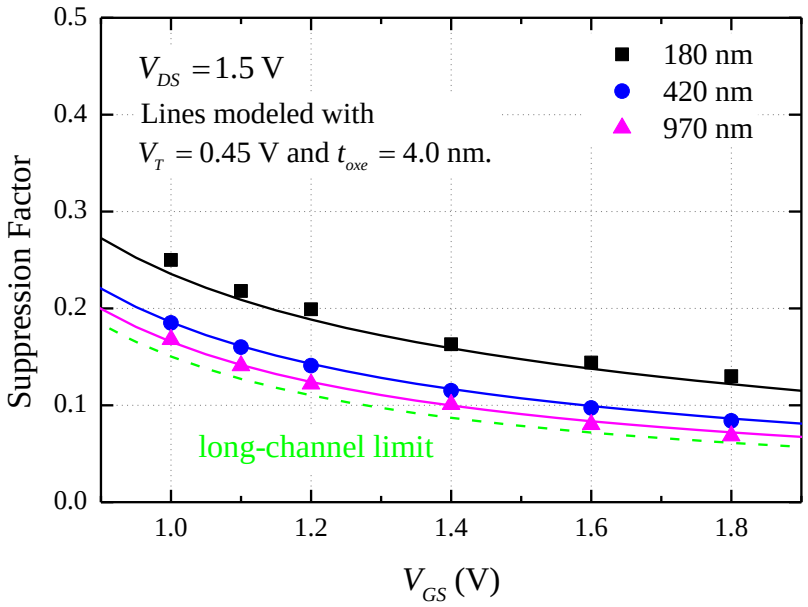


Fig. 5.3. Calculated (lines) and experimental (symbols) values of the channel noise suppression factor F versus V_{GS} for n-type MOSFETs manufactured in 180-nm CMOS technology. The symbols are calculated using experimental data extracted from [122] and [123]. The lines with same colors are calculated using the proposed suppression factor (5.31). The green dashed line represents the lower limit of the suppression factor by setting the channel length to ∞. (© 2017 IEEE.)

5.5 Experimental Results and Discussions

The first data set is measured on n-type MOSFETs with W = 60 um and L = 180, 420, and 970 nm fabricated in 180-nm CMOS technology [122],[123]. In Fig. 5.3, we plot both the calculated and experimental values of F versus the gate-source voltage V_{GS}. The calculated F are computed using (5.31) with fitted values of V_T = 0.45 V and t_{oxe} = 4.0 nm

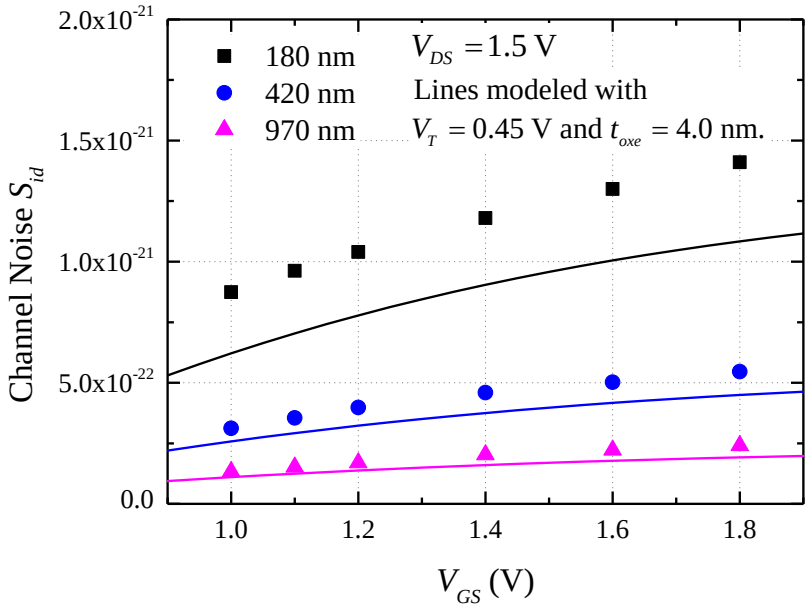


Fig. 5.4. Calculated (lines) and experimental (symbols) values of the channel noise power spectral density versus V_{GS} for n-type MOSFETs fabricated in 180-nm CMOS technology. The symbols are experimental data extracted from [122]. The lines with same colors are calculated using (5.10), with the exception that L_{elec} is replaced with L_{eff}, which causes the discrepancy between the calculated and experimental data. (© 2017 IEEE.)

whereas the experimental ones are obtained using the definition in (5.21) with the extracted noise and drain current. It results in a good match between the prediction and experimental data for the given bias range and channel lengths. Although fitted values of the two process parameters are used, they are within reasonable range and should be close, if not exactly the same, with the actual values. Furthermore, the long-channel limit of the suppression factor obtained by setting the channel length to ∞ is also shown by the green dashed line in Fig. 5.3. It is fairly close the curve that represents L = 970 nm, which suggests that F becomes almost independent of the channel length after it surpasses a certain value.

Using the same parameter values, we can also evaluate the traditional compact noise equation (5.10) except that L_{elec} is replaced by L_{eff} to skip the modeling of the CLM effect. The results are plotted in Fig. 5.4 using lines, together with the corresponding experimental data in symbols. A clear discrepancy happens between the calculated and experimental data, especially for the smallest length of 180 nm. The behavior has already been investigated in [122] and the reason is that CLM effect plays a significant role in the total-inversion-charge-based equation (5.10) with the presence of the electrical length L_{elec}. Therefore, to achieve accurate noise modeling using (5.10), it is essential to first have accurate modeling of the CLM effect. In contrast, the proposed suppression factor method works by letting the current take care of the CLM effect and only relies on easily accessible parameters from DC characterization.

So far, for channel lengths above 180 nm, it is experimentally verified that (5.10) still works if the CLM effect is properly taken care of. Therefore, it is expected that the suppression factor expression (5.31) also performs well in the same channel length range. However, as the channel length decreases down to sub-100-nm, thermal-noise-based (5.10) might lose its validity due to the non-equilibrium effect. The suppression factor method excels in this situation in that it starts from the concept of non-equilibrium transport and thus is inherently compatible with it.

