Title
Design studies of nanometer-gate low-noise amplifier near the limits of CMOS scaling

Author
Cai, Ming

Publication Date
2006

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DESIGN STUDIES OF NANOMETER-GATE LOW-NOISE AMPLIFIER NEAR THE LIMITS OF CMOS SCALING

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Electrical Engineering (Applied Physics)

by

Ming Cai

Committee in charge:

Professor Yuan Taur, Chair
Professor Peter M. Asbeck
Professor Yu-Hwa Lo
Professor Vitali Nesterenko
Professor Jan B. Talbot

2006

The dissertation of Ming Cai is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2006

DEDICATION

To my wife and mother

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

ACKNOWLEDGMENTS

The experience in UCSD is the most rewarding one for me. First and foremost, I would like to cordially thank my advisor Professor Yuan Taur for being the best advisor that a student could ever have, for being knowledgeable, insightful, supportive and considerate. Without his continuous guidance and mentoring, my research work towards this dissertation would never be possible. I am greatly indebted to his encouragement and help on my career development in the United States.

I wish to express my sincere gratitude to Professor S. S. Lau for bringing me to UCSD and giving me an exceptional educational experience during my stay in his group. I would also like to thank Professor Peter Asbeck and Yu-Hwa Lo for their generous guidance and valuable discussions. I am grateful to Professor Vitali Nesterenko and Jan Talbot for their efforts of serving on my committee.

This work has benefited from the contributions of many other individuals. Special thanks to Raj Samuel from SPAWAR Systems Center in San Diego for his support to my research project. Thanks are also due to Huaxin Lu, Minjian Liu, Weiyuan Lu, Nikhil Vaidya for their assiduous assistance in solving the CAD problems and sharing their knowledges.

Most of all, I thank my beloved wife, Jing Zhang. Without her love and understanding, I could not have finished my study at UCSD. I am forever grateful for her inspiration and encouragement. I thank my mother and parents-in-law who have given me endless spiritual support.

The text of Chapter Three to Six, in part, is a reprint of the material that has been submitted for publication: Cai, Ming; Liu, Minjian; Taur, Yuan. “A design study of nanometer-gate low-noise amplifier near the limit of CMOS scaling”, Solid-State Electronics; Cai, Ming; Liu, Minjian; Taur, Yuan. “Design study of power-efficient 60-GHz MOS low noise amplifiers near scaling limits”, Electronics Letters. The dissertation author was the primary author and the co-authors listed in these publications directed and supervised the research which forms the basis for these chapters.

VITA

1998	B. S. in Physics, Peking University, P. R. China
2001	M. S. in Physics, Peking University, P. R. China
2001-2006	Research Assistant, University of California, San Diego
2006	Ph. D. in Electrical Engineering, University of California, San Diego

PUBLICATIONS

1. **M. Cai**, M. Liu, and Y. Taur, "A design study of nanometer-gate low-noise amplifier near the limit of CMOS scaling", *Solid-State Electronics*, submitted.

2. **M. Cai**, M. Liu, and Y. Taur, "Design study of power-efficient 60-GHz CMOS low noise amplifiers near scaling limits", *Electronics Letters*, submitted.

3. **M. Cai**, M. Liu, and Y. Taur, "Noise analysis of a MOSFET RF amplifier using two-dimensional mixed-mode simulation", *Semiconductor Science and Technology*, in preparation.

4. M. Liu, **M. Cai**, and Y. Taur, "Scaling limit of CMOS supply voltage from noise margin considerations", *SISPAD 2006*, accepted.

5. M. Liu, **M. Cai**, B. Yu**,** and Y. Taur, "Effect of gate overlap and source/drain doping abruptness on 10 nm CMOS performance", *IEEE Trans. Electron Devices*, accepted.

6. P. Chen, P. K. Chu, T. Höchbauer, J.-K. Lee, M. Nastasi, D. Buca, S. Mantl, R. Loo, M.Caymax, T. Alford, J. W. Mayer, N. D. Theodore, **M. Cai**, B. Schmidt, and S. S. Lau, "Investigation of plasma hydrogenation and trapping mechanism for layer transfer", *Applied Physics Letters*, 86 (3), 031904 (2005).

7. P. Chen P. K. Chu, T. Höchbauer, M. Nastasi, D. Buca, S. Mantl, N. D. Theodore, T. L. Alford, J. W. Mayer, R. Loo, M. Caymax, **M. Cai**, and S. S.

Lau, "Plasma hydrogenation of strain-relaxed SiGe/Si heterostructure for layer transfer", *Applied Physics Letters*, 85 (21), 4944 (2004).

8. J. K. Lee, M. Nastasi, N. D. Theodore, A. Smalley, T.L. Alford, J. W. Mayer, **M. Cai,** and S. S. Lau. "Effects of hydrogen implantation temperature on ion-cut of silicon", *Journal of Applied Physics*, 96 (1), 280 (2004).

9. K. Henttinen, T. Suni, A. Nurmela, H. V. A. Luoto, I. Suni, V. M. Airaksinen, S. Karirinne, **M. Cai**, and S. S. Lau, "Transfer of thin Si layers by cold and thermal ion cutting", *Jounal of Materials Science: Materials in Electronics*, 14, 299 (2003).

10. F. Lu, D. Qiao, **M. Cai**, P. K. L. Yu, S. S. Lau, R. K. Y. Yu, L. S. Hung, C. P. Li, P. K. Chu, H. C. Chien, and Y. Liou, "Ion-cutting of Si onto glass by pulsed and direct-current plasma immersion ion implantation", *Journal of Vacuum Science Technology B*, 21 (5), 2109 (2003).

11. **M. Cai**, D. Qiao, L. S. Yu, S. S. Lau, C. P. Li, L. S. Hung, T. Haynes, K. Kenttinen, I. Suni, V. M. C. Poon, T. Marek and J. W. Mayer, "Single crystal Si layers on glass formed by ion cutting", *Journal of Applied Physics*, 92 (6), 3388 (2002).

ABSTRACT OF THE DISSERTATION

DESIGN STUDIES OF NANOMETER-GATE LOW-NOISE AMPLIFIER NEAR THE LIMITS OF CMOS SCALING

by

Ming Cai

Doctor of Philosophy in Electrical Engineering

(Applied Physics)

University of California, San Diego, 2006

Professor Yuan Taur, Chair

As many millimeter-wave (mm-wave) applications has been proposed for the next generation communication systems recently, CMOS has become a technology of choice due to its increasing RF performance by transistor scaling and many inherit

merits including low cost, low power and high level of integration. This dissertation presents a comprehensive design study of power-efficient CMOS low-noise amplifiers (LNAs) near the scaling limits.

Starting with an introduction to the mm-wave system applications and enabling technologies, the current status of research and development on technology candidates for mm-wave applications has been surveyed with the focus on the Si CMOS. The effect of CMOS scaling has been studied using state-of-the-art 40- and 20- nm MOSFETs, and 20-nm double-gate (DG) MOSFET (a generic form of FinFET), and a 10-nm DG MOSFET at the scaling limit. A two-dimensional device simulator TCAD is used to extract various RF performance parameters at 60 GHz. To take both the amplifier gain and noise figure into account, a new design methodology using noise measure as a figure of merit is developed.

Excellent noise performance has been achieved with the identified optimal biasing for power-efficient low-noise amplifications at 60 GHz. The effect of the gate resistance to the overall noise performance has been investigated, which has been found to be insignificant as long as the device width per gate finger is constrained. The performance trends of CMOS LNAs near the scaling limit has been studied with correlations made to the published hardware data. It is further shown that the RF performance of CMOS LNAs can be greatly enhanced by down scaling to 10 nm, using the more scalable structure of DG MOSFETs, with sub-1-dB noise figures at 60 GHz and beyond.

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

1.1.1 Millimeter-wave communication systems

Wireless applications, based in part on radio-frequency (RF) devices and integrated circuit (IC) technologies, have grown quickly to become significant markets for the semiconductor industry. Due to increasing demands for broadband wireless communication, high frequency bands such as microwave and millimeter-wave (mm-wave), approximately from 30 to 300 GHz, where there is sufficient spectrum are of growing interest. Recently 7 GHz of unlicensed bandwidth around 60 GHz was opened allowing for a variety of applications including cellular phone receivers, wireless local area networks (WLAN), wireless personal area networks (WPAN), automotive radar, gigabit wireless networks, and the like [1]. There are also increasing interests for the 94 GHz band arising from its applications for all weather landing and other security needs [2]. Moving to higher frequencies provides natural isolation from fast digital circuitry typical of today's technologies [3]. The greatest interest in mm-wave radio stems from the reduced antenna size and spacing at higher frequencies which make it possible to integrate the antenna or phased-array receiver, with the RF and digital circuitry on single silicon die [4]. This feature truly paves the road for wearable gigabit-rate multimedia platforms and mm-wave sensors by elegantly avoiding the potentially high cost of mm-wave packaging.

1.1.2 Technologies for millimeter wave applications

Until today, the fastest transistors are commonly fabricated using III/V-based technologies. Traditionally gallium arsenide (GaAs) and indium phosphide (InP) have demonstrated the capability of very high gain and ultra low noise at mm-wave frequencies below 100 GHz. For frequency range above 120 GHz, InP-based transistors are superior to those based on GaAs because the material they are made of provides higher electron mobility, higher saturation velocity and higher sheet electron density [5]. A cut-off frequency (f_T) up to 562 GHz has been reported for InAlAs/InGaAs pseudomorphic high electron mobility transistors (pHEMTs) based on InP [6]. However, due to the limited transistor yield, III/V-based circuits usually have high costs and allow only low circuit complexities.

In recent years, the speed gap between III/V- and silicon- (Si-) based technologies has been significantly reduced owing to the fast advances in Si-based technologies. A f_T of 375 GHz has been achieved using silicon germanium (SiGe) heterojunction bipolar transistors (HBTs) [7], and a silicon-on-insulator (SOI) nMOSFET with a f_T up to 330 GHz has been reported by IBM in IEDM 2005 [8]. These performances are expected to enable Si-based ultra wideband communication applications such as >1000 Gbps network systems and >100 GHz wireless systems. Along with the sufficient operation speed now available for mm-wave applications, Si-based technologies provide additional advantages such as CMOS-compatibility, large wafer size, high function density and system-on-a-chip (SOC) solution, all

leading to a strong cost efficiency. As an inevitable consequence of Moore's law, with the minimum feature size continuously and rapidly scaling downwards, CMOS technology will become a technology of choice for RF ICs [9]. Today 90-nm CMOS technology in production is capable of providing enough power gain with low noise figures at frequencies from 20 to 40 GHz while consuming a reasonable small amount of power. Future CMOS technologies of the 65-nm and 45-nm nodes, together with novel structures such as double-gate (DG) MOSFETs, are expected to provide even more gain with better noise performance at higher frequencies with less power consumption.

1.2 MOTIVATION

Geometrical scaling has resulted in a continuing improvement in the RF performance of CMOS transistors. As RF CMOS comes of age, performance metrics such as cut-off frequency (f_T), maximum oscillation frequency (f_{max}), and minimum noise figure (NF_{min}) have been widely studied [8, 10, 11]. However, the emphasis of these works was mostly placed on the extreme values of these parameters. Less consideration was given to the relevant circuit bias condition or power consumption, which is critical for portable wireless applications of low-power electronics. The urgent demands for low-power electronics originate from the stringent requirements of small size and weight, low cost, long operating life and reliability for proliferating portable wireless applications. Thus a detailed study of key performance metrics at bias conditions relevant for power-constrained RF circuit design is much needed. In

this dissertation, we attempt to carry out such a comprehensive design study to facilitate power-efficient low-noise operations at mm-wave frequencies.

In a RF receiver, the low-noise amplifier (LNA) is one of the key components of its front-end circuit, which directly amplifies the input signal from an antenna. It could have a significant influence on the performance of a wireless communication system as a whole. A low noise figure and a high gain are required along with low power consumption. Owing to those merits of the CMOS technology, many CMOS LNAs have been reported with operation frequencies up to 20-40 GHz by the year of 2005 [10, 12-14]. In order to facilitate the mm-wave applications in even higher frequencies, a detailed design study using CMOS LNAs optimized for power-efficient operations at 60 GHz will open up further opportunities for RF ICs at the mm-wave frequency range. In this dissertation, we would like to provide detailed design considerations and performance trends for CMOS LNAs near the scaling limits with an operation frequency of 60 GHz.

1.3 DISSERTATION ORGANIZATION

This dissertation consists of seven chapters:

Chapter 1 is the introduction of mm-wave communication systems with enabling technologies and the motivation of this dissertation.

Chapter 2 surveys the current status of research and development on mm-wave LNA technologies in details, including both the III/V- and Si- based ones with the focus on the Si CMOS. It also provides starting points for various design parameters and performance benchmarks to compare with the simulation results obtained from this work.

Chapter 3 describes the design details of CMOS devices used in this dissertation. Starting with an overview of 2-D mixed mode simulation that are used throughout the work, bulk MOSFETs with channel lengths from state-of-the-art 40 nm to the bulk scaling limit of 20 nm are carefully designed with good short-channel effect (SCE). The device design of DG MOSFETs with a channel length down to 10 nm, which is considered to be the ultimate scaling limit, is also presented.

Chapter 4 is to provide a complete solution of the RF small-signal design using a noise measure approach. It contains important relationships between a number of major design parameters such as circuit stability, power gains, and noise parameters all within the framework of two-port small-signal parameters. It also includes a newly developed methodology, which uses the noise measure as a figure of merit to indicate the overall noise performance. Linearity analysis for the CMOS LNA is also covered.

Chapter 5 follows up on the previous chapters to identify the specific conditions that lead to the optimal power-constrained LNA noise performance. Beginning with an introduction of intrinsic and extrinsic MOSFET models used in this

dissertation, a comprehensive design study of the 40-nm nMOS LNA at 60 GHz is carried out with particular considerations given to the design of DC bias conditions for power-efficient operations. The actual effect of gate resistance in current CMOS technologies has been accounted for better accuracy of simulation results. Correlation of simulation results with published hardware data is made to provide insights from the aspect of device technology. The frequency dependence and the trade-off between power and linearity are also discussed.

Chapter 6 extends the earlier work on the 40-nm nMOS LNA to advanced CMOS LNAs using MOSFETs near the scaling limits including: the 20-nm bulk and DG nMOSFETs, and the 10-nm DG nMOSFET. The performance trends along the roadmap under such a scaling scenario have been investigated.

Chapter 7 summarizes this dissertation with suggestions for future work.

REFERENCES

1. C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Brodersen, "Design of CMOS for 60 GHz applications", *IEEE Int. Solid-State Circuits Conf. Dig. Tech.*, vol. 1, pp. 440-538, Feb. 2004.

2. RF and AMS Technologies for Wireless Communications, 2005 International Technology Roadmap for Semiconductors.

3. C. H. Doan, S. Emami, D. Sobel, A. M. Niknejad and R. W. Brodersen, "60 GHz CMOS radio for Gb/s wireless LAN", *IEEE RFIC Symp. Dig.*, pp. 225-228, June 2004.

4. X. Guan, H. Hashemi, and A. Hajimiri, "A fully integrated 24-GHz eight-element phased-array receiver in silicon", *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2311-2320, 2004.

5. Y. L. Kok et al., "160-190-GHz monolithic low-noise amplifiers", *IEEE Microw. Guided Wave Lett.*, vol. 9, no. 8, pp. 311-313, 1999.

6. Y. Yamashita et al., "Pesudomorphic $In_{0.52}Al_{0.48}As/In_{0.7}Ga_{0.3}As$ HEMTs with an ultrahigh f_T of 562 GHz", *IEEE Electron Device Lett.*, vol. 23, No. 10, pp. 573-575, 2002.

7. J. Rieh et al., "Performance and design considerations for high speed SiGe HBTs of f_T/f_{max}=375GHz/210GHz", *Proc. Int. Conf. Indium Phosphide Related Materials*, pp. 374-377, 2003.

8. S. Lee et al., "Record RF performance of sub-46 nm Lgate NFETs in Microprocessor SOI CMOS technologies", *IEDM Tech. Dig.*, pp. 251-254, Dec. 2005.

9. H. S. Bennett et al., "Device and technology evolution for Si-based RF integrated circuits", *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1235-1258, 2005.

10. S. Shin, M. Tsai, R. Liu, K. Lin, and H. Wang, "A 24-GHz 3.9-dB NF low-noise amplifier using 0.18 μm CMOS technology", *IEEE Microw. Wireless Compon. Lett.*, vol. 15, no. 7, pp. 448-450, 2005.

11. L. F. Tiemeijer et al., "Record RF performance of standard 90 nm CMOS technology", *IEDM Tech. Dig.*, pp. 441-444, Dec. 2004.

12. F. Ellinger, "26-42 GHz SOI CMOS low noise amplifier", *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 522-528, 2004.

13. K. Yu, Y. Lu, D. Chang, V. Liang, and M. F. Chang, "K-band low-noise amplifiers using 0.18 um CMOS technology", *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 3, pp. 106-108, 2004.

14. X. Guan and A. Hajimiri, "A 24-GHz CMOS front-end", *IEEE J. Solid-State Circuits*, vol. 39, no. 2, pp. 369-373, 2004.

CHAPTER 2

CURRENT STATUS OF RESEARCH AND DEVELOPMENT ON MILLIMETER-WAVE TECHNOLOGIES

2.1 APPLICATION SPECTRUM OF RF TECHNOLOGIES

RF technologies at mm-wave frequencies now represent essential and critical technologies for the success of many semiconductor products. Such technologies serve the rapidly growing wireless communications market. They depend on many materials systems, some of which are based on semiconductors from group IV in the periodic table, such as Si and SiGe, and others are based on compound semiconductors composed of elements from group III and V, such as GaAs and InP, which are not compatible with CMOS processing.

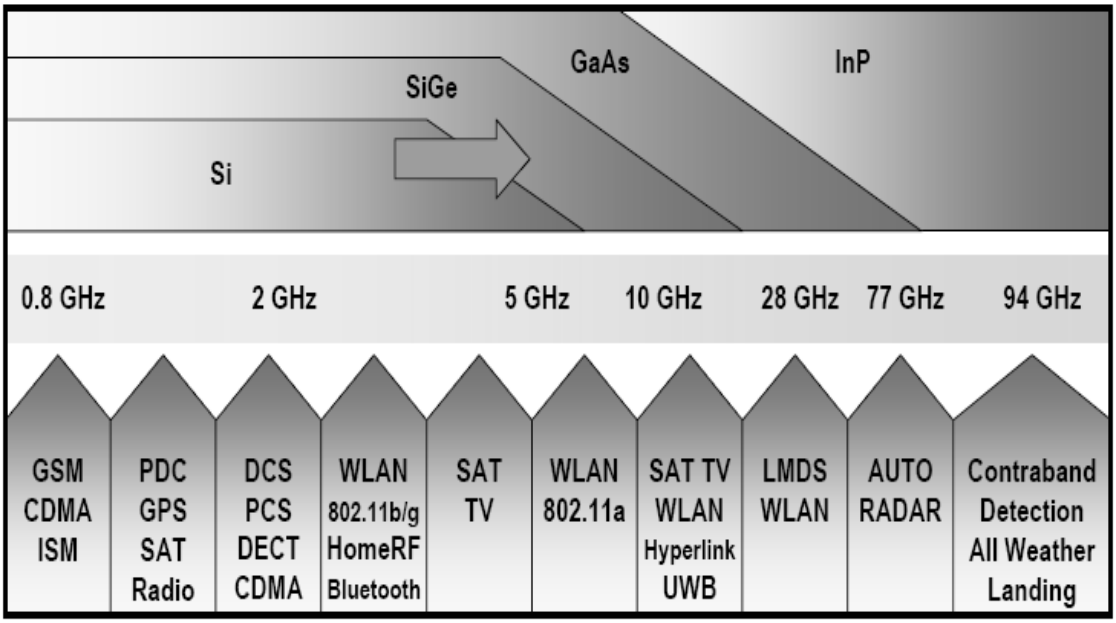

Fig. 2.1 Application spectrum of RF technologies for wireless communications (adapted from ITRS 2005 [1]).

Fig. 2.1 schematically presents the connection among commercial wireless communication applications, available spectrum, and the kinds of elemental and compound semiconductors likely to be used. At the same frequency, performance tends to increase in the following order: Si CMOS, SiGe, GaAs, and InP. Today group IV semiconductors (Si and SiGe) dominate below 10 GHz and III-V compound semiconductors dominate above 10 GHz. However, these boundaries are not as well defined as it may appear in Fig. 2.1, but are broad, diffuse, and change with time. The range of frequencies where competition exists between group IV and III-V compound semiconductors is expected to move towards higher frequencies [1]. Nevertheless, while SiGe has shown promising capability in mm-wave frequency range, it is unlikely to replace III-Vs in the near future for applications where either high power gain or ultra low noise is required.

The main drivers for wireless communications systems are cost, frequency bands, power consumption, functionality, size of mobile unit, and very high volume of product. During the last two decades, technologies based on III-V compound semiconductors have established new business opportunities for wireless communication systems. When high volumes of products are expected, Si-based technologies replace the III-Vs in those markets for which these CMOS-compatible processes can deliver appropriate performance at considerably low cost.

2.2 III/V-BASED TECHNOLOGIES

Recently many mm-wave system applications such as high-speed radio communications, automotive radar, and all weather landing have been proposed. These applications usually require high performance LNAs at mm-wave bands which typically utilize III-V technologies such as GaAs- or InP- based heterojunction bipolar transistors (HBTs), high electron mobility transistor (HEMTs), or metamorphic high electron mobility transistors (MHEMTs). InP-based technologies have shown very high gain, low noise figure, and high frequency capability among III-V compound semiconductor technologies, which makes it the dominating technology for frequency range above 120 GHz. An ultrahigh f_T of 562 GHz has been reported for a $In_{0.52}Al_{0.48}As/In_{0.7}Ga_{0.3}As$ HEMT based on a InP technology with a gate length of 25 nm [2]. Another exemplary demonstration of InP-based technology is a 160-190 GHz LNA with a 6-dB gain and 6-dB noise figure (*NF*) at 170 GHz, realized by using a 0.07-μm T-gate InGaAs/InAlAs/InP HEMT technology with a f_T of 230 GHz and a f_{max} of 500 GHz [3].

However, due to the typical high cost associated with InP-based technologies, it is very important to develop relatively low cost alternatives, such as GaAs-based technologies, as the consumer portions of wireless communications markets are very sensitive to cost. The benchmark f_T for GaAs-based transistors is reported to be 440 GHz with a lowest minimum noise figure of below 1.2 dB at 26 GHz, which utilized an $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As$ MHEMT fabricated on a GaAs substrate [4]. A 94-GHz

GaAs-based LNA has also been demonstrated, which achieves a *NF* of 2.3 dB over the band from 80 to 100 GHz. It was fabricated using an $In_{0.52}Al_{0.48}As/In_{0.80}Ga_{0.20}As/In_{0.53}Ga_{0.47}As$ HEMT technology [5].

2.3 SI-BASED TECHNOLOGIES

Si CMOS technologies have had acceptable device characteristics for RF wireless applications for frequencies from 1 to 10 GHz since early 1990s, and today the Si-based technologies (Si CMOS and more recently SiGe) dominate the frequency range below 10 GHz. A comparison of competing technologies over the past decades in Fig. 2.2 shows the steady improvement in cut-off frequency (f_T) for the CMOS, SiGe HBT, III-V HBT and HEMT. Increased RF performance for Si-based technologies is usually achieved by geometrical scaling whereas increased RF performance for III/V-based technologies is achieved by optimizing carrier transport properties through materials and bandgap engineering. While the III-V HEMT still shows the highest f_T among the technology candidates for mm-wave applications, the combination of low cost and superior integration with VLSI digital functions will make Si CMOS and SiGe HBT technologies of choice for RF wireless applications, as evident by the expanding wireless local area network (WLAN) connectivity market that is dominated by CMOS transceivers now and promising automotive radar applications at 77 GHz by SiGe HBT technology.

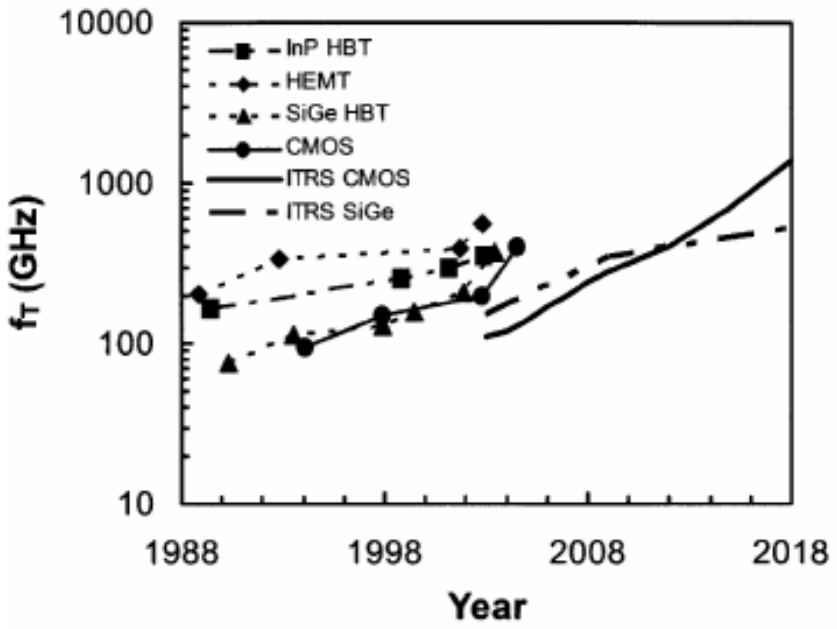

Fig. 2.2 Cut-off frequency (f_T) as a function of demonstration for CMOS, SiGe HBT, III-V HBT and HEMT technologies with 2003 ITRS roadmap (adapted from H. S. Bennett *et al.* [6])

SiGe HBT technology has been a leading contender among the Si-based technologies for high frequency applications owing to its advantage over CMOS technology such as higher operation speed, higher trans-conductance, higher output impedance, lower 1/*f* noise, better device matching and better power performance. Recent effort to improve the operation speed of SiGe HBT technology includes a record high f_T of 375 GHz [7], which is the highest reported value for any Si-based transistor. A 60-GHz SiGe transceiver circuit has also been realized including a LNA with a gain of 17 dB and a *NF* of 3.3 dB, using a 0.12-μm SiGe bipolar technology with a f_T of 207 GHz [8].

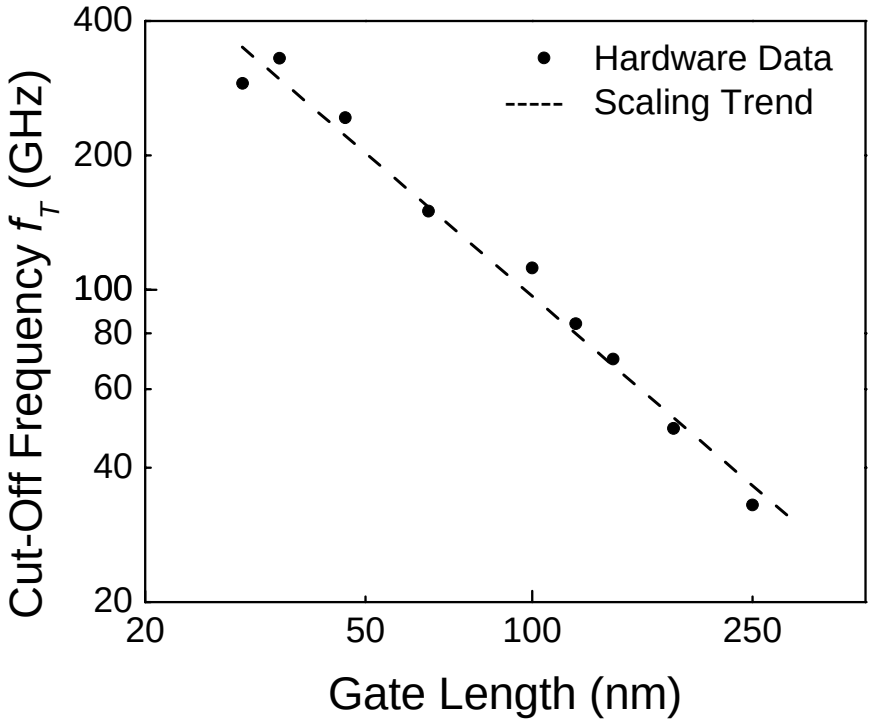

Fig. 2.3 Technology evolving trend for CMOS cut-off frequency (f_T) as a function of transistor gate length from reported CMOS technologies [9-13].

Cost has been and will continue to be the motivating factor in utilizing Si CMOS technology for RF ICs. CMOS has numerous unique benefits including low cost derived from dominance in the digital electronics market, superior integrated functions and continued improvement in performance as dictated by Moore's law. In recent years, the RF performance of CMOS transistors has been significantly improved through continuous technology scaling. Shown in Fig. 2.3 is the cut-off frequency (f_T) as a function of transistor gate length from reported RF CMOS technologies. No significant sign of slowing down has been shown in the scaling trend

of increasing f_T as the gate length is reduced, as a result of countless engineering efforts made at every possible level in materials, devices, circuits and systems. A record high f_T of 330 GHz has been demonstrated by IBM in IEDM 2005, which features a 65-nm SOI CMOS technology with a 1.05-nm gate oxide and 35-nm gate length [9]. The state-of-the-art CMOS LNA has been reported to have a *NF* of 4 dB and a gain of 9.5 dB at 40 GHz from a 90-nm SOI CMOS technology [14]. The availability of high performance CMOS, coupled with the higher resistive SOI substrate, makes these technologies very attractive for RF system-on-chip (SoC) applications.

Although LNAs using advanced SOI CMOS process demonstrated excellent RF performances, the standard bulk CMOS process is still attractive due to cost considerations. A 30-nm gate length nMOSFET with a record high f_T of 290 GHz is reportd in IEDM 2005 [10], which is the highest f_T reported for standard bulk CMOS process so far. Many attempts have been made to provide CMOS LNAs sufficient for frequencies above 20 GHz using standard sub-micron technologies. A CMOS LNA with a 13-dB gain and a 5.6-dB *NF* was fabricated using a 0.18-μm process at 23.5 GHz [15], and a 24-GHz LNA fabricated using a 0.18-μm CMOS technology has been reported to have a power gain of 15 dB and a *NF* of 6 dB [16]. The best CMOS LNA achieved using standard CMOS technology features a *NF* of 3.9 dB and a gain of 13.1 dB at 24 GHz using a TSMC 0.18-μm CMOS process with a f_T of 70 GHz [17]. It demonstrates the lowest noise figure among the reported LNAs at frequencies above 20 GHz using standard CMOS processes.

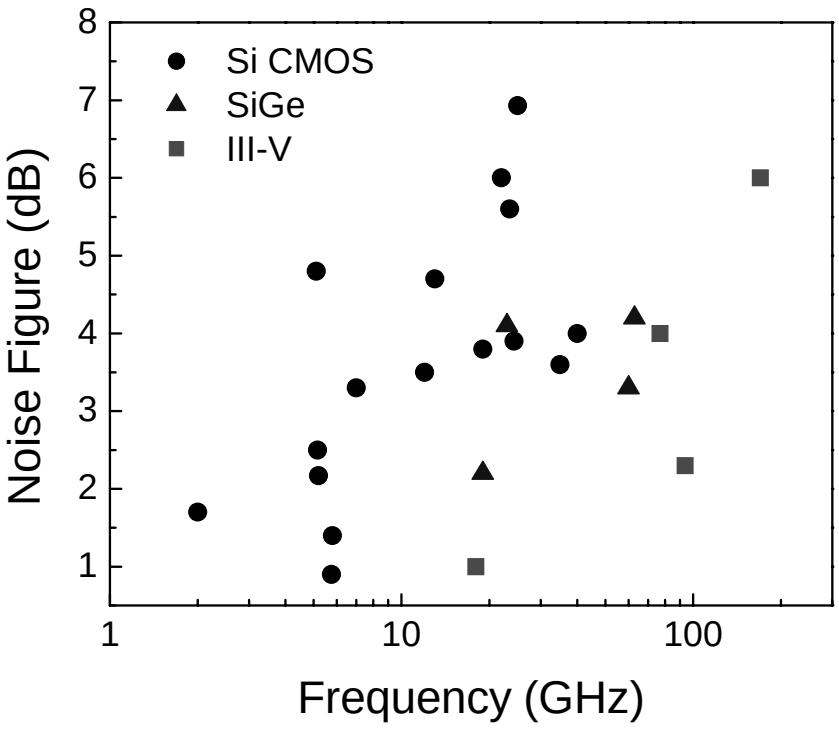

Fig. 2.4 Comparison of noise figure versus frequency of demonstration for state-of-the-art LNAs using technologies from Si CMOS, SiGe and III-V materials [3, 5, 8, 14-32].

Fig. 2.4 summarizes the noise figures for reported state-of-the-art LNAs based on Si CMOS, SiGe and III-V technologies from published hardware data up to date. Although the frequency range above 100 GHz is still dominated by the III-V technologies, Si-based technologies are now capable of providing low noise amplification up to 60 GHz, and more exciting results made by advanced CMOS technologies are being reported towards higher frequencies. It is also evident that the boundary between the Si-based and the III/V-based technologies has already been moved towards higher frequencies as compared with Fig. 2.1. Furthermore, as CMOS

scaling is approaching the limit imposed by short-channel effect, leakage current and other factors, double-gate (DG) MOSFET is becoming a subject of intense VLSI research because it can be scaled to the shortest channel length possible [33]. Although many researches have been done on the fabrication and characteristics of DG MOSFET or FinFET devices, there has been no report on the RF performance of the DG MOSFETs. However, it can be envisioned that as the RF performance of CMOS improves dramatically upon scaling, the question is not if, but when will CMOS become a viable alternative for mm-wave applications. Therefore, the main purpose of this dissertation is to design power-efficient CMOS LNAs with optimized noise performances at 60 GHz, utilizing advanced deep sub-micron device technologies including DG MOSFETs, and to project the performance trends with investigation on the viability of state-of-the-art CMOS technology as a potential candidate for building high performance LNAs for mm-wave applications.

REFERENCES

1. RF and AMS Technologies for Wireless Communications, 2005 International Technology Roadmap for Semiconductors.

2. Y. Yamashita et al., "Pesudomorphic $In_{0.52}Al_{0.48}As/In_{0.7}Ga_{0.3}As$ HEMTs with an ultrahigh f_T of 562 GHz", *IEEE Electron Device Lett.*, vol. 23, No. 10, pp. 573-575, 2002.

3. Y. L. Kok et al., "160-190-GHz monolithic low-noise amplifiers", *IEEE Microw. Guided Wave Lett.*, vol. 9, no. 8, pp. 311-313, 1999.

4. K. Elgaid, H. McLelland, M. Holland, D. A. J. Moran, C. R. Stanley and I. G. Thayne, "50-nm T-gate metamorphic GaAs HEMTs with fT of 440 GHz and noise figure of 0.7 dB at 26 GHz", *IEEE Electron Device Lett.*, vol. 26, no. 11, pp. 784-786, 2005.

5. A. Tessmann et al., "A coplanar 94 GHz low-noise amplifier MMIC using 0.07 um metamorphic cascade HEMTs", *IEEE Trans. Microw. Theory Tech.*, pp. 1581-1584, 2003.

6. H. S. Bennett et al., "Device and technology evolution for Si-based RF integrated circuits", *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1235-1258, 2005.

7. J. Rieh et al., "Performance and design considerations for high speed SiGe HBTs of f_T/f_{max}=375GHz/210GHz", *Proc. Int. Conf. Indium Phosphide Related Materials*, pp. 374-377, 2003.

8. S. Reynolds, B. Floyd, U. Pfeiffer and T. Zwick, "60GHz transceiver circuits in SiGe bipolar technology", *IEEE Int. Solid-State Circuits Conf. Dig. Tech.*, 2004.

9. S. Lee et al., "Record RF performance of sub-46 nm Lgate NFETs in Microprocessor SOI CMOS technologies", *IEDM Tech. Dig.*, pp. 251-254, Dec. 2005.

10. V. Dimitrov et al., "High performance, sub-50 nm MOSFETs for mixed signal applications", *IEDM Tech. Dig.*, pp. 213-216, Dec. 2005.

11. L. F. Tiemeijer et al., "Record RF performance of standard 90 nm CMOS technology", *IEDM Tech. Dig.*, pp. 441-444, Dec. 2004.

12. N. Zamdmer et al., "A 243-GHz Ft and 208-GHz Fmax, 90-nm SOI CMOS SoC Technology with Low-Power Millimeter-Wave Digital and RF Circuit Capability", *Symp. VLSI Tech. Dig.*, pp. 98-99, June 2004.

13. E. Morifuji, et al., "Future perspective and scaling down roadmap for RF CMOS", *Symp. VLSI Tech. Dig.*, pp. 163-164, 1999.

14. F. Ellinger, "26-42 GHz SOI CMOS low noise amplifier", *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 522-528, 2004.

15. K. Yu, Y. Lu, D. Chang, V. Liang, and M. F. Chang, "K-band low-noise amplifiers using 0.18 μm CMOS technology", *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 3, pp. 106-108, 2004.

16. X. Guan and A. Hajimiri, "A 24-GHz CMOS front-end", *IEEE J. Solid-State Circuits*, vol. 39, no. 2, pp. 369-373, 2004.

17. S. Shin, M. Tsai, R. Liu, K. Lin, and H. Wang, "A 24-GHz 3.9-dB NF low-noise amplifier using 0.18 μm CMOS technology", *IEEE Microw. Wireless Compon. Lett.*, vol. 15, no. 7, pp. 448-450, 2005.

18. J.-S. Goo, H.-T. Ahn, D. J. Ladwig, Z. Yu, T. H. Lee and R. W. Dutton, "A noise optimization technique for integrated low-noise amplifiers", *IEEE J. Solid-State Circuits*, vol. 37, pp. 994-1002, 2002.

19. Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, "The impact of scaling down to deep submicron on CMOS RF circuits", *IEEE J. Solid-State Circuits*, vol. 33, pp. 1023-1036, 1998.

20. B. A. Floyd and D. Qzis, "Low-noise amplifier comparison at 2 GHz in 0.25-μm and 0.18-μm RF-CMOS and SiGe BiCMOS", *IEEE RFIC Symp. Dig.*, pp. 185-188,

2004.

21. H. Samavati, H. R. Rategh, and T. H. Lee, “A 5-GHz CMOS wireless LAN receiver front end”, *IEEE J. Solid-State Circuits*, vol. 35, pp. 765-772, 2000.

22. T.-P. Liu and E. Westerwick, “5-GHz CMOS radio transceiver front-end chipset”, *IEEE J. Solid-State Circuits*, vol. 35, pp. 1927-1933, 2000.

23. H.-W. Chiu, S.-S. Lu, and Y.-S. Lin, “A 2.17-dB NF 5-GHz-band monolithic CMOS LNA with 10-mW DC power consumption”, *IEEE Trans. Microw. Theory Tech.*, vol. 53, pp. 813-824, 2005.

24. R. Fujimoto, K. Kojima, and S. Otaka, “A 7-GHz 1.8-dB NF CMOS low-noise amplifier”, *IEEE J. Solid-State Circuits*, vol. 37, no. 7, pp. 852-856, 2002.

25. D. J. Cassan and J. R. Long, “A 1-V transformer-feedback low-noise amplifier for 5-GHz wireless LAN in 0.18-um CMOS”, *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 427-435, 2003.

26. B. Mohammadi and C. A. T. Salama, “A 5.8 Ghz CMOS LNA for WLAN applications”, *IEEE RFIC Symp. Dig.*, pp. 113-116, 2004.

27. H. Yano, Y. Nakahara, T. Hirayama, N. Matsuno, Y. Suzuki, and A. Furukawa, “Performance of Ku-band on-chip matched Si monolithic amplifiers using 0.18-um-gatelength MOSFETs”, *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 6, pp. 1086-1093, 2005.

28. J. Gil, K. Han, and H. Shin, “13 GHz 4.67 dB NF CMOS low-noise amplifier”,. *Electron. Lett.*, Vol. 39, no. 14, pp. 1056-1058, 2003.

29. J. Bock et al., “Sub 5 ps SiGe Bipolar Technology”, *IEDM Tech. Dig.,* pp. 763-766, Dec. 2002.

30. G. Schuppener, T. Harada, and Y. Li, “A 23-GHz low-noise amplifier in SiGe heterojunction bipolar technology”, *IEEE RFIC Symp. Dig.*, pp. 177-180, 2001.

31. R. Eye and D. Allen, “77 GHz low noise amplifier for automotive radar applications”, *IEEE GaAs Dig.,* pp. 139-142, 2003.

32. Y. C. Lien et al.,“Low-noise metamorphic HEMTs with reflowed 0.1-μm T-gate”, IEEE Electron Device Lett., vol. 25, no. 6, pp. 348-350, 2004.

33. Y. Taur et al., “CMOS scaling into the nanometer regime”, *Proc. IEEE*, vol. 85, no. 4, pp. 486-504, 1997.

CHAPTER 3

MOSFET DESIGN NEAR THE SCALING LIMIT

USING TCAD

3.1 OVERVIEW OF 2-D MIXED-MODE SIMULATION

TCAD refers to using computer simulations to develop and optimize semiconductor processing technologies and devices. TCAD simulation tools solve fundamental and physical partial differential equations, such as diffusion and transport equations for discretized geometries, representing the silicon wafer or the layer system in a semiconductor device or even small circuits with a few devices. This deep physical approach gives TCAD simulation predictive accuracy. It is therefore possible to substitute TCAD computer simulations for costly and time-consuming test wafer runs when developing and characterizing a new semiconductor device or technology. The TCAD tools used in this work are provided by Integrated Systems Engineering (ISE), which is now part of SYNOPSYS. It consists of many powerful programs including MDRAW as a device editor and meshing tool, DESSIS as a device simulator, TECPLOT-ISE and INSPECT as plotting and data output tools.

In the mix-mode 2-D device and circuit simulation, a real semiconductor device, such as a MOSFET, is represented as a virtual device whose physical properties are discretized onto a user-defined mesh of grids. The electrostatic potential and carrier density at each grid are computed from the resolution of the Poisson equation coupled with carrier continuity equation in an iterative and self-consistent fashion. Nodal currents, voltages, and charges of such a single device in isolation or combined with other devices in a circuit can be simulated numerically using the mixed

device and circuit capability of SYNOPSYS TCAD. The small-signal high-frequency behavior of devices and circuits is usually analyzed in the framework of two-port networks. SYNOPSYS TCAD supports this type of analysis through a DESSIS small-signal simulation followed by a post-processing in INSPECT with the extension library TwoPortNetRFExtr.tcl. It computes the complex small signal admittance *Y* matrix with respect to the small-signal currents and voltages at observation nodes. For the numerical simulation of noise phenomena in physics-based device simulator, a numerically efficient Green function approach to the Langevin equation-based simulation of the Impedance Field Method [1] is the basis for the implementation in DESSIS [2]. In brief, an impulsive current source is injected into each mesh grid generated by 2-D device simulator, and drift-diffusion transport equations are solved to calculate the voltage fluctuations at the electrodes. The impedance field, originally defined by Shockley [3], is the transfer function between the stimuli current at location r_i and the response voltage at the *k*-th electrode

$$Z_{kr} = \frac{\text{Voltage fluctuation at the } k\text{-th electrode}}{\text{Inject current at location } r_i \text{ inside a device}}$$

The local noise generation can be modeled by injecting a current i_{in} at a location r_i and pulling out that current from another nearby location r_i', as shown in Fig. 3.1 (a) for a MOSFET in its common-source configuration. The total power spectral density of noise voltage u_{nk} at the *k*-th electrode, S_{vk}, is obtained by integrating the contribution from each location r_i over the whole area of the device:

$$S_{vk} = \iint \left|\nabla Z_k\right|^2 S_{in} dr$$

where ∇Z_k is the gradient of the impedance field at location r_i and S_{in} is the power spectral density of the local noise source. For thermal noise in equilibrium condition, S_{in} is equal to $k_B\, qn\mu T$. The accuracy of local noise source model is very important. Generally the thermal noise is directly proportionally to temperature. In thermal equilibrium, the lattice temperature and carrier temperature are identical. For high energy carriers in non-equilibrium conditions, this noise model needs to be modified by hydrodynamic models if needed. Finally, the results given by the simulator are in the format of self- and cross- power spectral densities of noise voltages at each observation nodes, as shown in Fig. 3.1 (b), where the MOSFET is noiseless with u_{ng} and u_{nd} representing the noise generated by all local noise sources within the device.

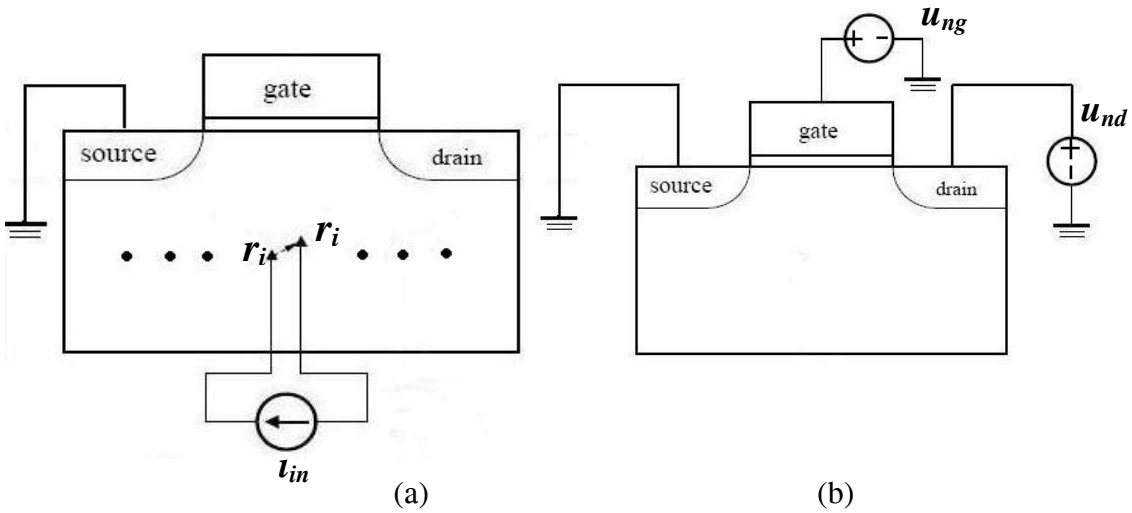

Fig. 3.1 (a): Schematic for impedance field method with local noise sources for 2D noise simulation; (b): Equivalent representation of noise voltages at the observation nodes using the result of impedance field method.

3.2 DEVICE DESIGN OF A 40-NM MOSFET

As dimensions of MOSFETs are scaled down to sub-100-nm regime, power supply voltage (V_{dd}) needs to be reduced as well in order to keep the electric field and power consumption of devices within reasonable limits. However, the threshold voltage (V_t) has not been scaled in proportion to the power supply voltage. In fact, a minimum V_t of ~0.2 V in worst case is required to turn off the device properly at room temperature. Such a minimum V_t leads to a minimum V_{dd} of approximately 1 V, since the delay of standard CMOS circuits increases rapidly when the V_t/V_{dd} ratio exceeds 1/4 [4]. The 90-nm CMOS technology in production today features a channel length of sub-50 nm, a V_{dd} of 1.0-1.2 V and a V_t of 0.3-0.4 V [5, 6]. Based on these parameters, we start our bulk MOSFET design with a channel length of 40 nm, a V_{dd} of 1V and a V_t of 0.2-0.3 V.

One of the key challenges to MOSFET device design is to control the short-channel effect (SCE) when the channel length is scaled down below 100 nm. SCE is the decrease of threshold voltage in a short-channel device due to the two-dimensional electrostatic charge sharing between the gate and the source-drain regions, and it becomes more severe as the channel length is reduced into deep submicron regime. The threshold voltage roll-off in a short-channel device is shown to be [4]

$$\Delta V_t = \frac{24 t_{ox}}{W_{dm}} \sqrt{\psi_{bi}(\psi_{bi} + V_{ds})} \cdot e^{-\frac{\pi L/2}{W_{dm}+3t_{ox}}} \tag{3.1}$$

where W_{dm} is the maximum gate depletion width. W_{dm} is related to the doping concentration in the channel N_a as given by (assuming constant doping)

$$W_{dm} = \sqrt{\frac{4\varepsilon_{Si} kT \ln(N_a / n_i)}{q^2 N_a}} \tag{3.2}$$

Because of the exponential factor in (3.1), the V_t roll-off as a function of channel length is very sensitive to the sum of gate depletion width (W_{dm}) and oxide thickness (t_{ox}). If a scale length λ is defined as

$$\lambda = W_{dm} + 3 \cdot t_{ox} \tag{3.3}$$

then

$$\Delta V_t \propto \exp(-\frac{\pi \cdot L / 2}{\lambda}) \tag{3.4}$$

In order to avoid excessive SCE, the minimum channel length (L_{min}) should be kept at least 1.5 λ so that the V_t roll-off due to the SCE is negligible, as suggested by (3.4). Therefore, a MOSFET with a desired channel length of 40 nm would require a scaling length no larger than 26 nm. To meet such requirement, both the gate oxide thickness (t_{ox}) and the gate-controlled depletion width (W_{dm}) need to be reduced simutaneously. The oxide thickness is usually 1-2 nm for current technology nodes determined by the lithography [6, 7], below which the tunneling current quickly becomes unmanageable unless a new gate material is developed. Therefore, the device design of the 40-nm bulk MOSFET starts with an oxide thickness of 2 nm. This will,

in turn, result in a W_{dm} of no larger than 20 nm for a scaling length of 26 nm, which is required for a minimum channel length of 40 nm.

When the channel length is scaled to below 100 nm, a much higher doping concentration is needed in the channel in order to reduce W_{dm} according to (3.2) if a uniform profile is used. Unfortunately the threshold voltage is also increasing with N_a as given by [4]

$$V_t = V_{fb} + 2\psi_B + \frac{qN_aW_{dm}}{C_{ox}} = V_{fb} + 2\psi_B + \frac{\sqrt{4\varepsilon_{Si}qN_a\psi_B}}{C_{ox}} \tag{3.5}$$

V_t would become too high at such high doping concentration for a device with a channel length below 100 nm. Therefore a retrograde channel doping is needed to allow the threshold voltage to be decoupled from the gate-controlled depletion width. In practice, such a profile can be formed by using higher-energy implants that peak below the surface and can be simulated by specifying the doping profile using MDRAW from the SYNOPSYS TCAD. The doping profile is usually described by the Gaussian distribution function as

$$g(y) = C_{peak} \exp\left[-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2\right]$$

where $g(y)$ is the doping concentration along the y direction (depth into Si substrate), C_{peak} is the peak concentration at the surface, and σ is the standard deviation. The value of C_{peak} for the source and drain of the 40-nm MOSFET device is chosen to be

10^{20} /cm^3 as commonly used in current CMOS processes.

The device width is another important device parameter. Although it does not affect the 2D simulation results as all the physical quantities in the TCAD are denoted as per unit width, the value of device width is critical at the time when impedance matching, maximum oscillation frequency, and minimum noise figure are considered. Since the width of the modern MOSFETs is usually on the order of several microns, and it was reported that an optimum channel width of ~ 1μm is found to maximize f_{max} of the nMOSFET [7], the device width is set to be 1 μm in this work.