In the following, we evaluate the suppression factor expression (5.31) for sub-100-nm devices using data from literature. Fig. 5.5 shows the calculated suppression factor in symbols using extracted data from [160] for *n*-type MOSFETs in 65-nm technology with

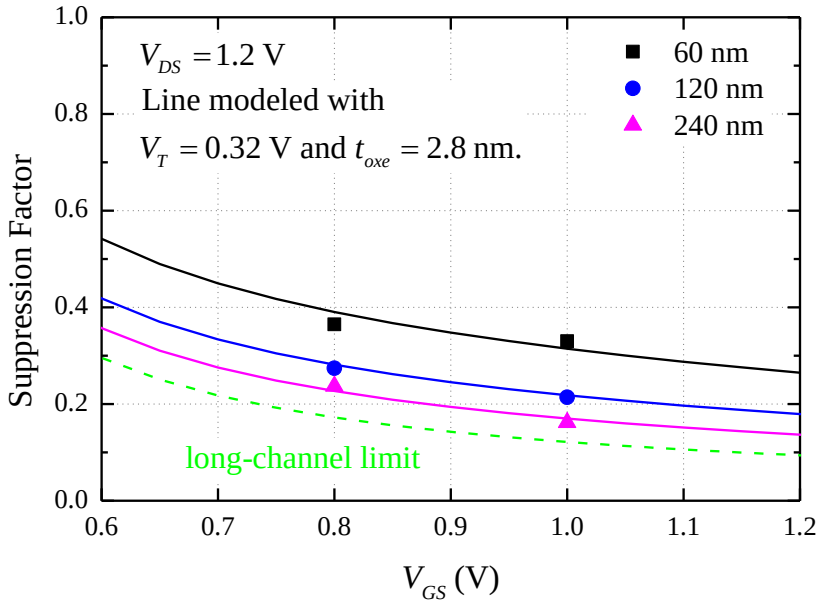


Fig. 5.5. Calculated (lines) and experimental (symbols) values of the channel noise suppression factor F versus V_{GS} for n-type MOSFETs in 65-nm technology. The experimental data are extracted from [160] for devices with channel lengths of 60 nm, 120 nm and 240 nm, respectively. The lines with same colors are calculated using (5.31). The green dash line is for the long-channel limit when the channel length is ∞.

channel lengths of 60 nm, 120 nm, and 240 nm, respectively. The lines with the corresponding colors shows the modeled suppression factors using (5.31) with V_T = 0.32 V and t_{oxe} = 2.8 nm. The green dashed line is for the long-channel limit when the channel length is ∞.

In addition, using extracted data from [127], we plot the experimental suppression factors of a 40-nm n-type MOSFET in symbols in Fig. 5.6. The modeled suppression factors using (5.31) with of V_T = 0.25 V and t_{oxe} = 2.4 nm for 3 channel lengths of 40, 60, and 200 nm, as well as the long-channel limit, are plotted in lines and green dashed line,

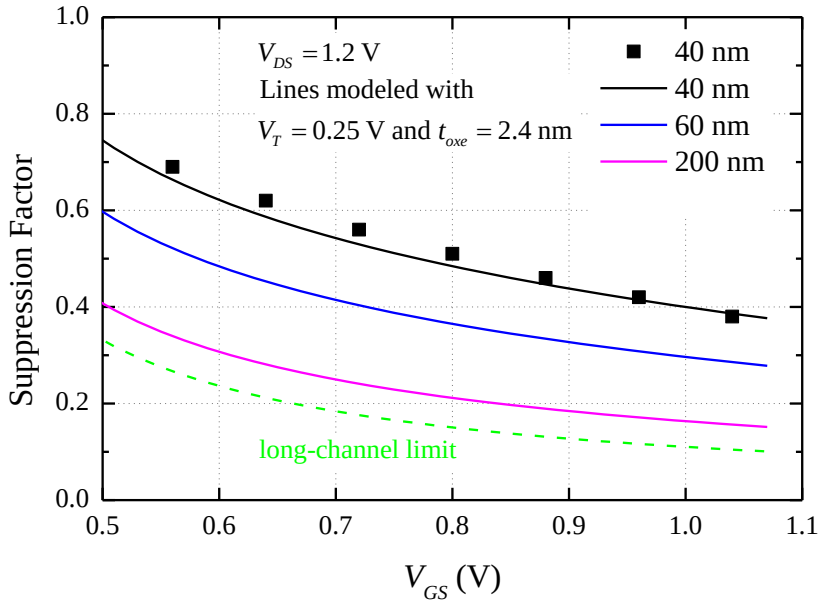


Fig. 5.6. Calculated (lines) and experimental (symbols) values of the channel noise suppression factor F versus V_{GS} for n-type MOSFETs in 40-nm technology. The experimental data is extracted from [127] and only available for L = 40 nm. For easy comparison and prediction, we plot modeled suppression factors for 3 more channel lengths of 40, 60, and 200 nm, as well as the long-channel limit. (© 2017 IEEE.)

respectively. In both figures, a good fitting is achieved with reasonable parameter values. It is a promising sign that the proposed suppression factor expression (5.31) may still work for other sub-100-nm MOSFETs. Furthermore, once the two process parameters are known, we can easily to predict the channel noise level for different biases and geometries using (5.31).

There are limitations in (5.31). Based on the concept of shot noise suppression, the suppression factor should have an upper limit of 1, which corresponds to unsuppressed shot noise. However, no such limit is suggested mathematically in (5.31), which means the use

of it should be limited to keep the result physical. First, when V_{GS} approaches V_T, and the channel becomes moderately inverted, the first term in (5.31) would increase dramatically and push F out of limit. This is the reason that we explicitly state that it should only be used when the MOSFET is in strongly inverted saturation region. Since (5.10) works in all bias regions, and the drain current I_D can be extended to all regions, further work can be done by extending (5.31) to work in more regions of operation.

Second, if we further decrease the channel length, the second term in (5.31) would eventually be big enough to contradict F with the upper limit of 1. In fact, the same problem of scaling also challenges (5.10) due to the presence of L_{elec}. In this situation, (5.31) and many existing transistor theories which are based on the assumption of thermal equilibrium becomes invalid. Fortunately, we think it is still valid to view the channel noise as suppressed shot noise, which is inherently compatible with non-equilibrium. Therefore, (5.32) will still be valid, but a new analytical equation for the suppression factor F is required.

5.6 Conclusion

In this chapter, we considered channel noise of MOSFETs as suppressed shot noise, which is not in conflict with traditional thermal-noise based approaches, but has some attractive advantages. First, the derived suppression factor is easy to use and accurately predicts the channel noise. It only depends on two process parameters - V_T and t_{oxe}, which are easily obtained from DC modeling and experiments. It avoids the use of the CLM model, which has a big impact on the channel noise but is complicated to obtain the parameter

values. Instead, the proposed suppression factor method leverages the drain current, which already takes care of the CLM effect. Second, the concept of suppressed shot noise is better equipped when facing the scaling challenges for noise modeling. Unlike traditional thermal noise based approaches, which constantly struggle with the "excess noise", the suppression factor concept has roots in mesoscopic physics research and has already been applied to noise modeling for various novel transistors.

Chapter 6.