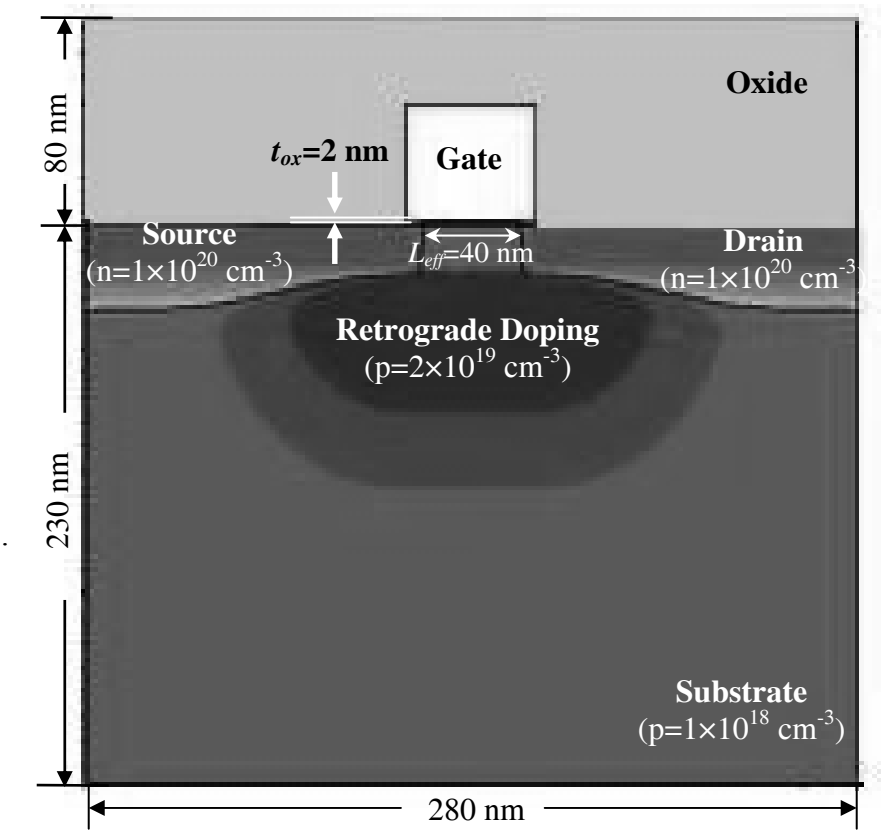

Fig. 3.2 Cross-section view of the 40-nm nMOSFET using SYNOPSYS TCAD with indicated peak doping concentrations and geometrical sizes.

A schematic cross-section view of the 40-nm nMOSFET following the above design considerations is shown in Fig. 3.2. For simplicity, the gate material is defined as a metal with a work function equal to that of n^+ polysilicon. It is capped by an oxide layer to account for the parasitic capacitances. The peak concentrations for the source, drain, substrate and the retrograde doping are indicated with the geometrical sizes as well. For the short-channel MOSFET, the channel length is not well defined as in the long-channel case [8]. In this work, the effective channel length (L_{eff}) is determined by the distance between the locations where the source-drain doping concentration along the channel direction falls to $2*10^{19}$ cm^{-3}. Fig. 3.3 shows the doping concentration in the surface of Si along the channel direction for the 40-nm nMOSFET from the source to the drain, and the effective channel length is found to be 40 nm in this case.

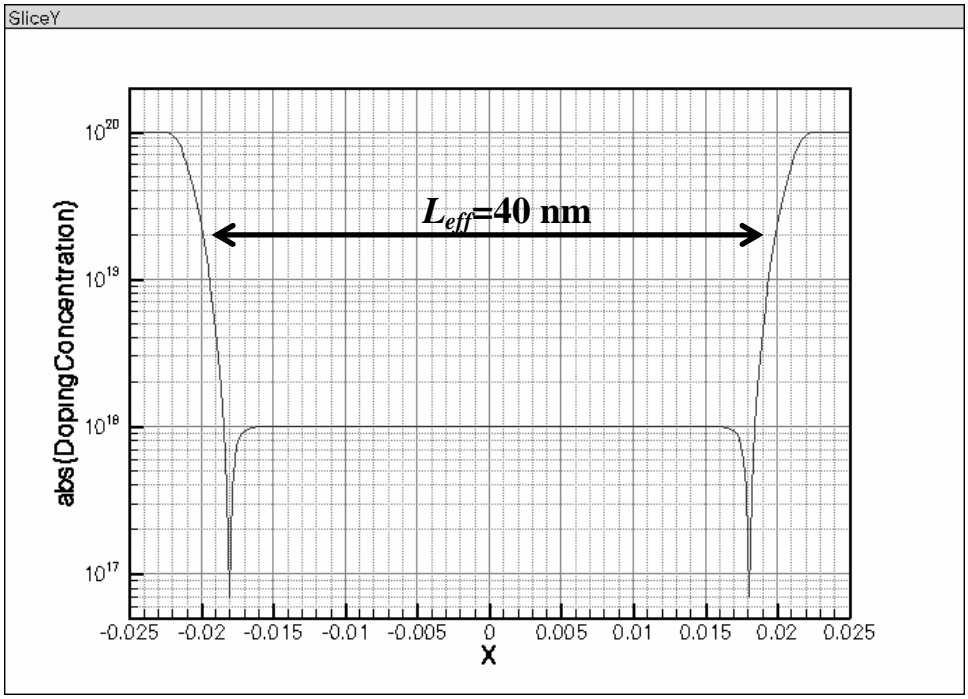

Fig. 3.3 Doping profile in the Si surface along the channel direction from source to drain for the 40-nm nMOSFET. The effective channel length is 40 nm for this device.

Fig. 3.4 shows the retrograde doping profile along the center of the 40-nm nMOSFET in the depth direction from the Si/SiO_2 interface to the Si substrate. The retrograde doping profile, along with those for the source, drain and substrate, are adjusted such that the key design parameters (W_{dm} and V_t) are within the required range for good SCE as discussed earlier in this section. The V_t is determined at the intersection point of the tangent on the steepest point of the drain current (I_d) - gate voltage (V_g) curve with the V_g axis at low drain conditions (with a drain voltage of 10 mV) as illustrated in Fig. 3.5. Finally, Table 3.1 summarizes the design parameters for the 40-nm nMOSFET.

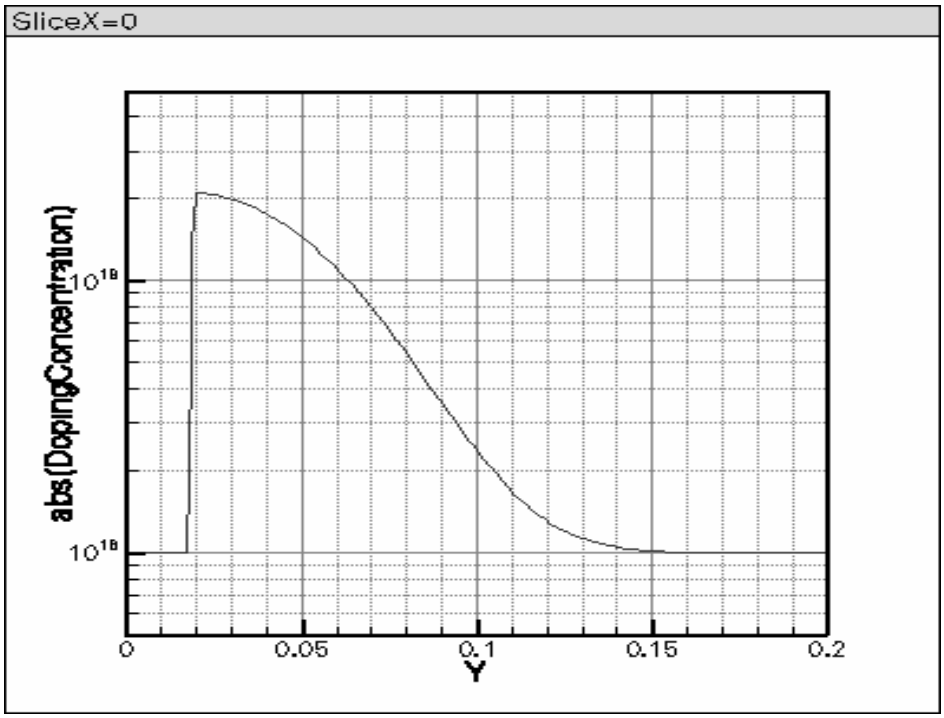

Fig. 3.4 Retrograde channel doping profile along the center of the 40-nm nMOSFET (the Y axis is the depth direction from Si/SiO_2 surface at Y=0 into Si substrate).

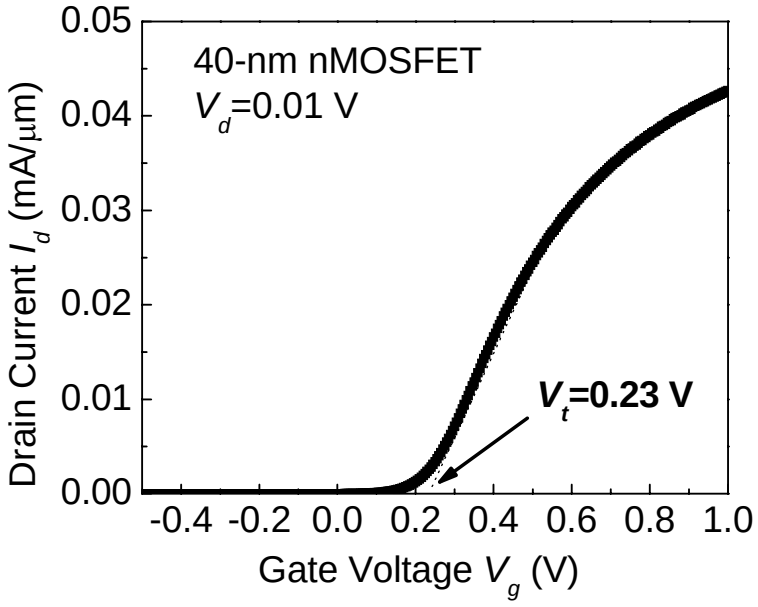

Fig. 3.5 Drain current (I_d) as a function of gate voltage (V_g) for the 40-nm nMOSFET at low drain condition (V_d=0.01 V) with indicated threshold voltage (V_t).

Table 3.1 Summary of key design parameters for the 40-nm nMOSFET

Parameter	**Unit**	**Value**
Effective Channel Length (L_{eff})	nm	40
Device Width (W)	μm	1
Gate Oxide Thickness (t_{ox})	nm	2
Maximum Depletion Width (W_{dm})	nm	20
Source/ Drain Doping	cm^{-3}	Gaussian, $1*10^{20}$
Channel Doping	cm^{-3}	Retrograde, $2*10^{19}$
Substrate Doping	cm^{-3}	$1*10^{18}$
Threshold Voltage (V_t)	V	0.23

3.3 MOSFET DEVICE DESIGN NEAR THE SCALING LIMIT

3.3.1 20-nm bulk MOSFET

Fig. 3.6 shows the CMOS scaling trends of the technology node and the transistor physical gate length as a function of time for the past and future. The performance of CMOS devices have been greatly improved by scaling in recent years and the current CMOS technology in production is quickly moving from the 90-nm node with an effective channel length of 50 nm, to the 65-nm node and beyond with a gate length shorter than 40 nm.

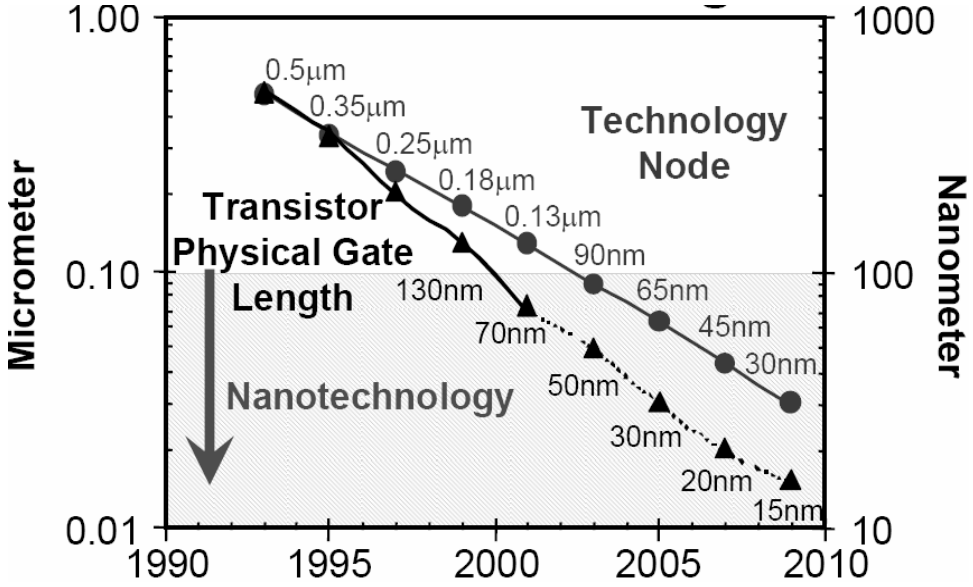

Fig. 3.6 CMOS scaling trends of technology node and transistor physical gate length for the past and future (adapted from R. Chau, Intel).

It is generally believed that the scaling limit for bulk CMOS will be reached at the time when the channel length is reduced to 20 nm [9]. In order to investigate the CMOS RF performance trends under such a scaling roadmap, a bulk MOSFET with an effective channel length of 20 nm was designed based on the foregoing 40-nm device using the constant-field scaling [4]. In constant-field scaling, it was proposed that one can keep the short-channel effect under control by scaling down the vertical and horizontal dimensions, while also proportionally decreasing the applied voltages and increasing the substrate doping concentration. This is shown schematically in Fig. 3.7. Therefore, for the scaling design from the 40-nm MOSFET to the 20-nm MOSFET, the device dimensions should reduced by a factor of 2 in general with the doping concentrations doubled.

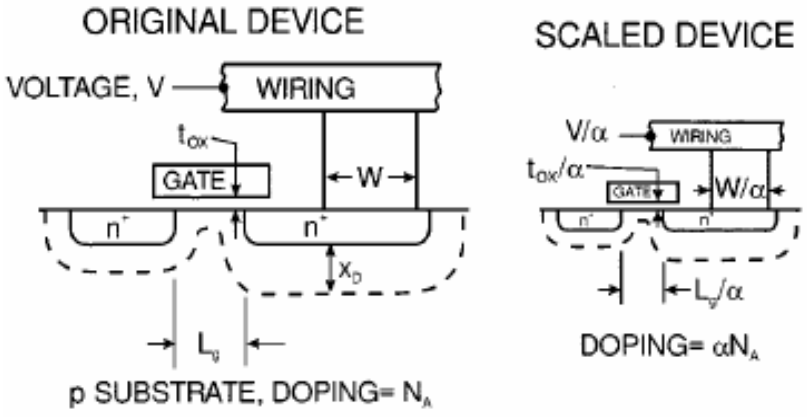

Fig. 3.7 Schematic illustration of the constant-field scaling principle of silicon technology (adapted from D. J. Frank *et al.* [10])

Fig. 3.8 compares the cross-section views of the 20-nm bulk MOSFET and the 40-nm device, from which the 20-nm device was designed by the constant-field scaling, using TECPLOT-ISE from SYNOPSYS TCAD. Similar design considerations that have been given to the 40-nm MOSFET were also applied to the 20-nm device. The top portions of the device including the gate height and the parasitic oxide layer, the thickness of the Si substrate, as well as the power supply voltage are not scaled due to practical considerations. Table 3.2 summarizes the key design parameters for the 20-nm bulk MOSFET, in comparison with those for the 40-nm bulk MOSFET.

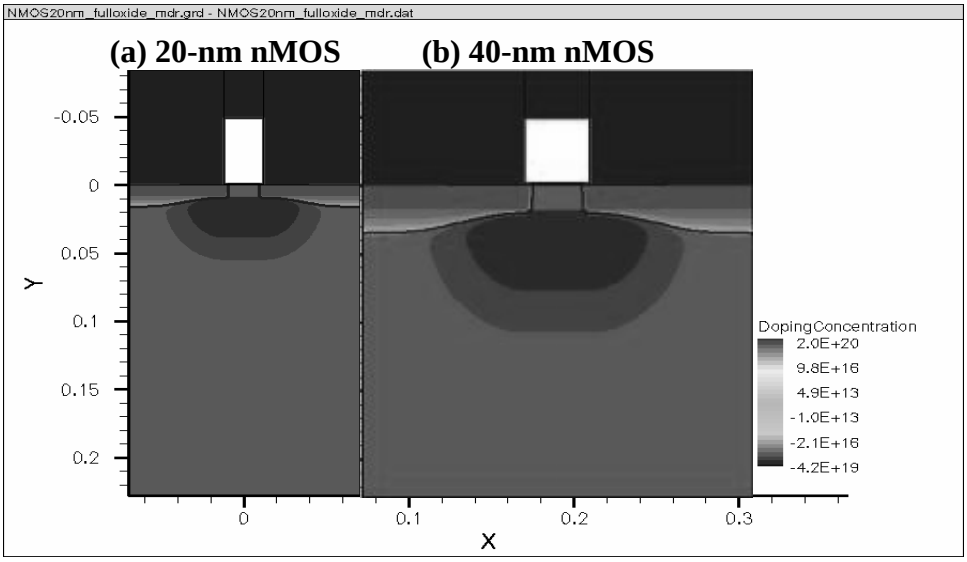

Fig. 3.8 Cross-view comparison of the 20-nm nMOSFET (a) and the 40-nm device (b) from which the 20-nm device was designed using the constant-field scaling.

Table 3.2 Summary of key design parameters for the 20- and 40- nm nMOSFETs.

Parameters	**Unit**	**Bulk MOSFET**	
Gate Length (L_{eff})	nm	40	20
Oxide Thickness (t_{ox})	nm	2	1
Depletion Width (W_{dm})	nm	20	10
S/D Doping	cm^{-3}	Gaussian, $1*10^{20}$	Gaussian, $2*10^{20}$
Channel Doping	cm^{-3}	Retrograde, $2*10^{19}$	Retrograde, $4*10^{19}$
Substrate Doping	cm^{-3}	$1*10^{18}$	$2*10^{18}$
Threshold Voltage (V_t)	V	0.23	0.20

3.3.2 Double-gate MOSFETs

As geometry of the CMOS device is approaching the scaling limit imposed by many limiting factors including SCE, double-gate (DG) MOSFET is becoming a subject of intense VLSI research because it offers immunity to short channel effects and decreased parasitic capacitances [11]. The channel length of the DG MOSFET can be scaled down further than the bulk scaling limit of 20 nm, given that some of the technological problems are solved. The advantages of the DG MOSFETs include, but not limited to: ideal sub-threshold slope, volume inversion for symmetric DG, setting of threshold voltage by the gate work function thus avoiding dopant fluctuation effects. DG MOSFETs in this work are designed with channel lengths of 20 nm for comparison with bulk devices, and 10 nm for prediction.

Fig. 3.9 shows the cross-section view of the 20-nm DG nMOSFET using the same geometry and doping parameters as the 20-nm bulk nMOSFET, except that the

constant doping in the channel. Although the retrograde profile is needed to allow the threshold voltage (V_t) to be decoupled from the gate-controlled depletion width (W_{dm}) in the bulk MOSFET device, it is not necessary for the DG MOSFET as the V_t can be easily adjusted by the work function. The gate work function for the 20-nm DG MOSFET is adjusted such that the 20-nm DG nMOSFET has the same V_t as the 20-nm bulk device. Since the body of the 20-nm DG MOSFET is designed to be full depleted, the role of depletion width for the scale length as in the bulk device is replaced by the Si thickness (T_{Si}). T_{Si} of the 20-nm DG MOSFET is set to be the same as the W_{dm} of the 20-nm bulk nMOSFET, which is 10 nm in this case.

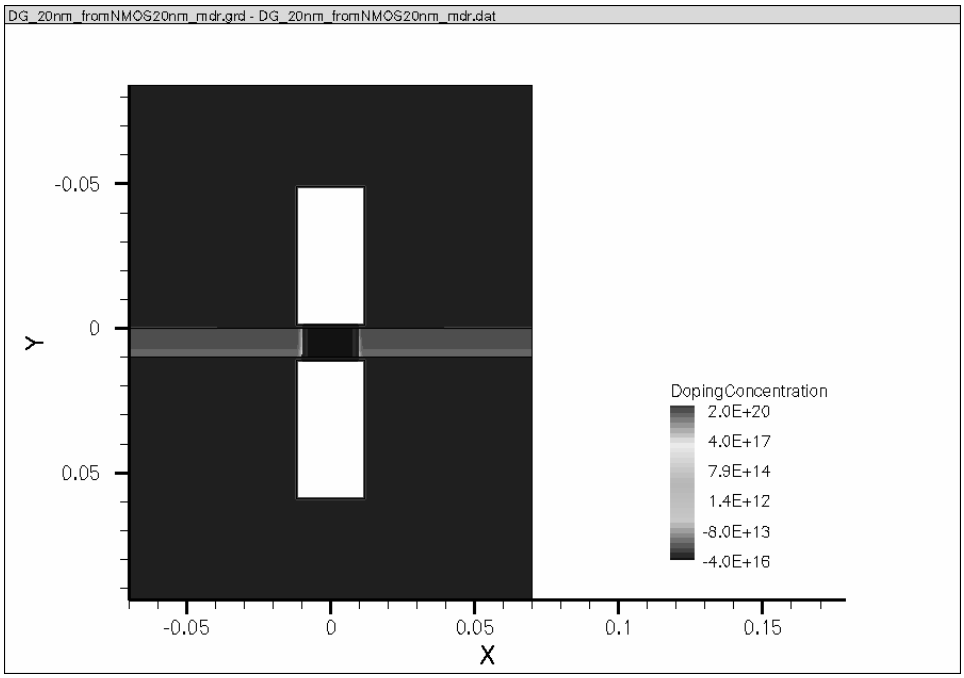

Fig. 3.9 Cross-section view of the 20-nm DG MOSFET using TECPLOT-ISE from SYNOPSYS TCAD.

Fig. 3.10 compares the relative geometry and configuration for all CMOS devices used in this study: (a) 40-nm bulk MOSFET, (b) 20-nm bulk MOSFET, (c) 20-nm DG MOSFET, and (d) 10-nm DG MOSFET. As the 20-nm DG MOSFET is scaled to 10 nm, the device dimensions are reduced by a factor of 2. It is noted that this would require an oxide thickness of 0.5 nm, which is unable to meet the gate leakage current requirement for low standby power applications if conventional SiO_2 gate dielectric is still to be used. However, such a thin thickness can be realized by high-k dielectric gate material in terms of EOT (equivalent oxide thickness) and it has been demonstrated that an ultra-thin (EOT=0.3 nm) $LaAlO_3$ gate dielectrics with excellent electrical characteristics can be fabricated by a novel high temperature deposition process [12]. As the dimensions are reduced by half from 20-nm DG MOSFET to the 10-nm DG device, the doping concentrations are not scaled up accordingly. This is because the source and drain doping have already been at a high level of $2*10^{20}$ /cm^3, and any concentration higher than that would degrade the mobility significantly. Therefore, the doping levels are kept the same while the DG MOSFET is scaled from 20 nm to 10 nm. The same gate work function is used for both the 20-nm and the 10-nm DG MOSFETs. Table 3.3 summarized the key design parameters for all CMOS devices near the scaling limit as studied in this work.

The text of this chapter, in part, is a reprint of the material that has been submitted for publication: Cai, Ming; Liu, Minjian; Taur, Yuan. “A design study of nanometer-gate low-noise amplifier near the limit of CMOS scaling”, Solid-State Electronics. The dissertation author was the primary author and the co-authors listed in

these publications directed and supervised the research which forms the basis for this chapter.

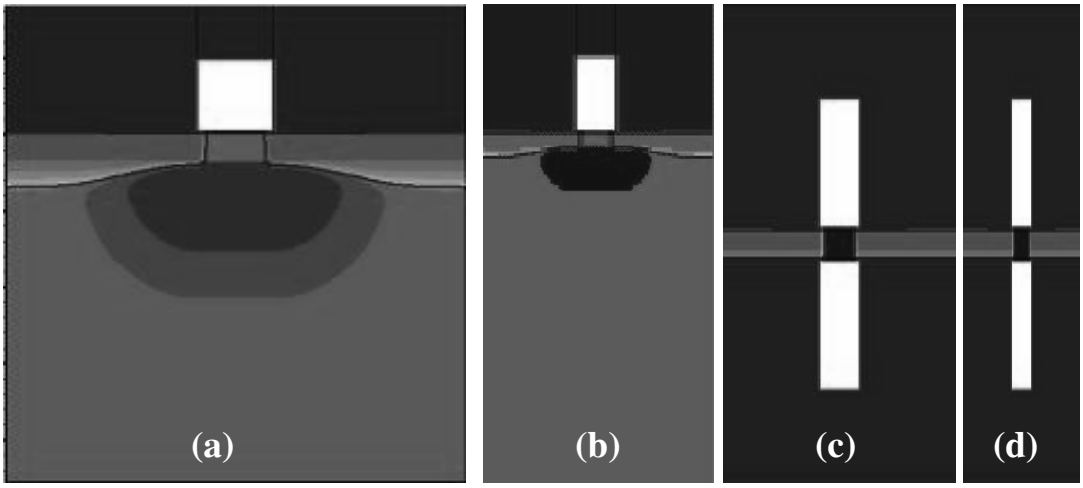

Fig. 3.10 A schematic comparison of CMOS devices used in this work: (a) 40-nm bulk MOSFET, (b) 20-nm bulk MOSFET, (c) 20-nm double-gate (DG) MOSFET, and (d) 10-nm DG MOSFET.

Table 3.3 Summary of key design parameters for the CMOS devices used in this work

Parameters	**Unit**	**Bulk MOSFET**		**DG MOSFET**	
Gate Length (L_{eff})	nm	40	20	20	10
Oxide Thickness(t_{ox})	nm	2	1	1	0.5
Depletion Width/ Si thickness (W_{dm} / T_{si})	nm	20	10	10	5
Source/Drain Doping	cm^{-3}	Gaussian $1*10^{20}$	Gaussian $2*10^{20}$	Gaussian $2*10^{20}$	Gaussian $2*10^{20}$
Channel Doping	cm^{-3}	Retrograde $4*10^{19}$	Retrograde $4*10^{19}$	Constant $2*10^{18}$	Constant $2*10^{18}$
Substrate Doping	cm^{-3}	$1*10^{18}$	$2*10^{18}$	NA	NA
Threshold Voltage (V_t)	V	0.23	0.20	0.20	0.16

REFERENCES

1. ISE Integrated Systems Engineering AG, “DESSIS 10.0 reference manual”, 2004.

2. F. Bonani, G. Ghione, M.R. Pinto, and R.K. Smith, “An efficient approach to noise analysis through multidimensional physics-based models”, *IEEE Trans. Electron Devices*, vol. 45, pp. 261-269, 1998.

3. W. Shockley, J. Copeland, R. James, *The Impedance Field Method of Noise Calculation in Active Semiconductor Devices*, Quantum Theory of Atoms, Molecules, and the Solid-State, Academic Press, New York,1966.

4. Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*, Cambridge University Press, New York, 1998.

5. P. Ranade et al., “High performance 35nm L_{GATE} CMOS transistors featuring NiSi metal gate (FUSI), uniaxial strained silicon channels and 1.2nm gate oxide”, *IEDM Tech. Dig.*, pp. 227-230, Dec. 2005.

6. M. Horstmann et al., “Integration and optimization of embedded-SiGe, compressive and tensile stressed liner films, and stress memorization in advanced SOI CMOS technologies”, *IEDM Tech. Dig.*, pp. 243-246, Dec. 2005.

7. S. Lee et al., “Record RF performance of sub-46 nm L_{gate} NFETs in microprocessor SOI CMOS technologies”, *IEDM Tech. Dig.*, pp. 251-254, Dec. 2005.

8. Y. Taur, “MOSFET channel length: extraction and interpretation”, IEEE Trans. Electron Devices, vol. 47, no. 1, pp. 160-170, 2000.

9. Y. Taur, “CMOS design near the limit of scaling”, *IBM J. Res. & Dev.*, vol. 46, No. 2/3, pp. 213-222, 2002.

10. D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, “Device scaling limits of Si MOSFETs and their application dependencies,” *IEEE Electron Device Lett.*, vol. 19, pp. 385-387, 1998.

11. Y. Taur et al., “CMOS scaling into the nanometer regime”, *Proc. IEEE*, vol. 85, no. 4, pp. 486-504, 1997.

12. M. Suzuki, M. Tomita, T. Yamaguchi, and N. Fukushima, “Ultra-thin (EOT=3A) and low leakage dielectrics of La_alminate directly on Si substrate fabricated by high temperature deposition”, *IEDM Tech. Dig.*, pp. 445-448, Dec. 2005.

CHAPTER 4

LNA DESIGN USING THE NOISE MEASURE APPROACH

4.1 COMMON-SOURCE LNA CONFIGURATION

The CMOS LNA used in this study is implemented with one nMOSFET in its common-source (CS) configuration, as shown in Fig. 4.1. nMOSFET is usually preferred than pMOSFET because of its higher mobility. The CS configuration is not only widely used in the CMOS RF IC design but also suits the purpose of this study, that is, to project the performance trends of CMOS LNAs through the device engineering instead of intricate circuit design techniques. The small-signal high-frequency behavior of transistor amplifiers is usually analyzed in the framework of two-port networks. The CS LNA can thus be viewed as a two-port with input port 1 at the gate and output port 2 at the drain with source grounded. Port 1 is connected to the signal source I_S with a source admittance of Y_S=G_S+j*B_S, where G_S is the conductance and B_S is the susceptance. In sections that follow, the small-signal high-frequency behavior of the LNA is analyzed in the generalized two-port parameters with results applicable to any circuit configurations.

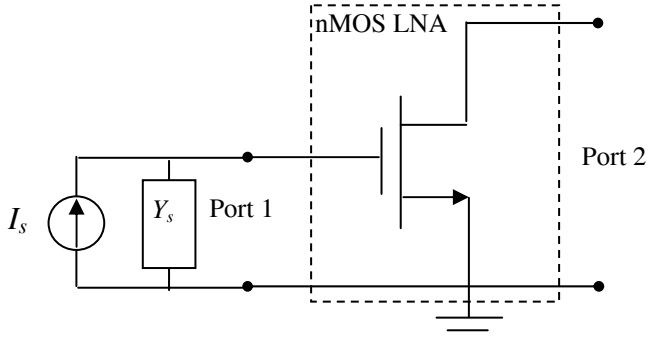

Fig. 4.1 Generalized representation of a two-port LNA using a common-source (CS) nMOSFET connected to a signal source I_S with its admittance Y_S.

4.2 SMALL-SIGNAL RF DESIGN USING TWO-PORT PARAMETERS

Design of solid-state small-signal RF transistor amplifier is usually based on the characterization of the active device as a linear two-port network, and such design is a systematic and mathematical procedure with exact computation for input / output admittance, stability, gain and noise figure available for the complete design solution free from approximation. The purpose of this chapter is to provide, in a single working reference, the important relationships necessary for the complete solution of the RF small-signal design using two-port parameters obtained from 2D device simulation tools.

4.2.1 Two-port network parameters

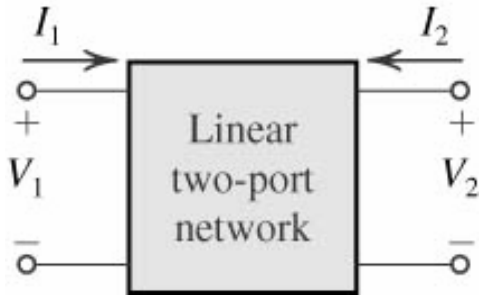

Fig. 4.2 A linear two-port network representation with its four port variables.

Fig. 4.2 shows a two-port network along with appropriate voltages and currents at its terminals. A two-port network has four port variables V_1, I_1, V_2 and I_2. If the two-port network is linear, two of these variables can be used as excitation

variables and the other two will be response variables. Depending on which two of these variables are used to represent the network excitation, different sets of equations (and a correspondingly different set of parameters) are obtained for characterizing the network. The most common representations are the impedance matrix (z parameters), the admittance matrix (y parameters), the hybrid matrix (h parameters), and the chain matrix (ABCD parameters). The major portion of this work presents design solutions in terms of y parameters, as they can be easily obtained from the small-signal equivalent circuit models of the transistor amplifier and the 2D device simulator SYNOPSYS TCAD as well. However, the results obtained from the y parameters are equally applicable to other representations using the direct conversion between these parameters [1].

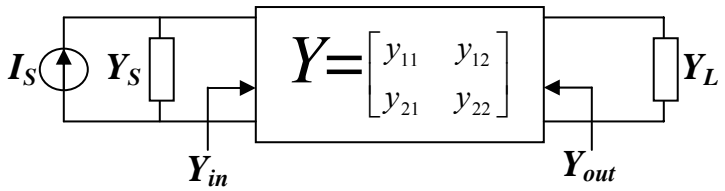

Fig. 4.3 Block diagram of a two-port amplifier and its y matrix with its input admittance Y_{in} and output admittance Y_{out}. Y_S is the admittance of the signal source (I_S), and Y_L is the admittance of the load.

A y-parameter representation of a two-port network with input and output admittances is shown in Fig. 4.3. The input and output admittances can be expressed in terms of the y parameters and arbitrary source and load terminations. The basic y-parameter equations for a two-port network are:

$$I_1 = y_{11}V_1 + y_{12}V_2$$
$$I_2 = y_{21}V_1 + y_{22}V_2 \tag{4.1}$$

where y_{11} is the input admittance at port 1 with port 2 short-circuited, and y_{22} is the admittance looking into port 2 with port 1 short-circuited. y_{21} represents the forward gain, and y_{12} represents the internal feedback in the network. The input and output admittances can be derived for the equivalent circuit representation shown in Fig. 4.3 as

$$Y_{in} = y_{11} - \frac{y_{12}y_{21}}{Y_L + y_{22}}$$
$$Y_{out} = y_{22} - \frac{y_{12}y_{21}}{Y_S + y_{11}} \tag{4.2}$$

Y_{in} can therefore be seen to be dependent on the load admittance Y_L and similarly Y_{out} is dependent on the source admittance Y_S. If y_{12}=0 (no internal feedback in the network), Y_{in}=y_{11} and Y_{out}=y_{22}.

4.2.2 Stability

Transistors are non-unilateral devices. The existence of internal feedback is indicated by the fact that a signal applied to the output port of a transistor amplifier results in a response at the input port, which is equivalently represented by y_{12} not

equal to zero. Therefore Y_{in} and/or Y_{out} are functions of Y_S and/or Y_L. Depending on the value of Y_S and Y_L, the real parts of Y_{in} and Y_{out} can be negative and therefore cause potential oscillations. The two-port network is said to be "unconditionally stable" if both $\mathrm{Re}(Y_{in})>0$ and $\mathrm{Re}(Y_{out})>0$ for all passive source and load terminations. It is "potentially unstable" if both $\mathrm{Re}(Y_{in})>0$ and $\mathrm{Re}(Y_{out})>0$ only for a certain range of passive source and load terminations. The stability of a transistor amplifier as a two-port can be determined by its y parameters using the stability factor K [2]

$$K = \frac{2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21})}{|y_{12}y_{21}|} \tag{4.3}$$

where g_{11} and g_{22} are the real part of y_{11} and y_{22}, respectively. When K is greater than 1, the amplifier is unconditionally stable, and it is potentially unstable when K is less than 1. Note that when an amplifier is designed, the stability should be checked at each bias condition and all frequencies.

4.2.3 Power gains

The power gain of an amplifier can be defined in the following different ways:

Transducer power gain $$G_T = \frac{P_L}{P_{AVS}} = \frac{\text{Power delivered to the load}}{\text{Power available from the source}}$$

Operating power gain $$G_P = \frac{P_L}{P_{in}} = \frac{\text{Power delivered to the load}}{\text{Power input to the network}}$$

Available power gain $$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{\text{Power available from the network}}{\text{Power available from the source}}$$

The available power gain (G_A) is a function of two-port parameters and source admittance. The output admittance of the input-terminated two-port is also a function of these variables as indicated by (4.2). If the load admittance Y_L matches the output admittance Y_{out} by complex conjugation, the power delivered to the load will equal the power available from the output and the available gain (G_A) is equal to the transducer gain (G_T). Otherwise G_T will be less than G_A. In other words, G_A provides an upper bound for G_T. Thus the maximum of the available gain (*MAG*) is a very important property of a two-port. If it is finite, specific source and load terminations can be selected to give $G_T = G_P = G_A$, all of which are equal to *MAG*. If the maximum of the available gain is infinite, the two-port is potentially unstable and can be made to oscillate with the connection of some source and load terminations. Therefore, in this section, we formulate the available power gain and its maximum value in terms of y parameters and stability factor.

Using the expression of P_{AVN} and P_{AVS} for the equivalent circuit representation shown in Fig. 4.3, the available power gain can be expressed as [4]

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{|y_{21}|^2 G_s}{g_{22}|y_{11} + Y_s|^2 - \mathrm{Re}\left[y_{12} y_{21} (y_{11} + Y_s)^*\right]} \tag{4.4}$$

where G_S is the real part of source admittance Y_S. The maximum available gain (*MAG*) is an often used figure of merit for RF transistor amplifiers. For a transistor amplifier,

the available gain is a function of source admittance for the given two-port as indicated by (4.4). However, if the amplifier is potentially unstable, it may oscillate and the maximum gain will be infinite. In the following the expression for *MAG* will be given to cases of both unconditional stability and potential instability.

If the two-port network is unconditionally stable as indicated by the stability factor $K>1$, the available gain from (4.4) can be rearranged as

$$\frac{1}{G_A} = \frac{1}{MAG} + \frac{R_{eg}}{G_S}\left[\left(G_S - G_{og}\right)^2 + \left(B_S - B_{og}\right)^2\right] \tag{4.5}$$

where

$$R_{eg} = \frac{g_{22}}{|y_{21}|^2} \qquad MAG = \left|\frac{y_{21}}{y_{12}}\right| \frac{1}{K + \sqrt{K^2 - 1}}$$

$$G_{og} = \frac{|y_{12} y_{21}|}{2g_{22}} \sqrt{K^2 - 1} \qquad B_{og} = \frac{\mathrm{Im}(y_{12} y_{21})}{2g_{22}} - b_{11} \tag{4.6}$$

It can be seen that the available gain (G_A) has a maximum value *MAG* at a particular source admittance $Y_{og}=G_{og}+j*B_{og}$ and this *MAG* only exist when $K>1$, i.e., the two-port network is unconditionally stable.

However, if the two-port is potentially unstable with a stability factor $K<1$, there is no maximum value for the available gain, and the expression of *MAG* in (4.5) is invalid. In this case, the general expression of the available gain in (4.4) has to be used. It can be further rearranged as

$$\frac{1}{G_A} = \frac{R_{eg}}{G_S}\left[\left(G_S - G_g\right)^2 + \left(B_S - B_g\right)^2 - C_g^2\right] \tag{4.7}$$

where $G_g = \dfrac{\mathrm{Re}(y_{12}y_{21})}{2g_{22}} - g_{11}$ $B_g = \dfrac{\mathrm{Im}(y_{12}y_{21})}{2g_{22}} - b_{11}$ $C_g = \dfrac{|y_{12}y_{21}|}{2g_{22}}$

Note that the expression of (4.7) is equally valid for both potential instability and unconditional stability. However, the expression of (4.5) is usually preferred in case of unconditional stability since both the maximum value of G_A and the dependence on the source termination are straightforward with its quadratic format.

Obviously the available gain expressed in the form of (4.7), can be infinite for some source terminations, such as $G_S=G_g+C_g$ and $B_S=B_g$. When the transistor amplifier is potentially unstable, *MAG*, as defined in (4.6), does not exist since *K* is less than 1. An alternate figure of merit, maximum stable gain (*MSG*), is usually used when the two-port is potentially unstable and it is set to be equal to *MAG* with K=1 as

$$MSG = MAG\big|_{K=1} = \left|\frac{y_{21}}{y_{12}}\right| \tag{4.8}$$

where the last step is from the expression of *MAG* in (4.6).

4.2.4 Noise parameters

A useful measure of the noise performance of a transistor amplifier is its noise

figure, which is defined as the ratio of the total output noise power to the output noise power due to the input source. It is well known that the noise figure of a linear two-port driven by a signal source with an admittance of Y_S=G_S+j*B_S can be expressed as [5]

$$NF = NF_{\min} + \frac{R_{ef}}{G_S}\left[\left(G_S - G_{of}\right)^2 + \left(B_S - B_{of}\right)^2\right] \tag{4.9}$$

where NF_{min} is the minimum noise figure, R_{ef} is the equivalent noise resistance, Y_{of} = G_{of} + j*B_{of} is the optimal source admittance resulting in NF_{min}. Thus the noise figure characteristics of a linear two-port can be completely described by the four noise parameters: $F_{\min}$, R_{ef}, G_{of} and B_{of}. These parameters can be transformed from noise power spectral densities, which are commonly used in the noise model development and circuit simulator implementation.

The circuit theory of linear noisy networks shows that any noisy two-port can be replaced by a noise equivalent circuit which consists of the original two-port (now assumed to be noiseless) and two additional noise sources. Several types of representations are available depending on the voltage / current noise sources that have been chosen. A physically significant description of these sources is given by their self- and cross- power spectral densities which are defined as the Fourier transform of their auto- and cross- correlation functions. Arranging these spectral densities in matrix form leads to the so-called correlation matrices. The correlation matrices belonging to admittance, impedance and chain representations are shown in

Fig. 4.4. The elements of matrices are denoted by Cs_1s_2*, where the subscript indicates that the spectral density refers to the noise sources s_1 and s_2.

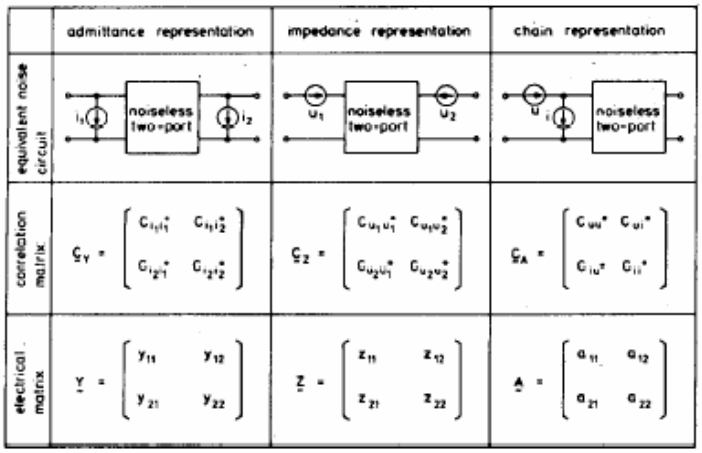

Fig 4.4 Correlation matrices of various representations for a linear noisy two-port network (adapted from H. Hillbrand *et al.* [6]).

It has been shown that the chain representation of the correlation matrix is related to the two-port noise parameters in (4.9) as given by [6]

$$C_A = \frac{1}{2\Delta f}\begin{bmatrix} \overline{u \cdot u^*} & \overline{u \cdot i^*} \\ \overline{i \cdot u^*} & \overline{i \cdot i^*} \end{bmatrix} = 2kT\begin{bmatrix} R_{ef} & \dfrac{F_{\min}-1}{2} - R_{ef}Y_{of} \\ \dfrac{F_{\min}-1}{2} - R_{ef}Y_{of}^* & R_{ef}\left|Y_{of}\right|^2 \end{bmatrix} \tag{4.10}$$

Once the correlation matrix is determined in its chain representation by circuit simulation, the noise parameters can be obtained using (4.10)

$$F_{\min} = 1 + \frac{C_{ui*} + C_{uu*} Y_{of}}{kT} \qquad R_{ef} = C_{uu*} / kT$$
$$G_{of} = \sqrt{\frac{C_{ii*}}{C_{uu*}} - \left[\mathrm{Im}\left(\frac{C_{ui*}}{C_{uu*}} \right) \right]^2} \qquad B_{of} = -\mathrm{Im}\left(\frac{C_{ui*}}{C_{uu*}} \right) \tag{4.11}$$

For the device simulator DESSIS in SYNOPSIS TCAD, the correlation matrix is given in the format of noise voltage spectral densities as in the impedance representation (Cz) shown in Fig. 4.4, and it can be transformed to the chain representation (C_A) by using transformation matrix T as

$$C_A = T \cdot C_Z \cdot T^H$$

where

$$T = \begin{bmatrix} 1 & -a_{11} \\ 0 & -a_{21} \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{y_{22}}{y_{21}} \\ 0 & -\dfrac{|y|}{y_{21}} \end{bmatrix}$$

Observe that the T matrix depends only on the two-port parameters and T^H is used to denote its Hermitian conjugation.

4.2.5 S parameter and the Smith chart

Scattering, or s parameters have greatly increased its popularity in RFIC design since the late 1960's, largely due to the appearance of sophisticated new equipment for performing s-parameter measurements. The results above are derived using y

parameters, and can be easily translated into s parameters using the conversion relationship as follows (all y parameters have been normalized with $1/Z_o$):

From y parameters to s parameters:

$$s_{11} = \frac{(1-y_{11})(1+y_{22})+y_{12}y_{21}}{(1+y_{11})(1+y_{22})-y_{12}y_{21}} \qquad s_{12} = \frac{-2y_{12}}{(1+y_{11})(1+y_{22})-y_{12}y_{21}}$$

$$s_{21} = \frac{-2y_{21}}{(1+y_{11})(1+y_{22})-y_{12}y_{21}} \qquad s_{22} = \frac{(1+y_{11})(1-y_{22})+y_{12}y_{21}}{(1+y_{11})(1+y_{22})-y_{12}y_{21}}$$

From s parameters to y parameters:

$$y_{11} = \frac{(1-s_{11})(1+s_{22})+s_{12}s_{21}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}} \qquad y_{12} = \frac{-2s_{12}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}}$$

$$y_{21} = \frac{-2s_{21}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}} \qquad y_{22} = \frac{(1+s_{11})(1-s_{22})+s_{12}s_{21}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}}$$

The Smith chart is the representation in the reflection coefficient plane, called the Γ plane, or the relation

$$\Gamma = \frac{Z - Z_o}{Z + Z_o}$$

for all values of Z such that Re(Z)>0. Z_o is the value of reference impedance, usually 50 Ω in RF communication systems. Defining the normalized impedance z as

$$z = \frac{Z}{Z_o} = \frac{R + jX}{Z_o} = r + jx$$

then

$$\Gamma = \frac{Z - Z_o}{Z + Z_o} = \frac{z - 1}{z + 1} = \frac{1 - y}{1 + y}$$

where y is the normalized admittance.

All previous results including input and output admittance, stability, power gains and noise figures derived above can be expressed in s parameters and plotted in Smith chart with circles for graphical representations. Below is a summary of the corresponding equations and formulas for the various design parameters:

Input/output reflection coefficient:

$$\Gamma_{in} = s_{11} + \frac{s_{12}s_{21}\Gamma_L}{1 - s_{22}\Gamma_L} \qquad \Gamma_{out} = s_{22} + \frac{s_{12}s_{21}\Gamma_S}{1 - s_{11}\Gamma_S}$$

Stability factor

$$K = \frac{1 - |s_{11}|^2 - |s_{22}|^2 + |\Delta|^2}{2|s_{12}s_{21}|} \qquad \text{where} \quad \Delta = s_{11}s_{22} - s_{12}s_{21}$$

Stability circles:

$$|\Gamma_S - C_S| = r_S \qquad \text{(input stability circle for } |\Gamma_{out}|=1)$$

$$|\Gamma_L - C_L| = r_L \qquad \text{(output stability circle for } |\Gamma_{in}|=1)$$

where

$$C_S = \frac{\left(s_{11} - \Delta s_{22}^*\right)^*}{\left|s_{11}\right|^2 - \left|\Delta\right|^2} \qquad r_S = \left|\frac{s_{12}s_{21}}{\left|s_{11}\right|^2 - \left|\Delta\right|^2}\right|$$

$$C_L = \frac{\left(s_{22} - \Delta s_{11}^*\right)^*}{\left|s_{22}\right|^2 - \left|\Delta\right|^2} \qquad r_L = \left|\frac{s_{12}s_{21}}{\left|s_{22}\right|^2 - \left|\Delta\right|^2}\right|$$

Power gains:

$$G_T = \frac{1-\left|\Gamma_S\right|^2}{\left|1-\Gamma_{in}\Gamma_S\right|^2}\left|s_{21}\right|^2 \frac{1-\left|\Gamma_L\right|^2}{\left|1-s_{22}\Gamma_L\right|^2}$$

$$G_P = \frac{1}{1-\left|\Gamma_{in}\right|^2}\left|s_{21}\right|^2 \frac{1-\left|\Gamma_L\right|^2}{\left|1-s_{22}\Gamma_L\right|^2}$$

$$G_A = \frac{1-\left|\Gamma_S\right|^2}{\left|1-s_{11}\Gamma_S\right|^2}\left|s_{21}\right|^2 \frac{1}{1-\left|\Gamma_{out}\right|^2}$$

Constant available power gain (G_A) circles

$$\left|\Gamma_S - C_a\right| = r_a$$

where

$$C_a = \frac{g_a C_1^*}{1 + g_a\left(\left|s_{11}\right|^2 - \left|\Delta\right|^2\right)} \qquad r_a = \frac{\left[1 - 2K\left|s_{12}s_{21}\right|g_a + \left|s_{12}s_{21}\right|^2 g_a^2\right]^{1/2}}{\left|1 + g_a\left(\left|s_{11}\right|^2 - \left|\Delta\right|^2\right)\right|}$$

$$g_a = \frac{G_A}{\left|s_{21}\right|^2} \qquad C_1 = s_{11} - \Delta s_{22}^*$$

Noise figure

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

where

$$r_n = R_{ef} / Z_o$$

Constant noise figure circles

$$\left|\Gamma_S - C_{F_i}\right| = r_{F_i}$$

where

$$C_{F_i} = \frac{\Gamma_{opt}}{1+N_i} \qquad r_{F_i} = \frac{1}{1+N_i}\sqrt{N_i^2 + N_i\left(1-\left|\Gamma_{opt}\right|^2\right)}$$

$$N_i = \frac{F_i - F_{\min}}{4r_n}\left|1+\Gamma_{opt}\right|^2$$

Since both the constant available power gain circles and the constant noise figure circles are function of Γ_S, they can be plotted together on the same Smith chart and the trade-offs between gain and noise figure can be easily analyzed in a graphical environment. These results will be presented in Chapter 5 in details.

4.3 METHODOLOGY OF THE NOISE MEASURE APPROACH

4.3.1 Noise measure and noise measure figure

It is a common practice to describe the noise performance of a two-port amplifier in terms of its noise figure. As shown in (4.9), this noise figure has a minimum value for some particular choice of its source admittance ($Y_S=Y_{of}$). If with this source admittance Y_{of}, the gain of a given amplifier remains sufficiently large, its noise figure will prescribe the overall noise figure of any amplifier cascade in which it is used as the first stage as a result of Friis' formula [7]. In this way, it is possible to build an amplifier cascade with any desired high gain, and with a noise figure set by the minimum noise figure (with respect to source admittance) of the first-stage amplifier.