Conclusion and Future Work

High-frequency noise modeling and characterization for nanoscale MSOFETs is an essential driving force for highly scaled CMOS technology to be used in RF applications. In this thesis, several important issues regarding noise modeling and characterization for nanoscale MOSFETs are studied. In this chapter, we conclude the work presented in previous chapters and also suggest future work that can be done to further the study on this topic.

6.1 Conclusions

In Chapter 2, a problem in noise characterization for on-wafer noise measurements is solved. The existing noise factor deembedding theories only work with two-port and three-port networks. We present a noise factor deembedding algorithm to extend existing theories to work for an active device surrounded by a four-port parasitic network. It also enables a new approach to obtain the intrinsic noise parameters by performing deembedding first and optimization last for on-wafer noise measurements. Using the new approach, it is possible to remove non-physical intrinsic noise factors before feeding them to optimization, which reduces the chance of the final noise parameters being non-physical. Statistical evaluation

of the new approach shows that it is more robust and produces less non-physical situations than the traditional optimization-first, deembedding-last approach. For the two evaluated designs of the 90-nm and 28-nm devices, the proposed approach works more effectively for a DUT with low gain and large parasitics in the 90-nm design.

In Chapter 3, we evaluate the noise performance for future nanoscale MOSFETs using the equivalent noise sheet resistance R_{nsh}, which is a figure-of-merit signifying the noise-to-signal ratio of the intrinsic device. It captures both the channel noise level and the transconductance of the transistor. Through experimental data, it is found that the noise performance of modern CMOS technologies degrades as the channel length decreases. Channel engineering techniques, for example strain-engineering, can improve the mobility of the carriers but do not change the saturation velocity, and thus have a deleterious effect on the noise performance. The reduction in power supply voltage is also negatively related to the noise performance. Increasing the gate oxide capacitance by reducing the oxide thickness or using high-k materials becomes the possible solution for future low noise technologies.

In Chapter 4, a novel Z-parameter-based approach to extract the gate resistance at low frequencies is presented. The gate resistance is important not only in DC and AC modeling, but also in noise modeling of MOSFETs. Not only does it generate noise by itself, it also plays an important role in the extraction of the noise sources of the intrinsic transistor. The proposed gate resistance extraction method is demonstrated to be statistically more stable and robust than other published methods in the literature. This statistical robustness and

stability are particularly important for the future technology nodes with smaller feature size and larger process variations. The dependence of the extracted gate resistance on channel length and process is analyzed and discussed. The accuracy of the proposed method reduces on the devices which have low critical frequencies. This issue ceases to exist for sub-100-nm devices which have sufficiently large critical frequencies. A method to extract the gate contact resistance is presented. The extracted contact resistance represents around half of the total gate resistance for the 55-nm *n*-type FETs.

In Chapter 5, a new perspective on MOSFET channel noise is proposed and evaluated. The channel noise is the most dominant noise source in the intrinsic transistor. Traditional modeling efforts are based on thermal noise that assumes the carriers in the channel resides in thermal equilibrium. However, this assumption is challenged by the physics of the modern-day extremely short-channel transistors. In light of this, we propose to consider the channel noise as suppressed shot noise and use the suppression factor F to quantify the level of suppression. It is not in conflict with traditional thermal-noise based approaches, which still works well with longer transistors, but has some attractive advantages. First, the derived suppression factor is easy to use and accurately predicts the channel noise. It only depends on two process parameters - V_T and t_{oxe}, which are easily obtained from DC modeling and experiments. It avoids the use of the CLM effect, which has a big impact on the channel noise but is complicated to model. Instead, the proposed suppression factor method leverages the drain current, which already takes care of the CLM effect. Second, the concept of suppressed shot noise is better equipped when facing the scaling challenges for noise modeling. Unlike traditional thermal noise based approaches, which constantly

struggle with the “excess noise”, the suppression factor concept has roots in mesoscopic physics research has already been applied to noise modeling for various novel transistors.

6.2 Future Work

Noise characterization and modeling of nanoscale MOSFETs is a relatively big topic. Its foundation is already built by contributions from many researchers. The work of this thesis only covers several issues in it. As the downscaling continues, the transition from drift-diffusion transport to ballistic transport is happening to current and future nanoscale transistors. The physics that governs how the transistors work is also shifting from classical theories to quantum mechanical theories. This causes a major challenge not only for DC and AC modeling, but also for noise modeling of the transistors. In this sense, this work is only one step towards that direction. More research efforts are needed to complete and unite the theory that can guide the future development of the semiconductor industry.

For the specific issues tackled in this work, there are also room for improvements and follow-up work, which are listed as follows.

1) For the proposed noise factor deembedding algorithm in Chapter 2, it works for the specific case of a two-port active device surrounded by a four-port passive network. A generalized theory or framework of noise factor deembedding for multi-port networks may be possible following the same spirit.
2) For the optimization last approach for noise parameter deembedding introduced in Chapter 2, it is evaluated and compared to traditional approach using Lane’s

optimization algorithm. However, the literature has reported many different optimization algorithms and they do have some impact on the final deembedded noise parameters. Therefore, it is possible to investigate the impact of those optimization algorithms on the two approaches and search for the theoretical proof for the statistical improvement of the proposed approach.

3) For the suppression factor expression derived in Chapter 5, it suggests no upper limit when V_{GS} approaches V_T and L_{eff} approaches zero. This cannot be true since known mechanisms in transistors can only suppress the noise with respect to full shot noise, which means F has an upper limit of 1. For the issue of V_{GS} approaching V_T, work can be done to extend the F expression to full range of operation at any bias. At the same time, V_{DS} dependence of F can also be investigated within the same theoretical framework.
4) In another aspect of the suppression factor in Chapter 5, the derivation of the analytical expression starts from a thermal noise based approach. However, the thermal-noise based approach may not be valid for extremely short transistors or alternative transistor technologies. In this case, it may be necessary to start from the physics of the suppression mechanisms and explore possible analytical equation for F that connects to macroscopic device parameters.

References

[1] A. van der Ziel, *Noise in Solid State Devices and Circuits*. Wiley New York, 1986.

[2] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Proc. IEEE*, vol. 86, no. 1, pp. 82–85, Jan. 1998.

[3] R. R. Schaller, "Moore's law: past, present and future," *Spectrum, IEEE*, vol. 34, no. 6, pp. 52–59, 1997.

[4] T. Skotnicki, J. A. Hutchby, and T. King, "The End of CMOS Scaling," *IEEE Circuits Devices Mag.*, vol. 21, no. 1, pp. 16–26, 2005.

[5] Moore's Law Diagram, 2004. [Online]. Available: https://en.wikipedia.org/wiki/File:Moore_Law_diagram_(2004).jpg

[6] *International Technology Roadmap for Semiconductors* (*ITRS*), 2012. [Online]. Available: http://www.itrs2.net/itrs-reports.html.