However, the validity of the above argument is based on the assumption that the amplifier has enough high gain when driven from the optimum source that yields the minimum noise figure. And it is clear that the condition of optimal source terminations for the minimum noise figure as given in (4.11) are not necessarily the same as that for the maximum available gain in (4.6). In practice, designers must generally accept less than maximum power gain if noise performance is to be optimized, and vice versa. Thus noise figure alone is a poor figure of merit since an amplifier designed to have the minimum noise figure may or may not have useful gain [8]. In communication systems, a low-noise amplifier usually contains several stages

to provide enough gain to signals received from the antenna. A compromise between stage gain and the noise figure must be made if the overall noise figure of such a multistage amplifier is to be minimized.

It appears that the characterization of the amplifier quality must include the specification of the gain in addition to the specification of the noise figure and source admittance. Consequently, noise measure has been proposed by Haus and Adler [9] as a single quantitative measure of the absolute quality of amplifier noise performance. It is defined as

$$M = \frac{NF - 1}{1 - \frac{1}{G_A}} \tag{4.12}$$

where F is the noise figure and G_A is the available gain, and M is the noise measure of the two-port amplifier. An alternative expression is the noise measure figure, which is given in the unit of decibels as

$$NMF = 10 \cdot \log_{10}(1 + M) = 10 \cdot \log_{10}(1 + \frac{NF - 1}{1 - 1/G_A}) \tag{4.13}$$

Since both the NF and G_A are functions of the same source admittance (Y_S), M will also be a function of that source admittance. For any transistor, the minimum noise measure figure (NMF_{min}) with respect to the source impedance is a key figure of merit for its overall noise performance. It provides the upper limit of the noise figure for an

identical amplifier cascade with any desired high gain, and the design approach using NMF_{min} is especially useful for circuit benchmarking since the trade-off between the noise and gain is automatically accounted for. The minimization algorithm that leads to NMF_{min} is presented in the follow section. Since the expression of G_A depends on the two-port stability, special consideration needs to be given as M is formulated.

4.3.2 Minimization algorithm

Solution of minimum noise measure as a function of source admittance with the implicit assumption that the two-port is unconditionally stable has been done by Fukui [4]. Substituting (4.5) and (4.9) into (4.12), the noise measure can be expressed in terms of gain parameters, noise parameters and the source admittance as follows:

$$M = \frac{NF - 1}{1 - \frac{1}{G_A}} = \frac{NF_{\min} - 1 + \frac{R_{ef}}{G_S}\left[\left(G_S - G_{of}\right)^2 + \left(B_S - B_{of}\right)^2\right]}{1 - \frac{1}{MAG} - \frac{R_{eg}}{G_S}\left[\left(G_S - G_{og}\right)^2 + \left(B_S - B_{og}\right)^2\right]} \tag{4.14}$$

Rearranging (4.14), the following equation can be derived

$$\left(G_s - G_M\right)^2 + \left(B_s - B_M\right)^2 = G_{RM}^2$$

For the condition $G_{RM} = 0$, M takes on its minimum value M_{min}, i.e.,

$$M_{\min} = \frac{M_2}{M_1}\left[1 + \left(1 - \frac{M_1 M_3}{M_2^2}\right)^{1/2}\right]$$

where

$$M_1 = \left(1 - \frac{1}{MAG}\right)^2 + 4\left(1 - \frac{1}{MAG}\right) R_{eg} G_{og}$$

$$M_2 = \left(1 - \frac{1}{MAG} + 2R_{eg} G_{og}\right)\left(NF_{\min} - 1 - 2R_{ef} G_{of}\right) + 2R_{eg} R_{of}\left(\left|Y_{og}\right|^2 + \left|Y_{of}\right|^2 - 2B_{og} B_{of}\right)$$

$$M_3 = \left(NF_{\min} - 1\right)^2 - 4\left(NF_{\min} - 1\right) R_{ef} G_{of}$$

The source conductance and susceptance which produce the minimum noise measure are given by

$$G_{om} = \frac{M_{\min}\left(2R_{eg} G_{og} - \frac{1}{MAG} + 1\right) + 2R_{ef} G_{of} - NF_{\min} + 1}{2\left(M_{\min} R_{eg} + R_{ef}\right)}$$

$$B_{om} = \frac{M_{\min} R_{eg} B_{og} + R_{ef} B_{of}}{M_{\min} R_{eg} + R_{ef}}$$

For today's transistors at multi-giga-herz frequency range, the unconditional stability can not be always guaranteed and designers are constantly facing RF circuit design under potential instability. As a result, it would be beneficial to develop a design algorithm without any assumption on the circuit stability. For that purpose, the available power gain can be expressed as (4.7) and the noise measure can thus be obtained as

$$M = \frac{NF - 1}{1 - \frac{1}{G_A}} = \frac{NF_{\min} - 1 + \frac{R_{ef}}{G_S}\left[\left(G_S - G_{of}\right)^2 + \left(B_S - B_{of}\right)^2\right]}{1 - \frac{R_{eg}}{G_S}\left[\left(G_S - G_g\right)^2 + \left(B_S - B_g\right)^2 - C_g^2\right]}$$

Following the similar optimization method used in the case of unconditional stability, it can be shown that the minimum noise measure M_{min} under arbitrary circuit stability is given by:

$$M_{\min} = \frac{M_2}{M_1}\left[1+\left(1-\frac{M_1 M_3}{M_2^2}\right)^{1/2}\right]$$

where

$$M_1 = 1 + 4R_{eg}G_g + 4R_{eg}^2 C_g^2$$

$$M_2 = \left(1+2R_{eg}G_g\right)\left(NF_{\min}-1-2R_{ef}G_{of}\right)+2R_{eg}R_{of}\left(\left|Y_g\right|^2+\left|Y_{of}\right|^2-C_g^2-2B_gB_{of}\right)$$

$$M_3 = \left(NF_{\min}-1\right)^2 - 4\left(NF_{\min}-1\right)R_{ef}G_{of}$$

The source conductance and susceptance which produce the minimum noise measure are

$$G_{om} = \frac{M_{\min}\left(2R_{eg}G_g+1\right)+2R_{ef}G_{of}-F_{\min}+1}{2\left(M_{\min}R_{eg}+R_{ef}\right)}$$

$$B_{om} = \frac{M_{\min}R_{eg}B_g + R_{ef}B_{of}}{M_{\min}R_{eg}+R_{ef}}$$

In this work, the CMOS LNAs are designed using the noise measure approach, and the NMF_{min} are calculated with no assumption on the circuit stability. This approach is particularly useful when the circuit stability changes as a function of operation conditions such as frequency, or other parameters that may affect the stability such as gate resistance of MOS transistors.

4.4 LNEARITY ANALSIS OF CMOS LNA

4.4.1 Power series expansion

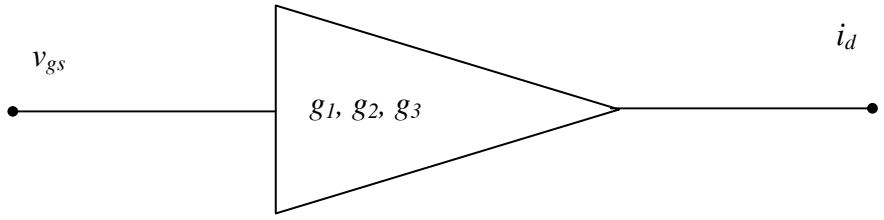

Fig. 4.5 Generalized representation of a CMOS LNA with its power series coefficients (g_1, g_2 and g_3).

Consider a CMOS LNA made by a nMOSFET in a common-source configuration as generalized in Fig. 4.5. Its small-signal output current (i_d) can be expanded into the following power series in terms of the small-signal voltage (v_{gs}) around the DC bias point (V_{GS}):

$$i_d(v_{gs}) = g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 + \cdots \tag{4.15}$$

where

$$g_1 = \frac{\partial I_D}{\partial V_{GS}} = g_m \qquad g_2 = \frac{1}{2}\cdot\frac{\partial^2 I_D}{\partial V_{GS}^2} = \frac{g_{m2}}{2} \qquad g_3 = \frac{1}{6}\cdot\frac{\partial^3 I_D}{\partial V_{GS}^3} = \frac{g_{m3}}{6} \tag{4.16}$$

g_1 is the small-signal trans-conductance (g_m), and the higher-order coefficients (g_2, g_3, etc.) are introduced by the transistor nonlinearity and they describe the magnitudes of the corresponding nonlinearities. g_{m2} and g_{m3} are defined as the 2nd and 3rd -order derivatives of g_m, respectively.

4.4.2 Harmonic distortion

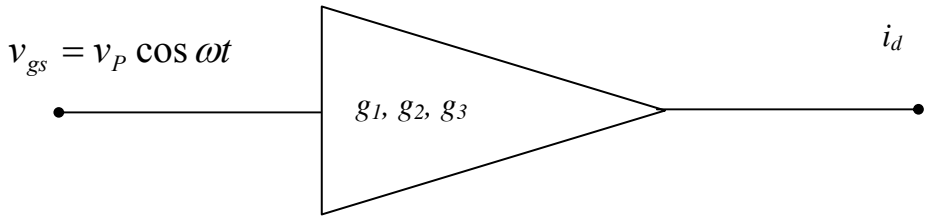

Fig. 4.6 Generalized representation of a CMOS LNA with its power series coefficients (g_1, g_2 and g_3) and a small-signal sinusoid input voltage v_{gs}.

The nonlinearity of a LNA can be characterized by applying a small sinusoid signal at the input and measuring the harmonic content of the output as shown in Fig. 4.6. Using (4.15) gives the output current as

$$\begin{aligned} i_d &= g_1 v_P \cos \omega t + g_2 (v_P \cos \omega t)^2 + g_3 (v_P \cos \omega t)^3 + \cdots \\ &= \frac{1}{2} g_2 v_P^2 + (g_1 v_P + \frac{3}{4} g_3 v_P^3) \cos \omega t + \frac{1}{2} g_2 v_P^2 \cos 2\omega t + \frac{1}{4} g_3 v_P^3 \cos 3\omega t + \cdots \end{aligned}$$

Magnitude of the desired signal (at frequency ω): $g_1 v_P + \frac{3}{4} g_3 v_P^3$

Magnitude of the 2nd-order harmonic distortion (at 2ω): $\frac{1}{2} g_2 v_P^2$

Magnitude of the 3rd-order harmonic distortion (at 3ω): $\frac{1}{4} g_3 v_P^3$

Although many measures of linearity are available, the most commonly used are 3rd-order input intercept point (*IIP3*) and 1-dB compression point [5]. The *IIP3* for harmonic distortion is defined as the input value where the extrapolated 3rd-order

harmonic component and the desired linear response intersect. It is straightforward from the forgoing equations to calculate the *IIP3* by setting the magnitude of the 3rd-order harmonic equal to that of the linear fundamental term:

$$\frac{1}{4}g_3 v_P^3 = g_1 v_P + \frac{3}{4}g_3 v_P^3 \approx g_1 v_P$$

which yields (using (4.16) in the last step)

$$V_{IIP3} = \sqrt{4 \cdot \left|\frac{g_1}{g_3}\right|} = \sqrt{24 \cdot \left|\frac{g_m}{g_{m3}}\right|} \tag{4.17}$$

The 1-dB compression point for the harmonic distortion is defined as the input level where the gain of the fundamental term has been reduced from the linear extrapolation by 1 dB. The V_{1dB} for the forgoing CS LNA can be calculated as follows:

$$20\log_{10}\frac{g_1 v_P + \frac{3}{4}g_3 v_P^3}{g_1 v_P} = -1$$

from which the 1-dB compression point can be determined as

$$V_{1dB} = \sqrt{0.11} \cdot \sqrt{\frac{4}{3}\left|\frac{g_1}{g_3}\right|} = \sqrt{0.11} \cdot \sqrt{8 \cdot \left|\frac{g_m}{g_{m3}}\right|} \tag{4.18}$$

The V_{1dB} characterizes the linearity on a high-level limitation because it describes the input level where the actual output is 11% below the linear extrapolation of small-signal behavior, whereas the V_{IIP3} is considered as a low-level measure of linearity since it is measured with the signal amplitude small enough to cause insignificant

departures from linear operation. The V_{IIP3} and V_{1dB} are related to each other as can be seen from the definitions of V_{IIP3} in (4.17) and V_{1dB} in (4.18), thus either one of them can be used as a figure of merit for the 1st-order linearity analysis of CMOS LNA.

4.4.3 Intermodulation distortion

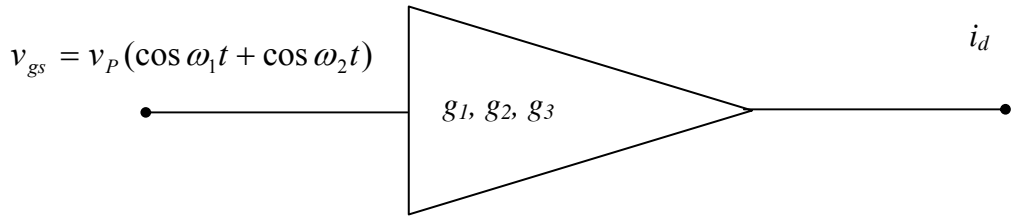

Fig. 4.7 Generalized representation of a CMOS LNA with its power series coefficients (g_1, g_2 and g_3) under a two-tone test input voltage

In practice wireless communication engineers are also concerned with another form of distortion known as intermodulation distortion, where the test is done with a multiple tone input and the tones intermodulate with each other. Fig. 4.7 shows a CS LNA under a two-tone test with equal tone magnitude. The various frequency components can be calculated in the similar way as that for the harmonic distortion. Below is a summary of the primary results:

Magnitude of the desired signal (at fundamental frequency ω_1 or ω_2): $g_1 v_P$

Magnitude of the 3rd-order inter-modulation products (at $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$): $\frac{3}{4} g_3 v_P^3$

The *IIP3* for the intermodulation distortion can be calculated by equating the

magnitude 3rd-order intermodulation products to that of the fundamental term as:

$$\frac{3}{4} g_3 v_P^3 = g_1 v_P$$

which yields

$$V_{IIP3} = \sqrt{\frac{4}{3} \cdot \left| \frac{g_1}{g_3} \right|} = \sqrt{8 \cdot \left| \frac{g_m}{g_{m3}} \right|} \tag{4.19}$$

Note that there is a difference by a factor of the square root of 3 between the V_{IIP3} for the harmonic distortion in (4.17) and intermodulation distortion in (4.19). Since both of them originate from the nonlinearity of the transistor *I-V* characteristics, either one of them is appropriate for a first-order linearity analysis of CMOS LNAs. In this work, the formula in (4.17) is used in line with the data presented in literatures [11, 12].

The text of this chapter, in part, is a reprint of the material that has been submitted for publication: Cai, Ming; Liu, Minjian; Taur, Yuan. “A design study of nanometer-gate low-noise amplifier near the limit of CMOS scaling”, Solid-State Electronics; Cai, Ming; Liu, Minjian; Taur, Yuan. “Design study of power-efficient 60-GHz MOS low noise amplifiers near scaling limits”, Electronics Letters. The dissertation author was the primary author and the co-authors listed in these publications directed and supervised the research which forms the basis for this chapter.

REFERENCES

1. G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey, 1997.

2. R. Hejhall, "RF small signal design using two-port parameters", Application note 215A, Motorola Semiconductor Product, Inc., 1993.

3. J. G. Linvill and J. F. Gibbons, *Transistors and active circuits*, McGraw-hill, 1961.

4. H. Fukui, "Available power gain, noise figure, and noise measure of two-ports and their graphical representations", *IEEE Trans. Circuit Theory*, vol. ct-13, no. 2, pp. 137-142, 1996.

5. T. H. Lee, *The design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed., Cambridge University Press, 2004.

6. H. Hillbrand and P. H. Russer, "An efficient method for computer aided noise analysis of linear amplifier networks", *IEEE Trans. Circuits Syst.*, vol. cas-23, no. 4, pp.235-238, 1976.

7. T. H. Lee, *Planar microwave engineering*, Cambridge University Press, New York, 2004.

8. C. R. Poole and D. K. Paul, "Optimum noise measure terminations for microwave transistor amplifiers", *IEEE Trans. Microw. Theory Tech.*, vol. MTT-33, no. 11, pp. 1254-1257, 1985.

9. H. A. Haus and R. B. Adler, "Optimum noise performance of linear amplifiers", *Proc. IRE*, vol. 46, pp. 1517-1533, 1958.

10. R. Van Langevelde et al., "RF-distortion in deep-submicron CMOS technologies", *IEDM Tech. Dig.*, pp. 807-810, Dec. 2000.

11. C. Yu, J. S. Yuan, and H. Yang, “MOSFET linearity performance degradation subject to drain and gate voltage stress”, *IEEE Trans. Device Mater. Rel.*, vol. 4, no. 4, pp. 681-689, 2004.

CHAPTER 5

40-NM NMOS LNA DESIGN AND ANALYSIS

5.1 MOSFET SMALL-SIGNAL MODELS

5.1.1 Intrinsic model

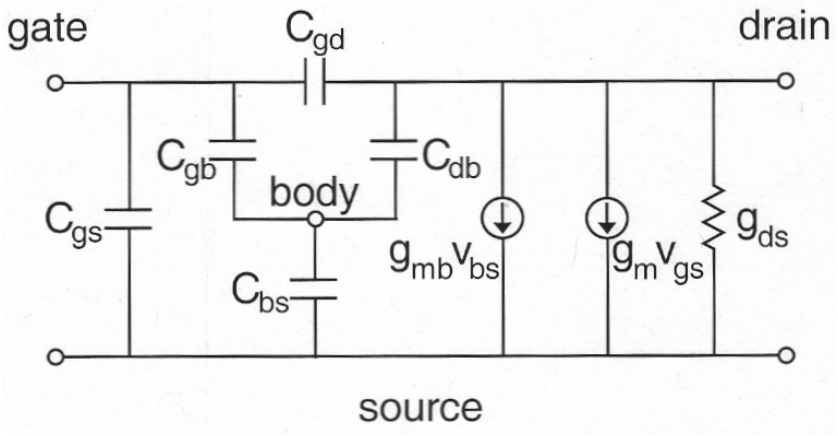

Fig. 5.1 A small-signal RF equivalent circuit for a intrinsic MOSFET with its four terminals: gate, source, drain and body.

Fig. 5.1 gives a small-signal RF equivalent circuit for an intrinsic MOSFET with its four terminals: gate, source, drain and body. It includes the conductance between source and drain (g_{ds}), the transconductance between gate and source (g_m), transconductance between body and source (g_{mb}), and various capacitances between its terminals (C_{gs}, C_{gb}, C_{gd}, C_{bs} and C_{bd}). For the common-source configuration of LNA, the body is tied with the source and both are grounded, thus the circuit in Fig. 5.1 can be simplified by lumping g_{mb} into g_m, C_{gb} into C_{gs} with C_{bs} eliminated altogether as shown in Fig. 5.2. There is a concern that the resistance between the body and ground is ignored in this model. In actual situations, the connection to the body is often a key issue in getting the correct model for mm-wave CMOS technologies. However, this

effect is not taken into consideration for the purpose of scaling trend prediction.

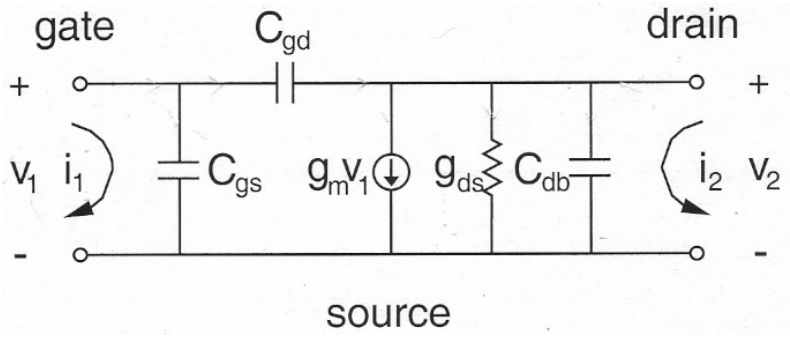

Fig. 5.2 A simplified small-signal RF equivalent circuit for an intrinsic MOSFET with its three terminals: gate, source, and drain. The body is tied with its source.

From the definition of two-port y parameters in (4.1), it can be derived that the y matrix for the equivalent circuit shown in Fig. 5.2 can be expressed as

$$[\boldsymbol{Y}] = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \begin{bmatrix} j\omega(C_{gs} + C_{gd}) & -j\omega C_{gd} \\ g_m - j\omega C_{gd} & g_{ds} + j\omega(C_{db} + C_{gd}) \end{bmatrix} \tag{5.1}$$

5.1.2 Extrinsic model

For a real MOSFET device, there are extrinsic resistances associated with the gate, drain and source terminals. In this work, the parasitic resistances are modeled by lumped resistors between the intrinsic model and corresponding terminals as shown in Fig. 5.3. It includes the extrinsic resistances associated the gate (R_g), drain (R_d) and source (R_s) terminals.

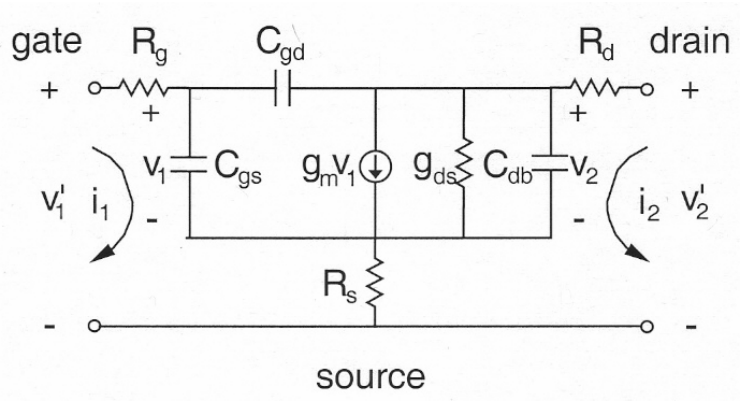

Fig. 5.3 A simplified small-signal RF equivalent circuit for an extrinsic MOSFET with the body tied with its source.

If it is assumed that the source / drain resistance is insignificant as compared with the gate resistance (R_g), then to the first order of R_g, the extrinsic y matrix [Y'] can be obtained from the intrinsic y matrix [Y] by matrix manipulation as given by

$$[Y'] = \left\{ [Y]^{-1} + \begin{bmatrix} R_g & 0 \\ 0 & 0 \end{bmatrix} \right\}^{-1} \tag{5.2}$$

The relationship between a matrix and its inverse is given by

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}^{-1} = \frac{1}{|Y|} \begin{bmatrix} y_{22} & -y_{12} \\ -y_{21} & y_{11} \end{bmatrix} \quad \text{, where} \quad |Y| = y_{11}y_{22} - y_{12}y_{21} \tag{5.3}$$

By using (5.3) in (5.2), we have

$$\left[Y'\right]=\left\{\left[Y\right]^{-1}+\begin{bmatrix}R_g & 0\\ 0 & 0\end{bmatrix}\right\}^{-1}=\left\{\frac{1}{|Y|}\begin{bmatrix}y_{22} & -y_{12}\\ -y_{21} & y_{11}\end{bmatrix}+\begin{bmatrix}R_g & 0\\ 0 & 0\end{bmatrix}\right\}^{-1}=\left\{\frac{1}{|Y|}\begin{bmatrix}y_{22}+R_g|Y| & -y_{12}\\ -y_{21} & y_{11}\end{bmatrix}\right\}^{-1}$$
$$=|Y|\cdot\begin{bmatrix}y_{22}+R_g|Y| & -y_{12}\\ -y_{21} & y_{11}\end{bmatrix}^{-1}$$

Using (5.3) again

$$\left[Y'\right]=\frac{|Y|}{R_g y_{11}|Y|+y_{11}y_{22}-y_{12}y_{21}}\begin{bmatrix}y_{11} & y_{12}\\ y_{21} & y_{22}+R_g|Y|\end{bmatrix}$$
$$=\frac{1}{1+R_g y_{11}}\begin{bmatrix}y_{11} & y_{12}\\ y_{21} & y_{22}+R_g|Y|\end{bmatrix} \quad (5.4)$$

where y_{ij} (i, j=1, 2) is the elements of the intrinsic y matrix given by (5.1).

Although the two-port parameters for the extrinsic MOSFET can be directly extracted from the 2D simulation, it is the small-signal model that provides insight to the physics behind the device operation as well as circuit performance. The general design analysis has been demonstrated by using the two-port y parameters in previous chapters, and more considerations with respect to the design of power-constrained LNA will be presented in this chapter with insights from the small-signal models shown in Fig. 5.2 for the intrinsic MOSFET with its y parameters in (5.1), and Fig. 5.3 for the extrinsic MOSFET with its y parameters in (5.4).

5.2 DESIGN OF DC BIAS CONDITIONS FOR POWER-EFFICIENT OPERATIONS

5.2.1 Drain current characteristic and constant power contours

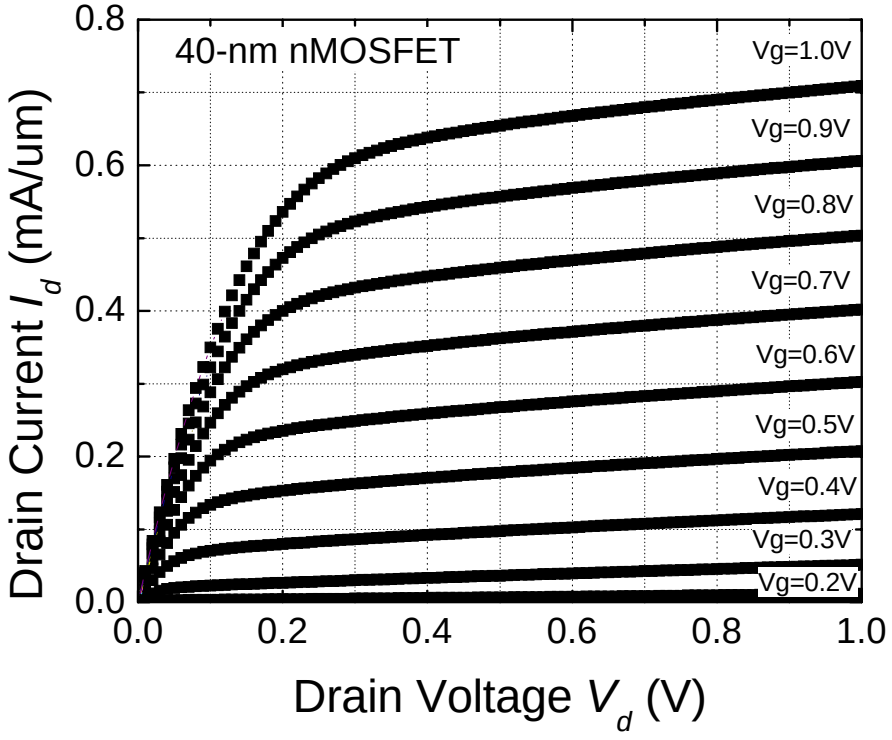

Fig. 5.4 Drain current as a function of grain voltage (I_d-V_d) at different values of gate voltage (V_g) for the 40-nm nMOSFET with a threshold voltage (V_t) of 0.23 V.