[7] L. A. Samoska, "An Overview of Solid-State Integrated Circuit Amplifiers in the Submillimeter-Wave and THz Regime," *IEEE Trans. Terahertz Sci. Technol.*, vol. 1, no. 1, pp. 9–24, Sep. 2011.

[8] C. Hu, *Modern Semiconductor Devices for Integrated Circuits*, Prentice Hall, 2010.

[9] D. Hisamoto *et al.*, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Trans. Electron Devices*, vol. 47, no. 12, pp. 2320–2325, Dec. 2000.

[10] G. K. Celler and S. Cristoloveanu, "Frontiers of silicon-on-insulator," *J. Appl. Phys.*, vol. 93, no. 9, pp. 4955–4978, 2003.

[11] J.-P. Colinge *et al.*, "Nanowire transistors without junctions," *Nat. Nanotechnol.*, vol. 5, no. 3, pp. 225–229, 2010.

[12] Y. Yoon, K. Ganapathi, and S. Salahuddin, “How good can monolayer MoS_2 transistors be?” *Nano Lett.*, vol. 11, no. 9, pp. 3768–73, 2011.

[13] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai, “Ballistic carbon nanotube field-effect transistors.,” *Nature*, vol. 424, no. 6949, pp. 654–657, 2003.

[14] K. Natori, “Ballistic/quasi-ballistic transport in nanoscale transistor,” *Appl. Surf. Sci.*, vol. 254, no. 19, pp. 6194–6198, Jul. 2008.

[15] M. S. Lundstrom and D. A. Antoniadis, “Compact Models and the Physics of Nanoscale FETs,” *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 225–233, Feb. 2014.

[16] P. H. Woerlee *et al.*, “RF-CMOS performance trends,” *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776–1782, 2001.

[17] Y. Cheng, C.-H. Chen, M. Matloubian, and M. J. Deen, “High-frequency small signal AC and noise modeling of MOSFETs for RF IC design,” *IEEE Trans. Electron Devices*, vol. 49, no. 3, pp. 400–408, 2002.

[18] C.-H. Chen, M. J. 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 Trans. Electron Devices*, vol. 48, no. 12, pp. 2884–2892, 2001.

[19] A. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers-van Duijnhoven, and V. C. Venezia, “Noise modeling for RF CMOS circuit simulation,” *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618–632, Mar. 2003.

[20] G. D. J. Smit, A. J. Scholten, R. M. T. Pijper, L. F. Tiemeijer, R. van der Toorn, and D. B. M. Klaassen, “RF-Noise Modeling in Advanced CMOS Technologies,” *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 245–254, Feb. 2014.

[21] **X. Chen**, C.-H. Chen, R. Lee, D. C. Chen, and D. Y. Wu, “Direct De-embedding of Noise Factors for On-Wafer Noise Measurement,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 3, pp. 916-922, March 2017.

[22] C.-H. Chen, **X. Chen**, D. Y. Wu, and C. S. Chen, "Future low-noise technologies for RF, analog and mixed-signal integrated circuits," *2015 IEEE 11th International Conference on ASIC (ASICON)*, Chengdu, 2015, pp. 1-4.

[23] **X. Chen**, N. Misljenovic, B. Hosein, C.-H. Chen, and C. Tsironis, "Calibration of a Noise Receiver Taking Care of its Gain Variations," *The 23rd International Conference on Noise and Fluctuations (ICNF 2015)*, Xi'an China (2-6 June, 2015).

[24] **X. Chen**, M. K. Tsai, C.-H. Chen, R. Lee, and D. C. Chen, "Extraction of Gate Resistance in Sub-100-nm MOSFETs with Statistical Verification," *IEEE Transactions on Electron Devices,* vol. 61, no. 9, pp. 3111-3117, Sept. 2014.

[25] **X. Chen**, C.-H. Chen, and M. J. Deen, "Shot Noise Suppression Factor for Nanoscale MOSFETs Working in the Saturation Region," *The 24th International Conference on Noise and Fluctuations (ICNF 2017)*, Vilnius, Lithuania (June 20-23, 2017).

[26] A. A. Abidi and J. C. Leete, "De-embedding the noise figure of differential amplifiers," *IEEE J. Solid-State Circuits*, vol. 34, no. 6, pp. 882–885, 1999.

[27] M. J. Deen and C.-H. Chen, "The impact of noise parameter de-embedding on the high frequency noise modeling of MOSFETs," *IEEE International Conference on Microelectronic Test Structures* (*ICMTS '99*), Goteborg, Sweden, pp. 34–39, 1999.

[28] Q. Liang, J. D. Cressler, G. Niu, Y. Lu, G. Freeman, D. C. Ahlgren, R. M. Malladi, K. Newton, and D. L. Harame, "A simple four-port parasitic deembedding methodology for high-frequency scattering parameter and noise characterization of SiGe HBTs," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 11, pp. 2165–2174, 2003.

[29] 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 Trans. Microw. Theory Techn.*, vol. 40, no. 11, pp. 2013–2024, 1992.

[30] V. Rizzoli and A. Lipparini, “Computer-aided noise analysis of linear multiport networks of arbitrary topology,” *IEEE Trans. Microw. Theory Techn.*, vol. 33, no. 12. pp. 1507–1512, 1985.

[31] IRE Subcommittee 7.9 on Noise, “Representation of noise in linear two-ports,” *Proc. IRE*, vol. 48, pp. 69-74, Jan. 1960.

[32] R. Q. Lane, “The determination of device noise parameters,” *Proc. IEEE*, vol. 57, no. 8, pp. 1461–1462, Aug. 1969.

[33] M. S. Gupta, “Determination of the noise parameters of a linear 2-port,” *Electron. Lett.*, vol. 6, no. 17, pp. 543–544, Aug. 1970.

[34] G. Caruso and M. Sannino, “Computer-aided determination of microwave two-port noise parameters,” *IEEE Trans. Microw. Theory Techn.*, vol. MTT- 26, no. 9, pp. 639–642, Sept. 1978.

[35] M. Mitama and H. Katoh, “An improved computational method for noise parameter measurement,” *IEEE Trans. Microw. Theory Techn.*, vol. MTT- 27, no. 6, pp. 612–615, Jun. 1979.

[36] M. Sannino, “On the determination of device noise and gain parameters,” *Proc. IEEE*, vol. 67, no. 9, pp. 1364–1366, Sept. 1979.

[37] G. I. Vasilescu, G. Alquie, and M. Krim, “Exact computation of two-port noise parameters,” *Electron. Lett.*, vol. 25, no. 4, pp. 292–293, Feb. 16, 1988.