Fig. 5.4 shows the drain current characteristics of the 40-nm nMOSFET with V_t=0.23 V. For a given gate voltage (V_g), Drain current (I_d) is plotted as a function of drain voltage (V_d). All results thereafter in this chapter are based on such a nMOSFET. If the power consumption of a MOSFET is indicated by the product of the drain

current (I_d) and drain voltage (V_d), then a set of constant power contours can be plotted in the I_d-V_d design space, as shown in Fig. 5.5 for three power levels at 0.01, 0.05 and 0.1 mW/μm. The construction of these constant power contours facilitates the choice of power-constrained bias conditions as will be shown in later sections.

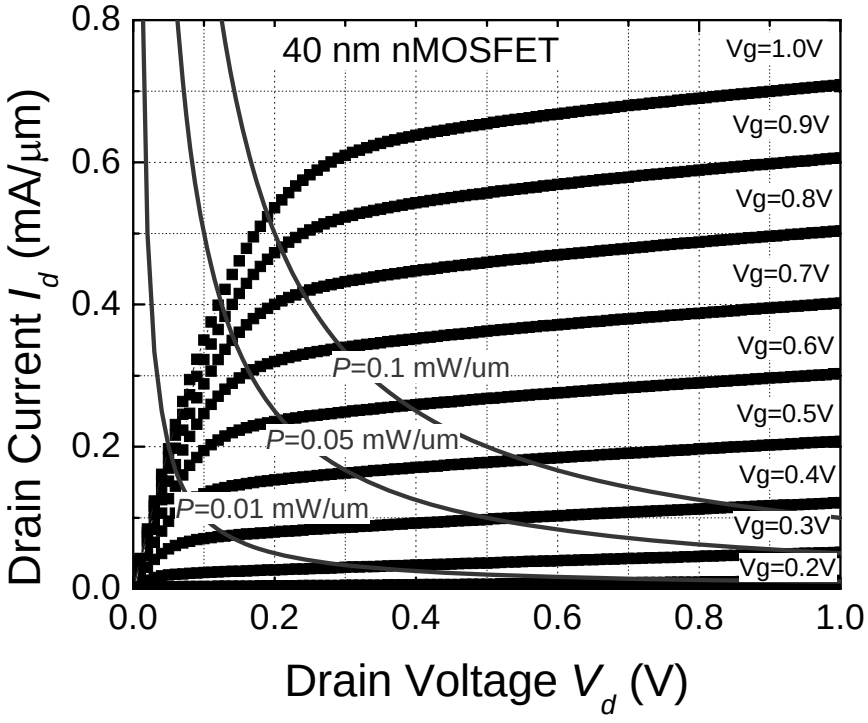

Fig. 5.5 Constant power contours in I_d-V_d design space for the 40-nm nMOSFET. Contours shown are for powers at 0.01, 0.05 and 0.1 mW/μm, respectively.

5.2.2 Current gain and cut-off frequency

Cut-off frequency (f_T) has been widely used to demonstrate device behavior at high frequencies and validate models. It is defined as the frequency at which the current gain (h_{21}), which is equal to a ratio of y_{21} to y_{11}, of the device equals to 1. Fig. 5.6 illustrates the frequency response of current gain (h_{21}) for the 40-nm nMOSFET at a gate bias of 0.5 V and a drain bias of 1 V. A direct search for the unit-gain point from Fig. 5.6 gives a cut-off frequency (f_T) of 207 GHz for this device under such particular bias.

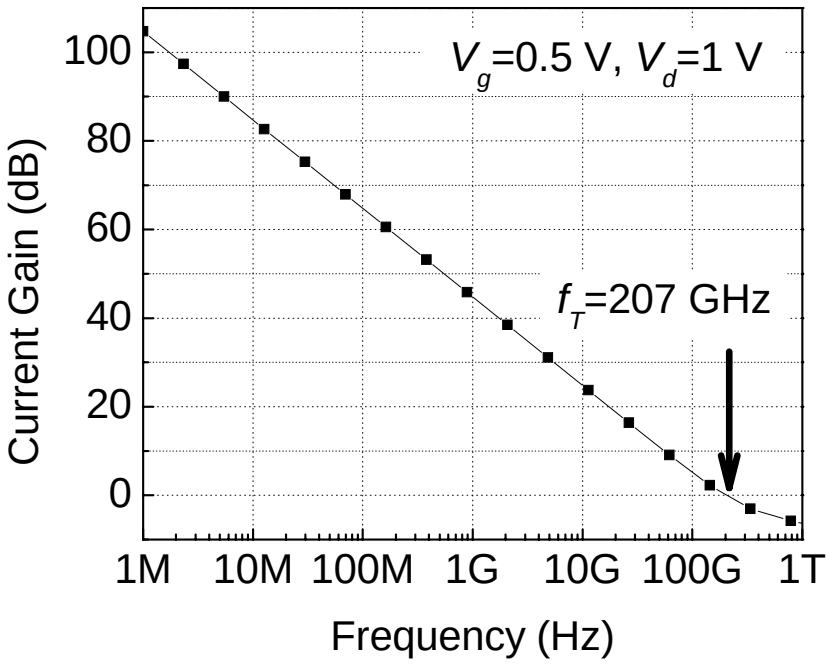

Fig. 5.6 Current gain as a function of frequency for the 40-nm nMOSFET under a gate bias V_g=0.5 V and drain bias V_d=1 V. The cut-off frequency (f_T) is determined to be 207 GHz in this case.

Let us look at the small-signal representation of current gain using the intrinsic two-port y parameters from (5.1):

$$|h_{21}| = \left|\frac{y_{21}}{y_{11}}\right| = \left|\frac{g_m - j\omega C_{gd}}{j\omega(C_{gs} + C_{gd})}\right| \approx \frac{g_m}{2\pi \cdot f \cdot (C_{gs} + C_{gd})} \tag{5.5}$$

where it is assumed that the device is operating at frequencies much lower than g_m/C_{gd}. A similar calculation of the current gain from the extrinsic model based on (5.4) reveals the same result as for the intrinsic case, indicating that, for the first-order analysis, the cut-off frequency has no dependence on the gate resistance. This allows the simplification of simulation and the following results are presented based on the intrinsic device unless stated otherwise. The expression of (5.5) not only confirms the slope of 20 dB/dec for the current gain at low frequencies shown in Fig. 5.6, but also enables us to find the expression of f_T at which the current gain is unity:

$$f_T = \frac{g_m}{2\pi \cdot (C_{gs} + C_{gd})} \tag{5.6}$$

(5.6) suggests that high transconductance and low capacitance are desirable for a MOSFET device to have a high cut-off frequency. For a given technology, the bias condition is very important as it can give a significant change to the transconductance. Fig. 5.7 plots the cut-off frequency (f_T) as a function of gate bias for the 40-nm nMOSFET. The simulation was done at a drain bias V_d = 1 V while sweeping the gate bias V_g.

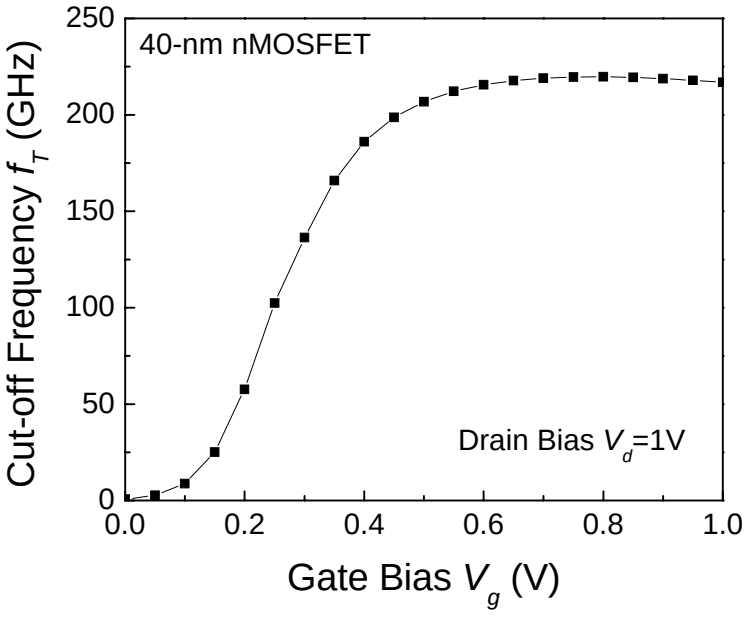

Fig. 5.7 Cut-off frequency (f_T) as a function of gate bias (V_g) for the 40-nm nMOSFET at a fixed drain bias V_d=1 V.

A f_T of over 200 GHz is obtained at V_g as low as 0.45 V for the 40-nm nMOSFET, which allows low-power mm-wave RF circuit design. The simulation result is consistent with the latest hardware data, which achieves a record f_T of 330 GHz from a 65-nm SOI CMOS technology featuring a 35 nm gate length [1]. Fig. 5.7 also shows that f_T increases with V_g until V_g~V_t+0.25 V where f_T starts to saturate at approximately 220 GHz. The decrease of f_T at very high V_g is caused by the mobility degradation at high field.

5.2.3 Maximum power gains

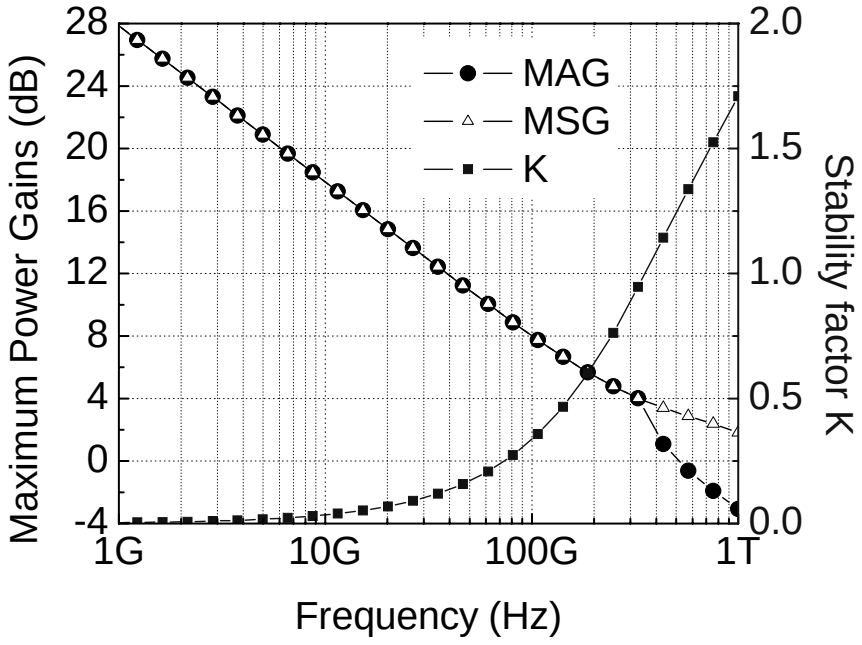

Fig. 5.8 Maximum available / stable power gains (*MAG* / *MSG*) and stability factor (*K*) as a function of frequency for the 40-nm nMOS LNA at a gate bias V_g=0.5 V and a drain bias V_d=1 V.

Fig. 5.8 plots the frequency responses of maximum available / stable power gains (*MAG* / *MSG*) and the stability factor (*K*), which are obtained from the two-port y parameters using (4.3), (4.6) and (4.8). The simulation was done for the 40-nm nMOS LNA at a gate bias V_g=0.5 V and a drain bias V_d=1 V. When circuit is potentially unstable (K<1), *MSG* needs to be used to describe the gain performance of the LNA as *MAG* is not defined there. However, if the circuit is unconditionally stable (K>1), only *MAG* is meaningful since *MSG*, by definition, assumes an unrealistic K=1.

Finally, for the mm-wave frequency around 60 GHz that we are interested, the *MSG* is 10 dB at this particular bias condition. In some part of this work, the *MAG* may be used interchangeably with *MSG* in case of $K<1$, where the notation of *MAG* is meant to be *MSG* with the same numerical value.

The focus of this study is to design a nMOS LNA for power-efficient operations. Fig. 5.9 shows the maximum available gain (*MAG*) as a function of power consumption for the 40-nm nMOS LNA. The *MAG* is plotted for different values of V_g at a fixed frequency of 60 GHz, with V_d varied to sweep a range of power levels at each V_g. If the transistor is biased in the sub-threshold condition (V_g=0.2 V as in Fig. 5.9), *MAG* is limited to be under 5 dB by the power supply of 1 V and the amplifier can not provide sufficient gain for practical use. For a given *MAG*, lowest power is needed when transistor is biased just slightly above V_t (V_g=0.4 V). Similarly, for a given low power, higher value of *MAG* is achieved by limiting V_g within 0.2-0.3 V above V_t with a sufficiently high V_d. Such a bias preference is also beneficial for output resistance, which is larger in the saturation region with low V_g and high V_d.

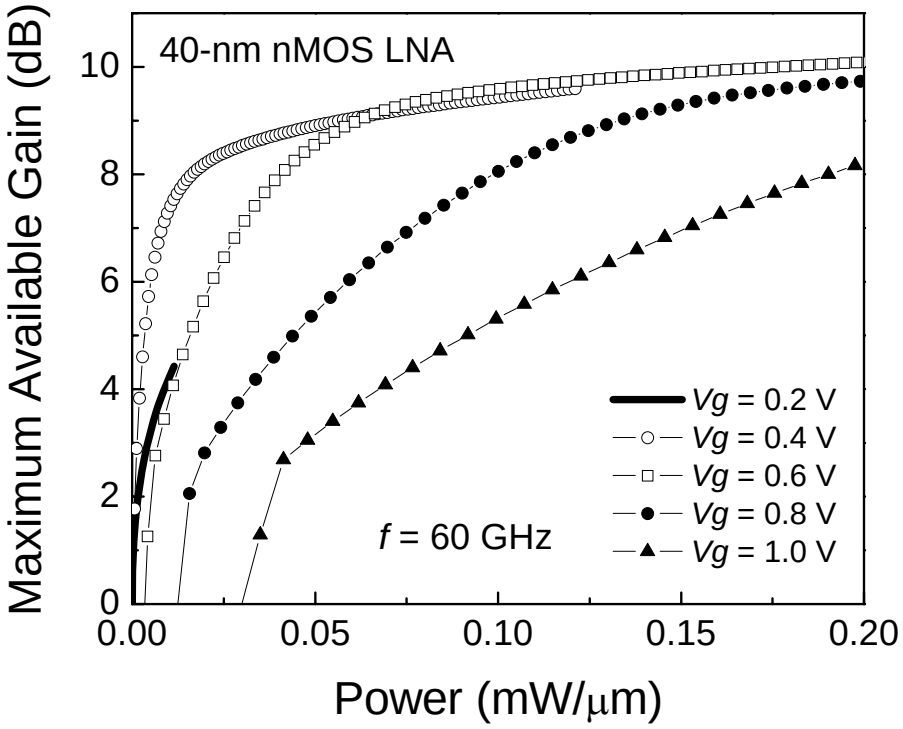

Fig. 5.9 Maximum available gain (*MAG*) at 60 GHz as a function of power consumption for the 40-nm nMOS LNA with a threshold voltage V_t=0.23 V for different gate bias V_g from 0.2 to 1 V.

5.2.4 Noise performance of a power-constrained LNA

It is a common practice to describe an amplifier's noise performance in terms of its noise figure (*NF*). However, as developed in Chapter 4, neither noise figure nor available gain alone is sufficient to characterize an amplifier's performance. Minimum noise measure figure (NMF_{min}) is proposed as a better figure of merit for the overall noise performance than minimum noise figure (NF_{min}), and the specific source terminations that lead to both of these figures are formulated there.

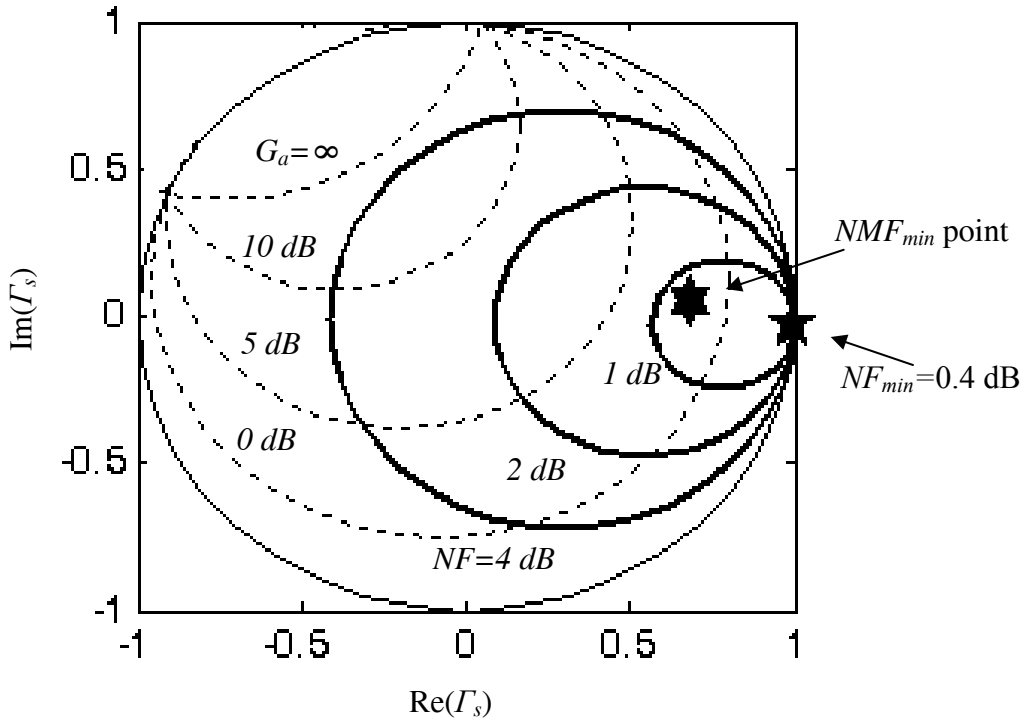

Fig. 5.10 Smith chart representation of constant noise figure *NF* circles (solid lines) and constant available gain G_a circles (dashed lines) in the plane of source reflection coefficient (Γ_s) for the 60-GHz LNA with a 40-nm nMOSFET biased at V_g=0.5 V and V_d= 1 V.

Fig. 5.10 illustrates the dependence of both the noise figure (*NF*) and the available power gain (G_a) on the source admittance for the 60-GHz nMOS LNA in a Smith chart. It plots both the constant noise figure and constant available gain circles in the plane of source reflection coefficient using equations derived in section 4.2.5 using s parameter representations. The 40-nm nMOSFET is biased at V_g=0.5 V and V_d=1 V. As shown in Fig. 5.10, an amplifier designed to have the minimum noise figure (NF_{min}), 0.4 dB as pointed out in this case, will have no useful gain (G_a<1). On the other hand, an amplifier designed to have near maximum gain will eventually have

excessively large noise figures. In practice, the maximum gain is usually traded off for acceptable noise figure. For example, at a gain of 15 dB, the minimum possible value of the noise figure is 3.8 dB whereas it is reduced to 2.5 dB at a gain of 10 dB, as determined using Fig. 5.10. The specific source termination that leads to the NMF_{min} for this 60-GHz nMOS LNA is also indicated in Fig. 5.10, where the single-stage noise figure is still less than 1 dB but the gain is now larger than 1. The overall noise figure is 1.6 dB for a LNA made up by an infinite chain of such amplifiers. In practice, there are possible losses associated with matching the output impedance of one amplifier to the input impedance of the next one. Coupled with the fact the LNA has only a few dB gain at the NMF_{min} point, it may be beneficial to design the amplifier have more gain with a larger noise to account for the gain loss if needed.

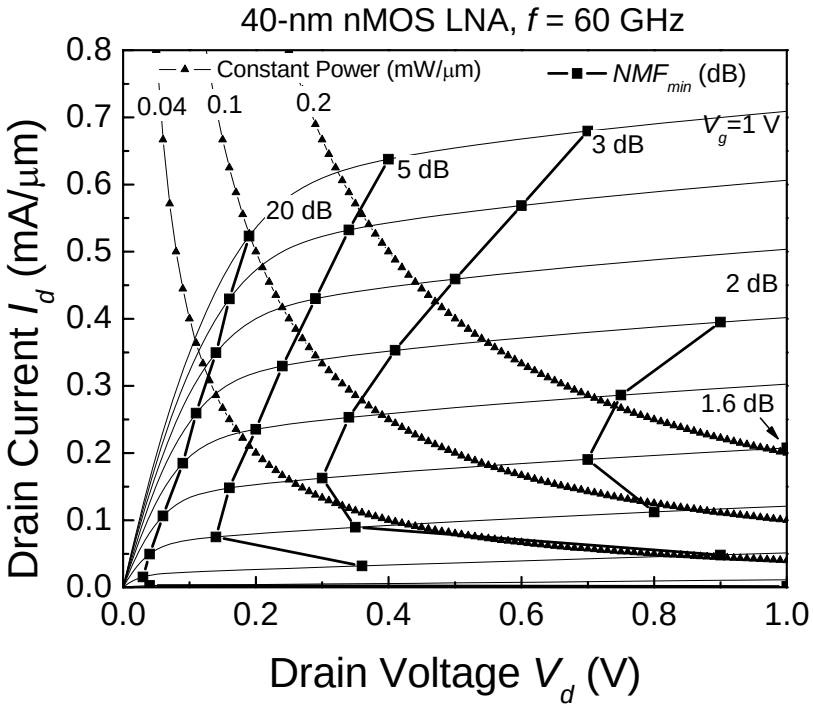

Fig. 5.11 Constant minimum noise measure figure (NMF_{min}) contours and constant power contours in I_d-V_d design space for the 40-nm nMOS LNA at 60 GHz with a power supply of 1 V.

Fig. 5.11 gives a graphical representation of power and noise trade-off in I_d-V_d space for the 40-nm nMOS LNA at 60 GHz. The solid lines with triangles are constant power curves. Those with squares are constant minimum noise measure figure (NMF_{min}) curves using the method developed in section 4.3.2. When power is constrained at a certain level, NMF_{min} varies depending on the bias choice of V_g and V_d along that constant power contour. The lowest NMF_{min} for a certain power is realized by moving the bias point towards the descending direction of constant NMF_{min} contours, until V_d becomes limited by the power supply voltage (1 V as in Fig. 5.11).

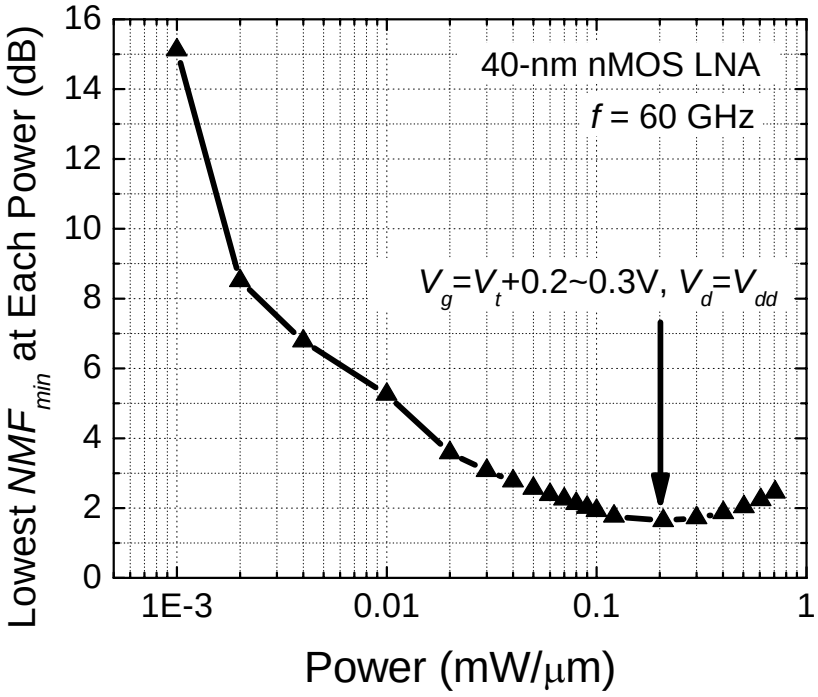

Fig. 5.12 The lowest minimum noise measure figure (NMF_{min}) at 60 GHz for a power-constrained 40-nm nMOS LNA as a function of power with a supply voltage of 1 V. The optimal bias condition for power-efficient operation is at V_g=V_t+0.2-0.3 V and V_d=V_{dd}=1 V.