[38] J. M. O’Callaghan and J. P. Mondal, “A vector approach for noise para- meter fitting and selection of source admittances,” *IEEE Trans. Microw. Theory Techn.*, vol. 39, no. 8, pp. 1376–1382, Aug. 1991.

[39] J. W. Archer and R. A. Batchelor, “Fully automated on-wafer noise characterization of GaAs MESFET’s and HEMT’s,” *IEEE Trans. Microw. Theory Techn.*, vol. 40, no. 2, pp. 209–216, Feb. 1992.

[40] A. Boudiaf and M. Laporte, "An accurate and repeatable technique for noise parameter measurements," *IEEE Trans. Instrum. Meas.*, vol. 42, no. 2, pp. 532–537, Apr. 1993.

[41] W. Wiatr and D. K. Walker, "Systematic errors of noise parameter determination caused by imperfect source impedance measurement," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 2, pp. 696–700, Apr. 2005.

[42] C.-H. Chen, "Accuracy issues of on-wafer noise measurements," *Fluctuation and Noise Letters*, vol. 8, issue 3-4, pp. L281-L303, 2008.

[43] P. J. van Wijnen, H. R. Claessen, and E. A. Wolsheimer, "A new straightforward calibration and correction procedure for 'on wafer' high frequency S-parameter measurements (45 MHz – 18 GHz)," in *IEEE Bipolar Circuits Technol. Meeting*, 1987, pp. 70–73.

[44] M. C. A. M. Koolen, J. A. M. Geelen, and M. P. J. G. Versleijen, "An improved de-embedding technique for on-wafer high-frequency characterization," in *IEEE Bipolar Circuits Technol. Meeting*, 1991, pp. 188–191.

[45] C.-H. Chen and M. J. Deen, "High frequency noise of MOSFETs I - modeling," *Solid. State. Electron.*, vol. 42, no. 11, pp. 2069–2081, 1998.

[46] C. E. Biber, M. L. Schmatz, T. Morf, U. Lott, E. Morifuji, and W. Bachtold, "Technology independent degradation of minimum noise figure due to pad parasitics," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, 1998, pp. 145–148.

[47] M. H. Cho, G. W. Huang, Y. H. Wang, and L. K. Wu, "A scalable noise De-embedding technique for on-wafer microwave device characterization," *IEEE Microw. Wirel. Components Lett.*, vol. 15, no. 10, pp. 649–651, 2005.

[48] C.-H. Chen and M. J. Deen, "A general noise and *S*-parameter de-embedding procedure for on-wafer high-frequency noise measurements of MOSFETs," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 5, pp. 1004 - 1005, May 2001.

[49] H. Rothe and W. Dahlke, "Theory of noisy fourpoles," *Proc. IRE*, vol. 2, 1956.

[50] R. Q. Twiss, “Nyquist’s and thevenin's theorems generalized for nonreciprocal linear networks,” *J. Appl. Phys.*, vol. 26, no. 5, pp. 599–602, 1955.

[51] H. Hillbrand and P. H. Russer, “An efficient method for computer-aided noise analysis of linear amplifier networks,” *IEEE Trans. Circuit Syst.*, vol. CAS-23, pp. 235–238, Apr. 1976.

[52] C. C. Enz and G. C. Temes, “Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization,” *Proc. IEEE*, vol. 84, no. 11, pp. 1584–1614, 1996.

[53] R. Kapusta, H. Zhu, and C. Lyden, “Sampling circuits that break the kT/C thermal noise limit,” *IEEE J. Solid-State Circuits*, vol. 49, no. 8, pp. 1694–1701, 2014.

[54] Y. Chiu, B. Nikolic, and P. R. Gray, “Scaling of analog-to-digital converters into ultra-deep-submicron CMOS,” in *Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005*, 2005, pp. 375–382.

[55] T. Sundstrom, B. Murmann, and C. Svensson, “Power dissipation bounds for high-speed Nyquist analog-to-digital converters,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 56, no. 3, pp. 509–518, 2009.

[56] L. B. Kish, “End of Moore’s law: thermal (noise) death of integration in micro and nano electronics,” *Phys. Lett. A*, vol. 305, no. 3, pp. 144–149, 2002.

[57] R. Poujois, B. Baylac, D. Barbier, and J. Ittel, “Low-level MOS transistor amplifier using storage techniques,” in *1973 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, 1973, vol. XVI, pp. 152–153.

[58] A. J. Williams, R. E. Tarpley, and W. R. Clark, “D-C Amplifier Stabilized for Zero and Gain,” *Trans. Am. Inst. Electr. Eng.*, vol. 67, no. 1, pp. 47–57, Jan. 1948.

[59] K. C. Hsieh and P. Gray, “A low-noise chopper-stabilized differential switched-capacitor filtering technique,” in *1981 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, 1981, vol. XXIV, pp. 128–129.

[60] R. P. Jindal, "High frequency noise in fine line NMOS field effect transistors," in *1985 International Electron Devices Meeting*, 1985, vol. 31, pp. 68–71.

[61] P. Klein, "An analytical thermal noise model of deep submicron MOSFET's," *IEEE Electron Device Lett.*, vol. 20, no. 8, pp. 399–401, Aug. 1999.

[62] C.-H. Chen and M. J. Deen, "Channel noise modeling of deep submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 8, pp. 1484–1487, Aug. 2002.

[63] A. O. Adan, M. Koyanagi, and M. Fukumi, "Physical Model of Noise Mechanisms in SOI and Bulk-Silicon MOSFETs for RF Applications," *IEEE Trans. Electron Devices*, vol. 55, no. 3, pp. 872–880, Mar. 2008.

[64] A. S. Roy and C. C. Enz, "Compact modeling of thermal noise in the MOS transistor," *IEEE Trans. Electron Devices*, vol. 52, no. 4, pp. 611–614, Apr. 2005.

[65] C.-H. Chen, R. Lee, G. Tan, D. C. Chen, P. Lei, and C. S. Yeh, "Equivalent Sheet Resistance of Intrinsic Noise in Sub-100-nm MOSFETs," *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2215–2220, Aug. 2012.

[66] M. J. Deen, C.-H. Chen, S. Asgaran, G. A. Rezvani, J. Tao, and Y. Kiyota, "High-Frequency Noise of Modern MOSFETs: Compact Modeling and Measurement Issues," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2062–2081, Sep. 2006.

[67] H. E. Halladay and A. van der Ziel, "On the high frequency excess noise and equivalent circuit representation of the MOS-FET with n-type channel," *Solid. State. Electron.*, vol. 12, no. 3, pp. 161–176, 1969.

[68] C.-H. Chen, M. J. 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 Trans. Electron Devices,* vol. 48, no. 12, pp. 2884–2892, 2001.

[69] C. G. Sodini, P.-K. Ko, and J. L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Devices*, vol. 31, no. 10, pp. 1386–1393, Oct. 1984.