Fig. 5.12 plots the lowest minimum noise measure figure (NMF_{min}) as a function of the constrained power level for the 40-nm nMOS LNA at 60 GHz. If operated at extreme low power, the LNA is very noisy. As the power is increased, the lowest NMF_{min} at each power level is reduced until it reaches an optimal value of 1.6 dB when the power is at 0.2 mW/μm. The corresponding optimal bias condition is for V_g to be 0.2-0.3 V above V_t with a drain bias of V_{dd} at 1 V. Beyond that optimal power condition the noise goes up as power is increased. Such a behavior is the combination of increasing power within a fixed supply limit as indicated by Fig. 5.11 and the fact the f_T decreases at high V_g. It is also worth noting that the power consumption is calculated per stage. In practice, more stages may be needed to account for the low gain at the NMF_{min} point, which calls for more power. In these cases, although the power from all stages needs to be taken into consideration, the biasing dependence that we found would not change at all.

5.3 GATE RESISTANCE EFFECT

5.3.1 Modeling of Gate Resistance at RF

For a MOSFET, the gate resistance consists mainly of the poly-silicon sheet resistance. However, at high frequencies, signal delay at the gate due to the distributed RC effect needs to be included for accurate CMOS RF noise modeling. This effect will become more severe as the operation frequency becomes higher. It can be modeled with different approaches for RF applications [2] and complex numerical models for the gate delay have been proposed [3]. In this work, a simple R_g approach with an effective gate resistance is used.

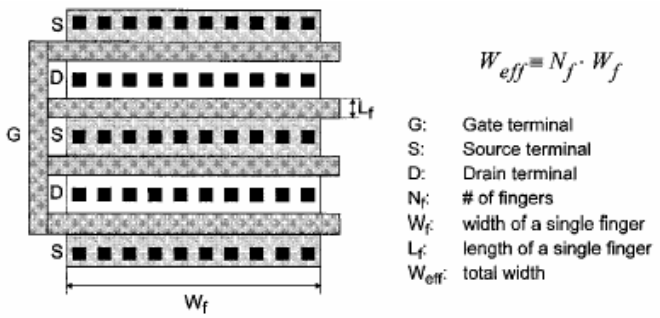

Fig. 5.13 Layout of a typical MOSFET with a multiple-finger gate structure (adapted from C. Enz [4])

RF MOS transistors are usually designed as large devices in order to achieve desired transconductances required to make them operate up to RF. As shown in Fig. 5.13, they are usually laid out as multi-finger devices, because in deep sub-micron CMOS processes, the maximum finger length (corresponding to the unit transistor width W_f) is limited. This is due to the so-called "narrow-line effect", which increases the silicided polysilicon sheet resistance as the finger width (corresponding to the transistor gate length L_f) decreases due to grain boundary problems [4]. The total effective width (W_{eff}) of a transistor is then simply N_f*W_f, where N_f is the number of fingers. When the multiple-gate finger structure is used, there are extra capacitances associated when interconnecting them. This could affect the noise and gain performance of LNA depending on the layout specifics. Many efforts are being made for layout optimization. For the CMOS layout shown in Fig. 5.13, the effective gate resistance associated with its multiple-finger structure is given by

$$R_g = k \frac{R_{g,sh}}{3} \frac{W_f}{N_f L_f} \tag{5.7}$$

where $R_{g,sh}$ is the sheet resistance of the gate material, W_f and L_f are the gate width and

length of a single finger, and N_f is the number of fingers. The factor of 1/3 accounts for the distributed nature of the RC line across the channel in the width direction. It is well known that if the gate is connected by metal lines on both sides, the gate resistance is four times smaller than what is obtained for a gate connected on one side. Factor k accounts for this, and is equal to 1 for gate contacted on one side and to 1/4 for a two-side contacted gate.

5.3.2 Effect on noise performance

From (5.6) we have

$$f_T = \frac{g_m}{2\pi \cdot (C_{gs} + C_{gd})}$$

An important observation that comes from the above equation is that the cut-off frequency is independent of the device width W, since both the transconductance and capacitances are proportional to W. In terms of noise figure, it can be shown that the noise figure for the fingered device can be expressed as [5]

$$NF^{(N)} = NF_{\min} + \frac{NG_n}{R_s}\left|R_s - \frac{Z_{opt}}{N}\right|^2$$

where the superscript N denotes the noise figure for the N-fingered device, NF_{min} is the minimum noise figure of a single finger, G_n is the noise conductance of a single finger, R_s is the driving source resistance, and Z_{opt} is the optimum noise impedance for a single finger. One of the important results here is that NF_{min} is independent of the

number of fingers and is only a function of one finger, which is given by

$$NF_{\min} = 1 + 2F_3 \frac{f}{f_T} \sqrt{\gamma g_{d0} R_g} \tag{5.8}$$

where γ is a bias-dependent parameter, g_{d0} is zero drain voltage conductance (~g_m). F_3 is a weak function of bias and less than 1. Since g_{d0} is directly proportional to finger width W whereas f_T has no width dependence, NF_{min} is a function of the product of gate resistance (R_g) with finger width (W) as indicated by (5.8).

Table 5.1 Comparison of minimum noise figure and noise measure measure figure at 60 GHz for the 40-nm nMOS LNA using a single-fingered nMOSFET with different combinations of device widths and gate resistances.

Device width	W (μm)	1	0.5		
Gate resistance	R_g (Ω)	100	50	100	200
Resistance-width product	$R_g \cdot W$ (Ω·μm)	100	25	50	100
Minimum noise figure	NF_{min} (dB)	0.44	0.39	0.41	0.44
Minimum noise measure figure	NMF_{min} (dB)	2.13	1.97	2.03	2.13

Table 5.1 lists the minimum noise figures (NF_{min}) and minimum noise measure figures (NMF_{min}) of a single-fingered 40-nm nMOS LNA at 60 GHz with different combinations of device widths and gate resistances. An important result from Table 5.1 is that, for the nMOS LNA with one single finger, the NF_{min} and NMF_{min} depend on both the device width and the gate resistance. Furthermore, whether the gate has a resistance of 100 Ω with a width of 1 μm or 200-Ω with 0.5 μm, as long as the product of gate resistance and width remains the same, the NF_{min} and NMF_{min} will keep unchanged. Therefore, the noise performance of the nMOS LNA can be characterized

as a function of the gate resistance-width product following (5.7):

$$R_g \cdot W = \frac{R_{g,sh}}{12}\frac{W_f}{N_f L_f} \cdot N_f W_f = \frac{R_{g,sh}}{12}\frac{W_f^2}{L_f}$$

where the gate is assumed to be contacted on both sides. Silicided gates, with a sheet resistance on the order of 10 Ω/□, are commonly used in current CMOS processes. The gate width for a single finger can be 2 μm or less. These typical device parameters give a gate resistance-width product of 100 Ω·μm for a MOSFET with a 40-nm gate length.

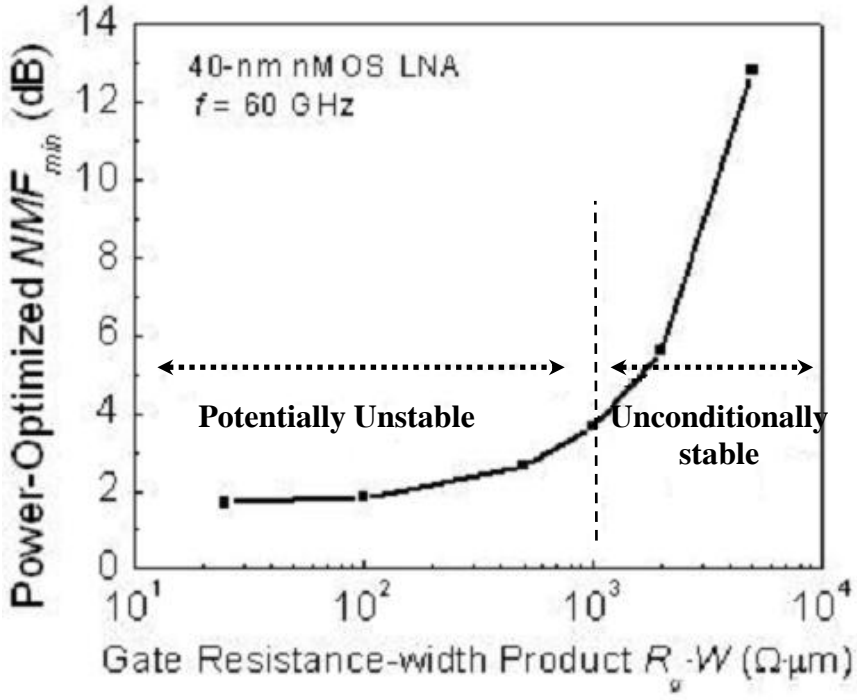

Fig. 5.14 The effect of gate resistance on the power-optimized minimum noise measure figure (NMF_{min}) for the 40-nm nMOS LNA at 60 GHz. The gate resistance is measured in Ω·μm, and its effect to the circuit stability is also indicated.

Fig. 5.14 plots the power-optimized minimum noise measure figure (NMF_{min}) as a function of gate resistance-width product ($R_g \cdot W$) for the 40-nm nMOS LNA at 60 GHz. The optimized value of NMF_{min} is obtained with the optimal bias condition indicated in Fig. 5.12. Gate resistance not only degrades amplifier's noise performance but also changes its stability as illustrated in Fig. 5.14. The design analysis would have been rather complicated if without the developed noise measure approach, which takes circuit stability into consideration. For a typical gate resistance-width product of 100 Ω·μm in the current silicide technology, the power-optimized NMF_{min} of the 40-nm nMOS LNA is 1.8 dB, which is only slightly degraded from its value of 1.6 dB with zero gate resistance. The slight degradation is also observed for noise figures shown in Fig. 5.10. For example, the minimum value of noise figure with a gain of 10 dB is increased from 2.5 to 2.8 dB, and from 3.8 to 4.0 dB with the gain of 15 dB due to the presence of the typical gate resistance.

5.4 FREQUENCY DEPENDENCE AND SUBTHRESHOLD OPERATION

The frequency dependence of the lowest minimum noise measure figure (NMF_{min}) as a function of power consumption for the power-constrained 40-nm nMOS LNA is shown in Fig. 5.15 (a) including three frequencies: 5, 30, 60, and 120 GHz. The noise performance of the nMOS LNA deteriorates once the operating frequency is increased to higher frequencies. There is a trade-off among the noise performance, power consumption and operation frequency. If the operation of the LNA is pushed to higher frequencies, the power consumption has to be increased in order to reduce the

degradation in its noise performance. On the other hand, the frequency has to be decreased if the power is reduced for low power applications without a significant degradation in its noise performance. However, for a very low frequency of 5 GHz, the increase of noise is minimal and the power reduction by operating into subthreshold is substantial. This is further indicated in Fig 5.15 (b).

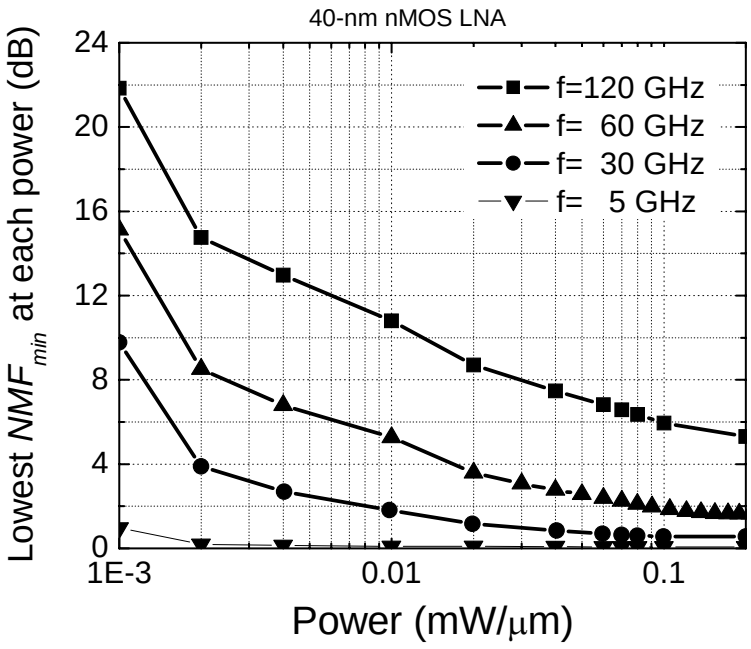

Fig. 5.15 (a) The lowest minimum noise measure figure (NMF_{min}) at different frequencies, from 5 to 120 GHz, for the power-constrained 40-nm nMOS LNA as a function of power with a supply voltage of 1 V.

Fig. 5.15 (b) shows the *NF* and G_a at the optimal source termination for NMF_{min} as a function of power at 5 GHz. The dividing line is for the power where $V_g=V_t$ and $V_d=V_{dd}$. As the power is reduced from above threshold to subthreshold, the gain goes down while the noise goes up. However, the increase of *NF* is minimal except for the last point and the gain remains to be higher than a few dB. This suggests that for low-

frequency operations, the 40-nm nMOS LNA can be operated at subthreshold condition for the ultra-low power and low noise.

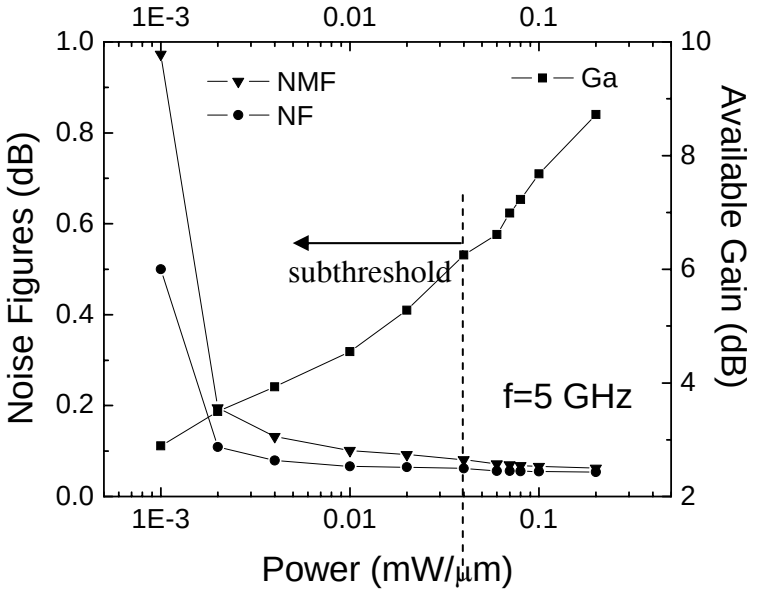

Fig. 5.15 (b) The noise figure (*NF*), minimum noise measure figure (NMF_{min}) and available gain (G_a) as a fuction of power for the 40-nm nMOS LNA at 5 GHz.

5.5 LINEARITY

With the increasing importance of CMOS technology in RF applications and the forward development of wideband wireless communication systems, linearity has become one of the most important issues in RF circuit design, which sets the upper limits of the spurious free dynamic range (SFDR) [6]. Many of these applications do require not only a low noise figure and sufficient power gain, but also a high linearity with low power consumption. In MOS RF IC, a purely sinusoidal input signal can produce a distorted output signal with higher-order harmonics due to the nonlinearity of MOS transistors. These harmonics are mainly induced by higher-order derivatives of the current-voltage (*I-V*) characteristics of the MOS transistors. As discussed in

section 4.4, an important figure of merit for linearity is the 3rd-order harmonic intercept voltage V_{IIP3}, which is the extrapolated input voltage amplitude at which the 1st- and 3rd- order output amplitudes are equal. It is given by (4.17) as

$$V_{IIP3} = \sqrt{24 \cdot \left| \frac{g_m}{g_{m3}} \right|}$$

For linearity, V_{IIP3} is used as a first-order design parameter, and should be as high as possible for low-distortion operations. It can be easily obtained from the DC characteristics and gives a good indication of the device linearity even at high frequencies [7]. Fig. 5.16 shows the transconductance (g_m) of the 40-nm nMOSFET with a threshold voltage (V_t) of 0.23 V as a function of gate voltages (V_g) with different drain voltages (V_d). For the same V_d, The g_m increases rapidly with V_g until V_g is well above V_t and saturates at high values of V_g. For the same V_g, an increase in V_d will cause an increase in g_m.

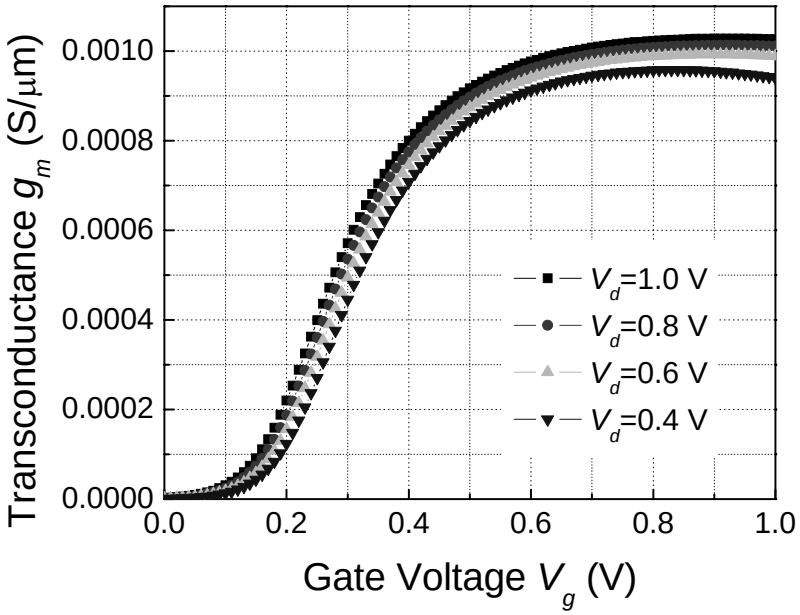

Fig. 5.16 Transconductance (g_m) of the 40-nm nMOSFET with a threshold voltage V_t= 0.23 V as a function of gate bias (V_g) under different values of drain bias (V_d).

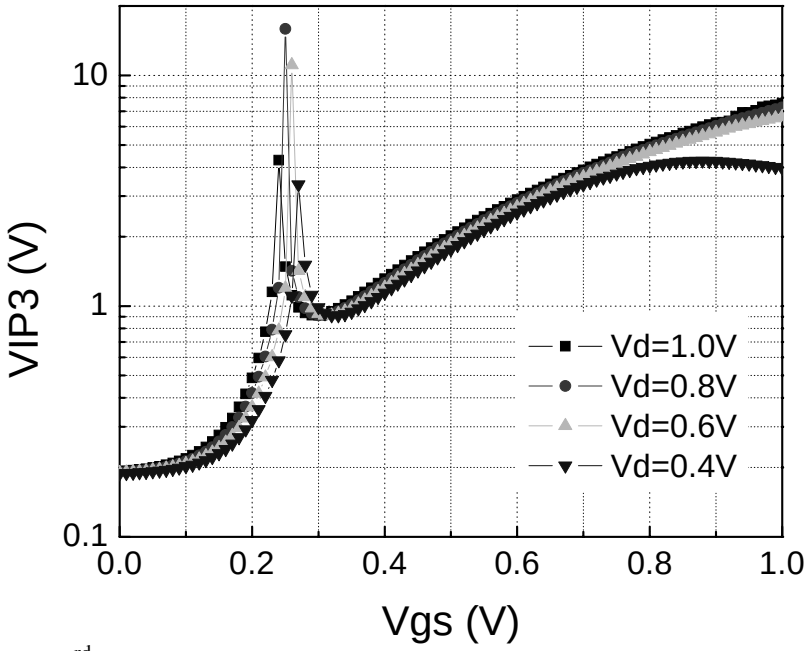

Fig. 5.17 The 3rd-order harmonic intercept voltage V_{IIP3} of the 40-nm nMOSFET (V_t=0.23 V) as a function of gate bias (V_g) under different values of drain bias (V_d).

Fig. 5.17 gives the 3rd-order harmonic intercept voltage V_{IIP3} from (4.17) as a function of gate bias (V_g) under different values of drain bias (V_d). The higher-order transconductances are obtained by numerical differentiation of g_m in Fig. 5.16. A sharp peak is observed near the threshold voltage as gate bias is varied, which reflects so-called "sweet spots" of gate biases for MOSFET linearity [8]. This is because, during the transition from subthreshold to strong inversion, the increase of g_m with V_g is at its highest, and the 2nd-order nonlinearity coefficient g_{m2} reaches its peak, while the 3rd-order nonlinearity coefficient g_{m3} becomes zero. In addition, the peak positions shift to lower V_g as V_d increases, because of the drain induced barrier lowering (DIBL) effect. Although the peak at the sweet spot is desirable for high linearity of the MOSFET, two important considerations prevent the nMOS LNA operating at such a bias. First, the V_g is very close to V_t where the LNA has significantly lower f_T, less

gain and larger noise figures from previous results. Second, the peak of V_{IIP3} is very narrow and usually requires an accurate biasing within ±10 mV of the optimum gate bias. Such a tight requirement is generally hard to achieve in reality due to the presence of process and temperature variations [9]. Therefore, it is preferred to design the nMOS LNA with a gate voltage well beyond V_t to ensure a predictably high linearity, and there is a trade-off between linearity and power, as the power is increased with either increasing V_g or V_d.

The requirement of linearity for a communication system is different upon applications. Table 5.2 compares the general requirement on the *IIP3* for both the GSM and CDMA communication standards, which are usually specified in power with a unit of dBm (0 dBm=1 mW). The power in dBm can be converted to voltage in Volt by using the following relationship:

$$P = \frac{1}{2}\frac{V_P^2}{R} \qquad dBm = 10\log_{10}\frac{P}{1mW}$$

Table 5.2 The requirement of IIP3 for the GSM and CDMA communication systems in units of power (dBm) and voltage (V). A 50-Ω system is assumed for conversion.

IIP3	GSM	CDMA
Power	- 5 dBm	5 dBm
Voltage	0.18 V	0.56 V

From section 5.2.4, it is shown that the optimal gate bias for overall noise performance is for V_g to be 0.2-0.3 V above V_t and V_d=V_{dd}. The V_{IIP3} under this optimal condition is 2 V from Fig. 5.17 with V_g=0.5 V and V_d=1 V. Considering the general requirement in Table 5.3, the 40-nm nMOS LNA remains higher linear with

its optimal bias conditions for low-noise and low-power operations. If more linearity is needed, the V_g can be increased to provide a higher V_{IIP3} at the expense of more power consumption. As for the subthreshold operation, the linearity is poor with a V_{IIP3} of 0.3 V at Vg=0.2 V. Such a low V_{IIP3} is one of the tradeoffs if subthreshold operation is involved.

5.6 CORRELATION WITH PUBLISHED HARDWARE DATA

In this section, a survey of published hardware data on CMOS LNAs is presented with the correlation to the simulation results obtained from the 40-nm nMOSFET. The purpose of this section is to identify the design trade-offs among important design parameters including operation frequency, gain and noise figure from the technology scaling point of view.

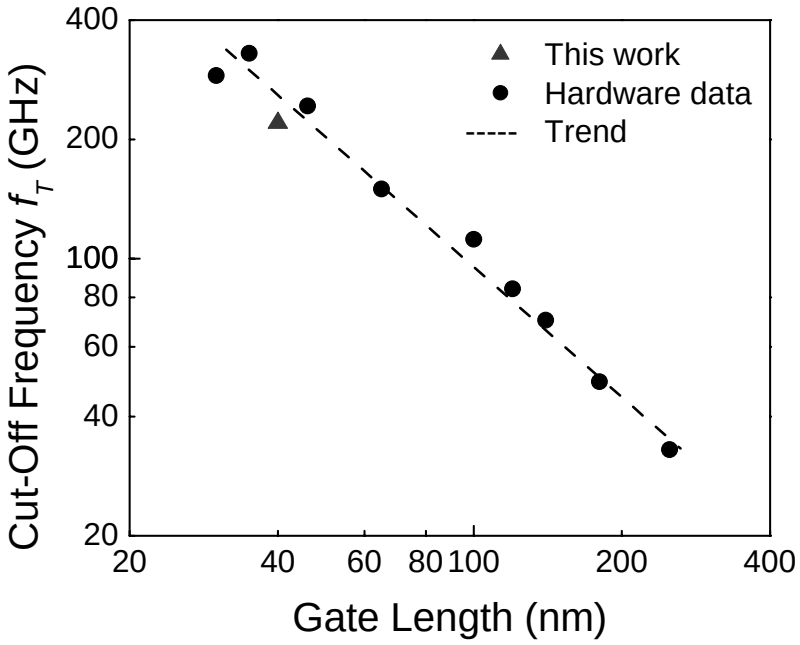

Fig. 5.18 Comparison of cut-off frequency (f_T) as a function of gate length from the simulated 40-nm nMOSFET (this work) and published hardware data [1, 10-13].

Fig. 5.18 shows the peak cut-off frequency (f_T) as a function of gate length for RF MOSFETs from both the published hardware data (denoted by solid circles) and simulation results from the 40-nm nMOSFET in this work (denoted by the solid triangle). At the time when this work started, the highest reported f_T for RF CMOS is below 150 GHz, which makes a f_T of 200 GHz a good starting point for our device design at that time. However, as the CMOS technology moves forward quickly in recent years, the f_T is increased rapidly to 243 GHz as a result of the many efforts in the CMOS scaling [12]. The record RF CMOS performance up-to-date is achieved by the SOI CMOS technology with a peak f_T of 330 GHz from IBM, which features a 65-nm technology with a gate length of 35 nm and a gate oxide of 1.05 nm [1]. The simulation result shown in Fig. 5.15 is for the 40-nm nMOSFET with a peak f_T of 220 GHz, and it agrees well with the performance trends from the published hardware data.