[70] M. V. Dunga *et al.*, *BSIM4.6.0 MOSFET Model – User's Manual*, UC Berkeley (2006).

[71] T. Yamada, J.-R. Zhou, H. Miyata, and D. K. Ferry, "In-plane transport properties of Si/Si1-xGex structure and its FET performance by computer simulation," *IEEE Trans. Electron Devices*, vol. 41, no. 9, pp. 1513–1522, Sep. 1994.

[72] T. Krishnamohan, Z. Krivokapic, K. Uchida, Y. Nishi, and K. C. Saraswat, "High-mobility ultrathin strained Ge MOSFETs on bulk and SOI with low band-to-band tunneling leakage: experiments," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 990–999, May 2006.

[73] S. M. Sze., *Physics of Semiconductor Devices,* John Wiley & Sons, 2nd ed., p. 46 (1985).

[74] L. Chang, M. Ieong, and M. Yang, "CMOS circuit performance enhancement by surface orientation optimization," *IEEE Trans. Electron Devices*, vol. 51, no. 10, pp. 1621–1627, Oct. 2004.

[75] S. Mayuzumi *et al.*, "Comparative study between Si (110) and (100) substrates on mobility and velocity enhancements for short-channel highly-strained PFETs," in *2009 Symposium on VLSI Technology*, 2009, pp. 14–15.

[76] G. Dewey *et al.*, "Carrier Transport in High-Mobility III-V Quantum-Well Transistors and Performance Impact for High-Speed Low-Power Logic Applications," *IEEE Electron Device Lett.*, vol. 29, no. 10, pp. 1094–1097, Oct. 2008.

[77] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed., Cambridge Univ. Press, 2004.

[78] Y. Sun *et al.*, "Self-aligned III-V MOSFETs: Towards a CMOS compatible and manufacturable technology solution," in *2013 IEEE International Electron Devices Meeting*, 2013, p. 2.7.1-2.7.4.

[79] Y. Sun *et al.*, "High-performance CMOS-compatible self-aligned In0.53Ga0.47As MOSFETs with GMSAT over 2200 S/m at VDD = 0.5 V," in *2014 IEEE International Electron Devices Meeting*, 2014, p. 25.3.1-25.3.4.

[80] J. Lin, X. Zhao, T. Yu, D. A. Antoniadis, and J. A. del Alamo, "A new self-aligned quantum-well MOSFET architecture fabricated by a scalable tight-pitch process," in *2013 IEEE International Electron Devices Meeting*, 2013, p. 16.2.1-16.2.4.

[81] T. W. Kim *et al.*, "ETB-QW InAs MOSFET with scaled body for improved electrostatics," in *2012 International Electron Devices Meeting*, 2012, p. 32.3.1-32.3.4.

[82] S. Lee *et al.*, "Highly Scalable Raised Source/Drain InAs Quantum Well MOSFETs Exhibiting ION = 482 μA/μm at IOFF= 100 nA/μm and Vdd= 0.5 V," *IEEE Electron Device Lett.*, vol. 35, no. 6, pp. 621–623, Jun. 2014.

[83] M. J. H. van Dal *et al.*, "Demonstration of scaled Ge p-channel FinFETs integrated on Si," in *2012 International Electron Devices Meeting*, 2012, p. 23.5.1-23.5.4.

[84] C.-H. Chen, D. Chen, R. Lee, P. Lei, and D. Wan, "Thermal noise modeling of nano-scale MOSFETs for mixed-signal and RF applications," in *Proceedings of the IEEE 2013 Custom Integrated Circuits Conference*, 2013, pp. 1–8.

[85] C.-H. Chen, D. Y. Wu, Y. C. Cheng, and C. S. Chen, "Future CMOS technology for low noise integrated circuit designs," *2014 12th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, 2014, pp. 1–4.

[86] C.-H. Chen, D. Y. Wu, and M. K. Tsai, "The Impact of Material Property on the Noise Performance of Future Nanoscale Field-Effect-Transistors," *Collaborative Conference on 3D and Materials Research (CC3DMR) 2015*, Busan, South Korea, pp. 277-280.

[87] C.-H. Jan *et al*., "RF CMOS technology scaling in high-*k*/metal gate era for RF SoC (System-on-Chip) applications," *IEDM Tech. Dig*., pp. 604 – 607, 2010.

[88] C.-H. Chen *et al.*, "Equivalent sheet resistance of intrinsic noise in sub-100nm MOSFETs," *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2215 – 2220, Aug. 2012.

[89] Keysight Technologies, *Noise Figure Measurement Accuracy – The Y-Factor Method*, Keysight Technologies Application Note 5952-3706E.

[90] L. F. Tiemeijer, R. Havens, R. Kort, and A. J. Scholten, "Improved Y-factor method for wide-band on-wafer noise parameter measurements," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 9, pp. 2917 – 2925, 2005.

[91] C.-H. Chen, Y. L. Wang, M. Bakr, and Z. Zeng, "Novel noise parameter determination for on-wafer microwave noise measurements," *IEEE Trans. Instrum. Meas.*, vol. 57, issue 11, pp. 2462 – 2471, Nov. 2008.

[92] *On-Wafer Noise Parameter Measurements Using Cold-Noise Source and Automatic Receiver Calibration*, 1994, Focus Microwaves Inc. Application Note 19.

[93] Leonid Belostotski, "A calibration method for RF and microwave noise sources," *IEEE Trans. Microwave Theory Tech.*, vol. 59, no. 1, pp. 178 –187, 2011.

[94] Z. Wang, P.-Y. Chiang, P. Nazari, C.-C. Wang, Z. Chen, and P. Heydari, "A 210 GHz fully integrated differential transceiver with fundamental- frequency VCO in 32 nm SOI CMOS," in *IEEE Int. Solid-State Circuit Conf.* (*ISSCC*) *Dig. Tech. Papers*, 2013, pp. 136-137.

[95] *International Technology Roadmap for Semiconductors* (*ITRS*), 2012. [Online]. Available: http://www.itrs2.net/itrs-reports.html.

[96] C.-H. Chen, R. Lee, G. Tan, D. C. Chen, P. Lei, and C. S. Yeh, "Equivalent sheet resistance of intrinsic noise in sub-100nm MOSFETs," *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2215–2220, 2012.

[97] W. S. Kwan, C.-H. Chen, and M. J. Deen, "Hot-Carrier Effects on RF Noise Characteristics of LDD NMOSFET," *Journal of Vacuum Science and Technology A*, vol. 18(2), pp. 765–769, March/April 2000.

[98] S. Naseh, M. J. Deen, and C.-H. Chen, "Effects of hot-carrier stress on the performance of CMOS low noise amplifiers," *IEEE Trans. Device and Materials Reliability*, vol. 5, no. 3, pp. 501–508, Sept. 2005.

[99] S. Naseh, M. J. Deen, and C.-H. Chen, "Hot-carrier reliability of submicron NMOSFETs and integrated NMOS low noise amplifiers," *Microelectronics Reliability*, vol. 46, pp. 201–212, 2006.

[100] C.-H. Chen and M. J. Deen, "Channel noise modeling of deep sub-micron MOSFETs," *IEEE Trans. Electron Device*, vol. 49, no. 8, pp. 1484–1487, Aug. 2002.

[101] C.-H. Chen, F. Li, and Y. Cheng, "MOSFET drain and induced-gate noise modeling and experimental verification for RF IC design," in *IEEE International Conference on Microelectronics Test Structures*, Awaji, Japan, 2004, pp. 51-56.

[102] M. J. Deen, C.-H. Chen, S. Asgaran, G. Ali Rezvani, J. Tao, and Y. Kiyota, "High-frequency noise of modern MOSFETs: compact modeling and measurement issues," *IEEE Trans. Electron Devices*, vol. 53, no.9, pp. 2062–2081, Sept. 2006.

[103] S. Lee, H. K. Yu, C. S. Kim, J. G. Koo, and K. S. Nam, "A novel approach to extracting small-signal model parameters of silicon MOSFETs," *IEEE Microw. Guided Wave Lett.*, vol. 7, no. 3, pp. 75–77, May 1997.

[104] X. Jin, J.-J. Ou, C.-H. Chen, W. Liu, M. J. Deen, P. R. Gray, and C. Hu, "An effective gate resistance model for CMOS RF and noise modeling," in *IEDM Tech Dig*., 1998, pp. 961–964.

[105] S. H.-M. Jen, C. C. Enz, D. R. Pehlke, M. Schröter, and B. J. Sheu, "Accurate modeling and parameter extraction for MOS transistors valid up to 10 GHz," *IEEE Trans. Electron Devices*, vol. 46, no. 11, pp. 2217-2227, Nov. 1999.

[106] C. C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, no. 2, pp. 186-201, Feb. 2000.

[107] Y. Cheng and M. Matloubian, “On the high-frequency characteristics of substrate resistance in RF MOSFETs,” *IEEE Electron Device Lett*., vol. 21, no. 12, pp. 604-606, Dec. 2000.

[108] Y. Cheng and M. Matloubian, “High frequency characterization of gate resistance in MOSFETs,” *IEEE Electron Device Lett.*, vol. 22, no. 2, pp. 98–100, Feb. 2001.

[109] S. H. Lee and H. K. Yu, “Parameter extraction technique for the small-signal equivalent circuit model of microwave silicon MOSFETs,” in *Proceedings IEEE/Cornell Conference on Advanced Concepts in High Speed Semiconductor Devices and Circuits CORNEL-97*, 1997, pp. 182–191.

[110] S. H. Lee, C. S. Kim, and H. K. Yu, “A small-signal RF model and its parameter extraction for substrate effects in RF MOSFETs,” *IEEE Trans*. *Electron Devices*, vol. 48, no. 7, pp. 1374–1379, Jul. 2001.

[111] R. Torres-Torres, R. S. Murphy-Arteaga, and S. Decoutere, “MOSFET gate resistance determination,” *Electron. Lett.*, vol. 39, no. 2, pp. 248–250, Jan. 2003.

[112] Y. Cheng, M. J. Deen, and C.-H. Chen, “MOSFET modeling for RFIC design,” *IEEE Trans. Electron Devices*, vol. 52, pp. 1286–1303, 2005.

[113] M. Kang, I. M. Kang, Y. H. Jung, and H. Shin, “Separate extraction of gate resistance components in RF MOSFETs,” *IEEE Trans. Electron Devices*, vol. 54, no. 6, pp. 1459-1463, Jun. 2007.

[114] K.-L. Yeh and J.-C. Guo, “Narrow-width effect on high-frequency performance and RF noise of sub-40-nm multifinger nMOSFETs and pMOSFETs,” *IEEE Trans. Electron Devices*, vol. 60, no. 1, pp. 109–116, Jan. 2013.

[115] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*, 3rd ed., New York, Oxford Univ. Press, 2011.

[116] D. M. M.-P. Schreurs, Y. Baeyens, B. K. J. C. Nauwelaers, W. D. Raedt, M. V. Hove, and M. V. Rossum, “S-parameter measurement based quasistatic large-signal cold

HEMT model for resistive mixer design," *Int. J. Microw. Millim.-Wave Comput.-Aided Eng.*, vol. 6, no. 4, pp. 250–258, Jul. 1996.

[117] K. J. Kuhn, M. D. Giles, D. Becher, P. Kolar, A. Kornfeld, R. Kotlyar, S. T. Ma, A. Maheshwari, and S. Mudanai, "Process technology variation," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197-2208, Aug. 2011.

[118] P. Magnone *et al.*, "Gate voltage and geometry dependence of the series resistance and of the carrier mobility in FinFET devices," *Microelectron. Eng.*, vol. 85, no. 8, pp. 1728–1731, 2008.

[119] P. H. Woerlee *et al.*, "RF-CMOS performance trends," *IEEE Trans. Electron Devices*, vol. 48(8), pp. 1776–1782, 2001.

[120] Y. Cheng, C.-H. Chen, M. Matloubian, M.J. Deen, "High-frequency small signal AC and noise modeling of MOSFETs for RF IC design,", *IEEE Trans. Electron Devices*, vol. 49(3), pp. 400-408, 2002.

[121] A. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers-van Duijnhoven, and V. C. Venezia, "Noise modeling for RF CMOS circuit simulation," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618–632, Mar. 2003.

[122] C.-H. Chen, M. J. Deen, "Channel noise modeling of deep submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 49(8), pp. 1484–1487, 2002.

[123] C.-H. Chen, "Noise characterization and modeling of MOSFETs for RF IC applications," Ph.D. thesis, McMaster University, 2002.

[124] M. J. Deen, C.-H. Chen, S. Asgaran, G. A. Rezvani, J. Tao, and Y. Kiyota, "High-Frequency Noise of Modern MOSFETs: Compact Modeling and Measurement Issues," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2062–2081, Sep. 2006.

[125] S. Asgaran, M. J. Deen, and C.-H. Chen, "Analytical modeling of MOSFETs channel noise and noise parameters," *IEEE Trans. Electron Devices*, vol. 51(12), pp. 2109-2114, 2004.

[126] A. S. Roy and C. C. Enz, “Compact modeling of thermal noise in the MOS transistor,” *IEEE Trans. Electron Devices*, vol. 52(4), pp. 611–614, 2005.

[127] G. D. J. Smit, A. J. Scholten, R. M. T. Pijper, L. F. Tiemeijer, R. van der Toorn, and D. B. M. Klaassen, “RF-noise modeling in advanced CMOS technologies,” *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 245–254, Feb. 2014.

[128] Y. M. Blanter and M. Büttiker, “Shot noise in mesoscopic conductors,” *Phys. Rep.*, vol. 336, no. 1–2, pp. 1–166, Sep. 2000.

[129] A. Betti, G. Fiori, and G. Iannaccone, “Shot noise suppression in quasi-one-dimensional field-effect transistors,” *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 2137–2143, Sep. 2009.

[130] G. Iannaccone, “Analytical and numerical investigation of noise in nanoscale ballistic field effect transistors,” *J. Comput. Electron.*, vol. 3, no. 3–4, pp. 199–202, Oct. 2004.

[131] T. González *et al.*, “Universality of the 1/3 shot-noise suppression factor in nondegenerate diffusive conductors,” *Phys. Rev. Lett.*, vol. 80, no. 13, pp. 2901–2904, Mar. 1998.

[132] M. S. Lundstrom and D. A. Antoniadis, “Compact models and the physics of nanoscale FETs,” *IEEE Trans. Electron Devices*, vol. 61(2), pp. 225–233, 2014.

[133] J. B. Johnson, “Thermal agitation of electricity in conductors,” *Phys. Rev.*, vol. 32, no. 1, p. 97, 1928.

[134] H. Nyquist, “Thermal agitation of electric charge in conductors,” *Phys. Rev.*, vol. 32, no. 1, p. 110, 1928.

[135] F. M. Klaassen and J. Prins, "Thermal noise of MOS transistors", *Philips Res. Reps.*, vol. 22, pp. 504-514, 1967.

[136] A. van der Ziel, “Noise in solid-state devices and lasers,” *Proc. IEEE*, vol. 58, no. 8, pp. 1178–1206, 1970.

[137] A. A. Abidi, "High-frequency noise measurements on FET's with small dimensions", *IEEE Trans. Electron Devices*, vol. ED-33, pp. 1801-1805, 1986.

[138] A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers-van Duijnhoven and V. C. Venezia, "Noise modeling for RF CMOS circuit simulation," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618-632, 2003.

[139] K. Han, H. Shin, and K. Lee, "Analytical drain thermal noise current model valid for deep submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 51(2), pp. 261–269, Feb. 2004.

[140] G. D. J. Smit, A. J. Scholten, R. M. T. Pijper, R. van Langevelde, L. F. Tiemeijer, and D. B. M. Klaassen, "Experimental Demonstration and Modeling of Excess RF Noise in Sub-100-nm CMOS Technologies," *IEEE Electron Device Lett.*, vol. 31, no. 8, pp. 884–886, Aug. 2010.

[141] W. Schottky, "Spontaneous current fluctuations in various conductors," *Ann. Phys.*, vol. 57, pp. 541–567, 1918.

[142] G. Reimbold and P. Gentil, "White noise of MOS transistors operating in weak inversion," *IEEE Trans. Electron Devices*, vol. 29, no. 11, pp. 1722–1725, 1982.

[143] R. Sarpeshkar, T. Delbruck, and C. A. Mead, "White noise in MOS transistors and resistors," *IEEE Circuits Devices Mag.*, vol. 9, no. 6, pp. 23–29, Nov. 1993.

[144] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1853–1864, 2003.

[145] Y. Yoon, K. Ganapathi, and S. Salahuddin, "How good can monolayer MoS_2 transistors be?" *Nano Lett.*, vol. 11, no. 9, pp. 3768–73, 2011.

[146] L. Liu, Y. Lu, and J. Guo, "On Monolayer MoS_2 Field-Effect Transistors at the Scaling Limit," *IEEE Trans. Electron Devices*, vol. 60, no. 12, pp. 4133–4139, 2013.

[147] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai, "Ballistic carbon nanotube field-effect transistors.," *Nature*, vol. 424, no. 6949, pp. 654–657, 2003.

[148] A. C. Seabaugh and Q. Zhang, “Low-voltage tunnel transistors for beyond CMOS logic,” *Proc. IEEE*, vol. 98, no. 12, pp. 2095–2110, 2010.

[149] S. Datta, H. Liu, and V. Narayanan, “Tunnel FET technology: A reliability perspective,” *Microelectron. Reliab.*, vol. 54, no. 5, pp. 861–874, 2014.

[150] D. Sarkar *et al.*, “A subthermionic tunnel field-effect transistor with an atomically thin channel,” *Nature*, vol. 526, no. 7571, pp. 91–95, Sep. 2015.

[151] P. K. Ko, R. S. Muller, and C. Hu, “A unified model for hot-electron currents in MOSFET’s,” in *1981 International Electron Devices Meeting*, 1981, vol. 27, pp. 600–603.

[152] A. Betti, G. Fiori, and G. Iannaccone, “Shot Noise Suppression in Quasi-One-Dimensional Field-Effect Transistors,” *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 2137–2143, Sep. 2009.

[153] G. M. Landauer and J. L. Gonzalez, “A compact noise model for carbon nanotube FETs,” in *2012 International Semiconductor Conference Dresden-Grenoble (ISCDG)*, 2012, pp. 53–56.

[154] R. Pandey, B. Rajamohanan, H. Liu, V. Narayanan, and S. Datta, “Electrical noise in heterojunction interband tunnel FETs,” *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 552–560, 2014.

[155] U. Fano, “Ionization yield of radiations. II. the fluctuations of the number of ions,” *Phys. Rev.*, vol. 72, no. 1, pp. 26–29, 1947.

[156] K.-Y. Toh, P.-K. Ko, and R. G. Meyer, “An engineering model for short-channel MOS devices,” *IEEE J. Solid-State Circuits*, vol. 23, no. 4, pp. 950–958, Aug. 1988.

[157] C Hu, Modern Semiconductor Devices for Integrated Circuits. Prentice Hall, 2010.

[158] K. Chen, H. C. Wann, P. K. Ko, and C. Hu, “The impact of device scaling and power supply change on CMOS gate performance,” *IEEE Electron Device Lett.*, vol. 17, no. 5, pp. 202–204, 1996.

[159] S. Takagi, M. Iwase, and A. Toriumi, “On the universality of inversion-layer mobility in n- and p-channel MOSFETs,” in *Technical Digest., International Electron Devices Meeting*, 1988, pp. 398–401.

[160] S. Wang, P. Su, K. Chen, K. Liao, and B. Chen, “Comprehensive Noise Characterization and Modeling for 65-nm MOSFETs for Millimeter-Wave Applications,” *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 4, pp. 740–746, Apr. 2010.