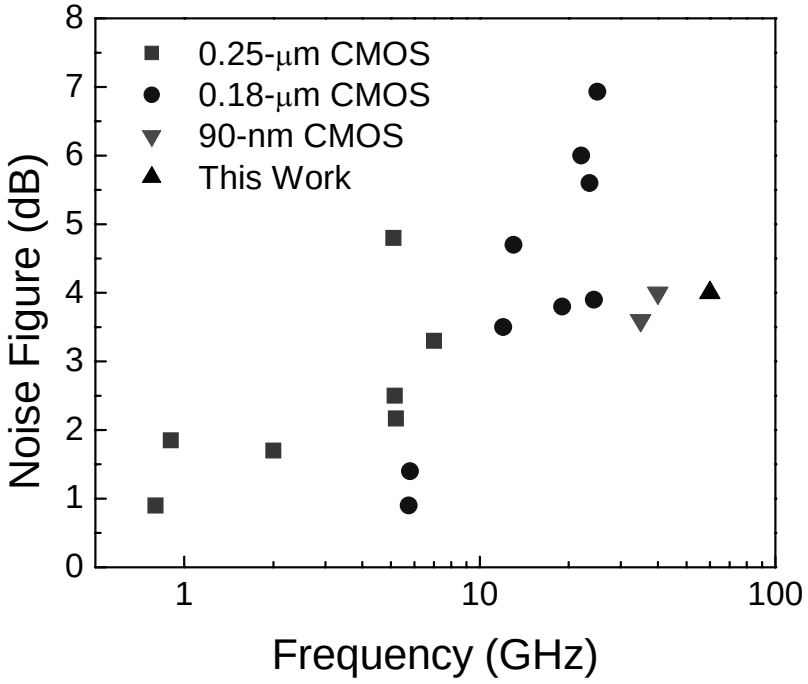

Fig. 5.19 Noise figure as a function of frequency for various RF nMOS LNAs from the simulated 40-nm device in this work and published hardware data [14-28].

Fig. 5.19 shows the noise figure as a function of frequency for RF nMOS LNAs from both published hardware data and the simulation result from the 40-nm nMOSFET. The noise figures from the published data are taken as the ones measured by the actual hardware from various nodes of CMOS technologies. For frequencies below 5 GHz, 0.25-μm CMOS technology is capable of providing an amplifier with a very low noise figure less than 2 dB. However, as the operation frequency is increased from 5 to 10 GHz, the noise figure increased substantially and a technology with a shorter length, 0.18-μm in this case, is needed in order to bring the noise figure down below 2 dB. When the frequency is further increased to around 20 GHz, the noise figures soar inevitably to well above 4 dB. Although there are some excellent results that bring the noise figure down below 4 dB, the low noise operation at even higher frequencies is generally difficult without using the advanced sub-100-nm technology. Recently, noise figures of 3.6 dB at 35 GHz and 4 dB at 40 GHz are reported using a 90-nm SOI CMOS technology from IBM [28], which are the best results for a Si-based mm-wave LNA reported up to date. As indicated by the comparison of these results, the general trends for the LNA noise performance is to increase the operation frequency while keeping the noise figure at a relative low level, and such a trend is made possible by the steady downscaling and improvement of CMOS RF technologies. Also shown in Fig. 5.19 is the simulation result from this work, indicated as the noise figure with a gain of 15 dB and typical gate resistances under the optimal bias conditions. The 40-nm nMOS LNA has a noise figure of 4 dB at 60 GHz, as compared to a noise figure of 4 dB at 40 GHz with a gain of 9.5 dB for the 90-nm

technology node.

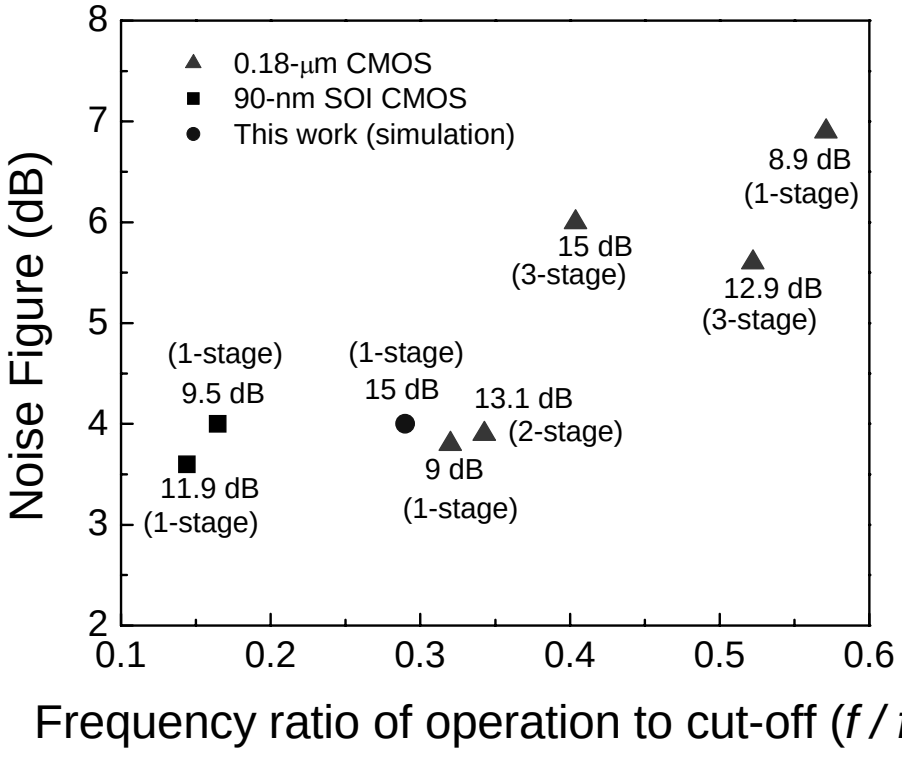

Fig. 5.20 Noise figure as a function of the ratio of operation frequency to cut-off frequency (f/f_T) of CMOS LNAs from published hardware results and simulated 40-nm device in this work. The numbers adjacent to symbols are the corresponding gain at the same frequency [24-28].

Table 5.3 Performance comparisons of CMOS LNAs for frequencies above 10 GHz from published hardware data and the simulated 40-nm device in this work.

Ref.	L_{gate} (nm)	f_T (GHz)	Freq. (GHz)	Gain (dB)	NF (dB)	Circuit Topology*
[24]	180	45	23.5	12.9	5.6	3 stages, CS+CS+CS
[24]	180	45	25.7	8.9	6.9	3 stages, CS+CS+CS
[25]	180	50	16	9	3.8	1 stage, CS
[26]	180	54	21.8	15	6	3 stages, CGRF+CS+CS
[27]	180	70	24	13.1	3.9	2 stages, CS+CS
[28]	90	243	35	11.9	3.6	1 stage, cascode
[28]	90	243	40	9.5	4	1 stage, cascode
This work	40	220	60	15	4	1 stage, CS

*CS: common source, CGRF: common gate resistive feedthrough

Fig. 5.20 compares the noise figure and gain as a function of the ratio of operation frequency to cut-off frequency (f/f_T) from reported CMOS LNAs and simulated 40-nm device in this work, and Table 5.3 listed the relevant data from the reported hardware results including the circuit topology. The cut-off frequency (f_T) is determined primarily by the technology node, and the device is usually operated at a frequency much lower than f_T as the gain drops with a slope of -10 dB/dec as the operation moves towards higher frequencies, and the noise becomes worse as well. For the 0.18-μm CMOS technology node, f_T is limited to be less than 100 GHz and the ratio of f/f_T will be greater than 0.3 in order to meet the operation at frequencies above 20 GHz using such technology. The low gain at high ratio of f/f_T is usually compensated by using multistage topology for required gain performance, however, this further aggregates the degradation of the noise figure at higher frequencies, due to the presence of the additional stages as evident by Fig. 5.20. One way to reduce the ratio of f/f_T without sacrificing the operation frequency is to increase f_T by technological scaling. As shown in Fig 5.20, the 90-nm technology node provides noise figures less than 4 dB with gains around 10 dB at frequencies up to 40 GHz, all within small f/f_T ratios of less than 0.2 using single stage topology. This is made possible by the high f_T of 243 GHz associated with this 90-nm SOI CMOS technology [28]. However, for mm-wave applications, the LNA made from this technology would be difficult to achieve a low noise figure and a simultaneously high gain at 60 GHz.

The simulation result from this work is also shown in Fig. 5.20. The noise figure is for the 40-nm nMOS LNA with a gain of 15 dB and the typical gate

resistance under optimal bias condition. Although the ratio of f/f_T is higher than that of the reported 90-nm technology node, it achieves the comparable noise figure with a larger gain at a higher frequency, all using a single-stage CS amplifier. Such an improvement, as compared with the 90-nm technology, is partly due to the shorter channel length of the 40-nm device, and partly due to the optimal biasing design in the I_d-V_d space as illustrated in Fig. 5.11.

The text of this chapter, in part, is a reprint of the material that has been submitted for publication: Cai, Ming; Liu, Minjian; Taur, Yuan. “A design study of nanometer-gate low-noise amplifier near the limit of CMOS scaling”, Solid-State Electronics; Cai, Ming; Liu, Minjian; Taur, Yuan. “Design study of power-efficient 60-GHz MOS low noise amplifiers near scaling limits”, Electronics Letters. The dissertation author was the primary author and the co-authors listed in these publications directed and supervised the research which forms the basis for this chapter.

REFERENCES

1. S. Lee et al., "Record RF performance of sub-46 nm Lgate NFETs in Microprocessor SOI CMOS technologies", *IEDM Tech. Dig.*, pp. 251-254, Dec. 2005.

2. Y. Cheng, "Advanced MOSFET modeling for RF IC design", *NSTI-nanotech WCM*, pp. 253-258, May 2005.

3. E. Abou-Allam and T. Manku, "A small-signal MOSFET model for radio frequency IC applications", *IEEE Trans. Computer-aided Design integr. Circuits Syst.*, vol. 16, pp. 437-447, 1997.

4. C. Enz, "An MOS transistor model for RF IC design valid in all regions of operation", *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 5, pp. 342-359, 2002.

5. T. Manku, "Microwave CMOS—device physics and design", *IEEE J. Solid State Circuits*, vol. 34, no. 3, pp. 277-285, 1999.

6. T. H. Lee, *The design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed., Cambridge University Press, 2004.

7. L. F. Tiemeijer *et al.*, "RF distortion characterization of sub-micron CMOS", *Proc. ESSDERC*, pp. 464-467, 2000.

8. G. Niu, J. Pan, X. Wei, S. S. Taylor, and D. Sheridan, "Intermodulation linearity characteristics of CMOS transistors in a 0.13 um Process", *IEEE RFIC Symp. Dig.*, pp. 65-68, 2005.

9. V. Aparin, G. Brown, and L. E. Larson, "Linearization of CMOS LNA's via optimum gate biasing", *ISCAS*, pp. 748-751, 2004.

10. V. Dimitrov et al., "High performance, sub-50 nm MOSFETs for mixed signal applications", *IEDM Tech. Dig.*, pp. 213-216, Dec. 2005.

11. L. F. Tiemeijer et al., "Record RF performance of standard 90 nm CMOS technology", *IEDM Tech. Dig.*, pp. 441-444, Dec. 2004.

12. N. Zamdmer et al., "A 243-GHz Ft and 208-GHz Fmax, 90-nm SOI CMOS SoC Technology with Low-Power Millimeter-Wave Digital and RF Circuit Capability", *Symp. VLSI Tech. Dig.*, pp. 98-99, 2004.

13. E. Morifuji et al., "Future perspective and scaling down roadmap for RF CMOS", *Symp. VLSI Tech. Dig.*, pp. 163-164, 1999.

14. J.-S. Goo, H.-T. Ahn, D. J. Ladwig, Z. Yu, T. H. Lee and R. W. Dutton, "A noise optimization technique for integrated low-noise amplifiers", *IEEE J. Solid-State Circuits*, vol. 37, pp. 994-1002, 2002.

15. Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, "The impact of scaling down to deep submicron on CMOS RF circuits", *IEEE J. Solid-State Circuits*, vol. 33, pp. 1023-1036, 1998.

16. B. A. Floyd and D. Qzis, "Low-noise amplifier comparison at 2 GHz in 0.25-um and 0.18-μm RF-CMOS and SiGe BiCMOS", *IEEE RFIC Symp. Dig.*, pp. 185-188, 2004.

17. H. Samavati, H. R. Rategh, and T. H. Lee, "A 5-GHz CMOS wireless LAN receiver front end", *IEEE J. Solid-State Circuits*, vol. 35, pp. 765-772, 2000.

18. T.-P. Liu and E. Westerwick, "5-GHz CMOS radio transceiver front-end chipset", *IEEE J. Solid-State Circuits*, vol. 35, pp. 1927-1933, 2000.

19. H.-W. Chiu, S.-S. Lu, and Y.-S. Lin, "A 2.17-dB NF 5-GHz-band monolithic CMOS LNA with 10-mW DC power consumption", *IEEE Trans. Microw. Theory. Tech.*, vol. 53, pp. 813-824, 2005.

20. R. Fujimoto, K. Kojima, and S. Otaka, "A 7-GHz 1.8-dB NF CMOS low-noise amplifier", *IEEE J. Solid-State Circuits*, vol. 37, no. 7, pp. 852-856, 2002.

21. D. J. Cassan and J. R. Long, "A 1-V transformer-feedback low-noise amplifier for 5-GHz wireless LAN in 0.18-μm CMOS", *IEEE J. Solid-State Circuits*, vol. 38, no.

3, pp. 427-435, 2003.

22. B. Mohammadi and C. A. T. Salama, “A 5.8 Ghz CMOS LNA for WLAN applications”, *IEEE RFIC Symp. Dig.*, pp. 113-116, 2004.

23. J. Gil, K. Han, and H. Shin, “13 GHz 4.67 dB NF CMOS low-noise amplifier”, *Electron. Lett.*, Vol. 39, no. 14, pp. 1056-1058, 2003.

24. K.-W. Yu, Y.-L. Lu, D.-C. Chang, V. Liang and M. F. Chang, “K-band low-noise amplifiers using 0.18 μm CMOS technology”, *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 3, pp. 106-108, 2004.

25. H. Yano, Y. Nakahara, T. Hirayama, N. Matsuno, Y. Suzuki, and A. Furukawa, “Performance of Ku-band on-chip matched Si monolithic amplifiers using 0.18-um-gatelength MOSFETs”, *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 6, pp. 1086-1093, 2005.

26. X. Guan and A. Hajimiri, “A 24-GHz CMOS front-end”, *IEEE J. Solid-State Circuits*, vol. 39, no. 2, pp. 369-373, 2004.

27. S.-C. Shin, M.-D. Tsai, R.-C. Liu, K.-Y. Lin, and H. Wang, “A 24-GHz 3.9-dB NF low-noise amplifier using 01.8 μm CMOS technology”, *IEEE Microw. Wireless Compon. Lett.*, vol. 15, no. 7, pp. 448-450, 2005.

28. F. Ellinger, “26-42 GHz SOI CMOS low noise amplifier”, *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 522-528, 2004.

CHAPTER 6

CMOS LNA DESIGN NEAR THE SCALING LIMIT

6.1 CMOS DEVICE DESIGN NEAR THE SCALING LIMIT

The performance of CMOS devices were greatly improved by scaling in recent years. The current CMOS technology in production is at the 90-nm node with an effective channel length less than 50 nm [1], and the cutting-edge 65-nm node has a channel length below 40 nm [2]. As the channel length is rapidly reduced by forward development of aggressive scaling, it is generally believed that the bulk scaling limit for CMOS will be reached at the time when the channel length is reduced to 20 nm. Once the channel length is further reduced to below 20 nm, double-gate (DG) MOSFET (a generic form of FinFET) technology becomes the solution, since it is more immune to the short-channel effect than the standard bulk MOSFET [3]. In order to investigate the performance trends of CMOS LNAs under such a scaling roadmap, three more devices were investigated in addition to the 40-nm bulk MOSFET discussed in Chapter 5: a 20-nm bulk MOSFET, a 20-nm DG MOSFET, and a 10-nm DG MOSFET. As a summary, Fig. 6.1 compares the structure and relative geometry of these devices using TECPLOT-ISE from SYNOPSYS-TCAD. The concepts and details involved in the design of these scaled MOSFET devices have been covered in Chapter 3, and this chapter presents the performance trends of both cut-off frequency and noise figures for these CMOS LNAs near the scaling limits.

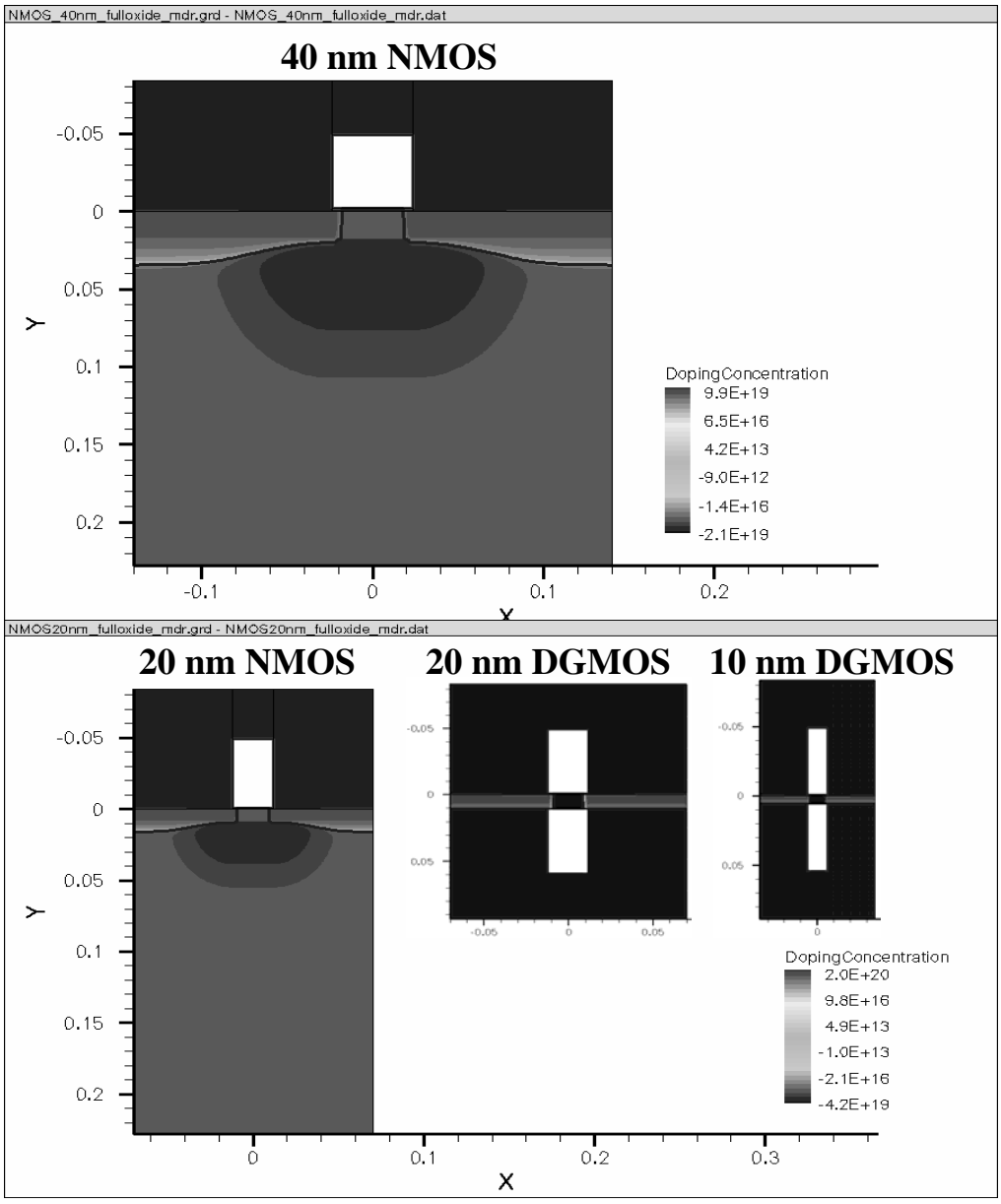

Fig. 6.1 Cross-section comparison of the MOSFETs near the CMOS scaling limits used in this study: (a) 40-nm bulk MOSFET; (b) 20-nm bulk MOSFET; (c) 20-nm double-gate (DG) MOSFET; (d) 10-nm DG MOSFET.

6.2 PERFORMANCE TRENDS OF CMOS LNAS

The design of various scaled nMOS LNAs are carried out in much the same way as for the 40-nm bulk nMOS LNA, which is studied exclusively in Chapter 5. As a result, Fig. 6.2 compares the peak cut-off frequency (f_T) as a function of channel length for all scaled MOSFETs in this work and the published hardware data, and Table 6.1 summarizes the key design parameters obtained from various nMOS LNAs used in this study. In terms of device design, since the threshold voltage (V_t) of the DG MOSFETs can be easily modified by the work functions of the gate material instead of several interacting parameters as in (3.5) for the bulk MOSFETs, the work functions of the DG MOSFETs are designed such that the 20-nm DG MOSFET has the same V_t as the 20-nm nMOSFET for a fair comparison. For 40- and 20- nm bulk nMOS LNAs, the RF parameters are extracted at corresponding optimal bias conditions for power-efficient noise operations, which is V_g=0.5 V and V_d=1 V for the 40-nm nMOS LNA as shown in Fig. 5.12, and it is V_g=0.45 V and V_d=1 V for the 20-nm nMOS LNA. For DG nMOS LNAs, the bias conditions are kept the same as 20-nm bulk nMOS LNAs, i.e., V_g=0.45 V and V_d=1 V. The minimum noise measure figures are shown as the lowest values under each optimal bias for power with the same typical gate resistance-width product, 100 Ω·μm, for all LNAs at 60 GHz.

Table 6.1 Summary of key design parameters for various nMOS LNAs near the CMOS scaling limits used in this study.

MOSFET Type	**Bulk**		**Double-Gate**	
Effective channel length (nm)	40	20	20	10
Gate oxide thickness (nm)	2	1	1	0.5
Threshold voltage (V)	0.23	0.20	0.20	0.16
Cut-off frequency (GHz)	210	370	430	770
Power-optimized NMF_{min} (dB)	1.8	0.8	0.8	0.6
3rd-order Harmonic V_{IIP3} (V)	2.0	1.8	1.8	2.5

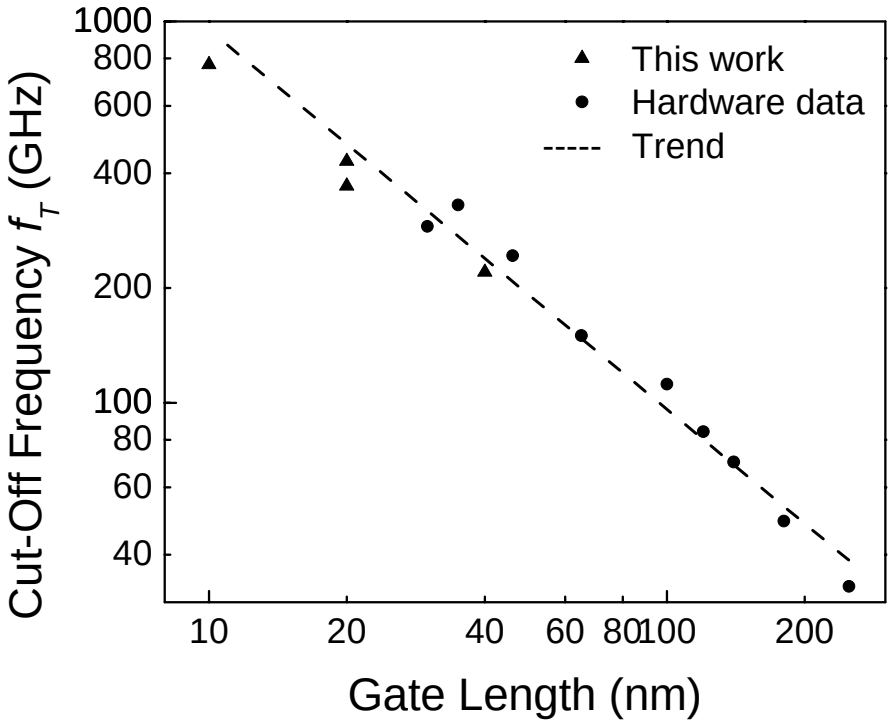

Fig. 6.2 Comparison and the general trend of cut-off frequency (f_T) as a function of the gate length for various RF MOSFETs from simulation (this work) and published hardware data [4-8].

Fig. 6.2 shows the peak cut-off frequency (f_T) as a function of the gate length for various RF MOSFETs from both the published hardware data (denoted by solid circles) and simulation in this work (denoted by solid triangles). The scaling trend of f_T from the hardware data is also shown with projection into the scaling limit at 10 nm. As shown in Fig. 6.2, the simulation result for the 40-nm nMOSFET agrees well with the published hardware data, and those for the various scaled MOSFETs studied in this work fall on the projected scaling trend. As the lithography technology is quickly moving forward from 90-nm to 65-, 45-, 32-, and 22-nm nodes, the f_T of the CMOS transistors will be increased above 1 THz as suggested by the scaling trend in Fig. 6.2. Such a high RF performance makes these transistors well sufficient for the applications at mm-wave frequency range and further beyond.

As shown in Table 6.1, when the bulk MOSFET is scaled from a channel length of 40 nm to 20 nm with a concurrent reduction of oxide thickness from 2 nm to 1 nm, the cut-off frequency (f_T) is increased from 210 GHz to 370 GHz whereas the power-optimized noise measure figure is reduced from 1.8 dB to 0.8 dB. There is no substantial performance improvement of the 20-nm DG nMOS LNA over the 20-nm bulk nMOS LNA. This is because while the DG MOSFET has higher transconductance, its gate capacitance is also higher, resulting in nearly the same AC performance. However, when the CMOS LNA is scaled from a channel length of 20 nm to 10 nm with more scalable DG device structure, it outperformed the limiting 20 nm bulk LNA by a factor of two in terms of cut-off frequency (f_T), and substantial

reduction in its noise figure at 60 GHz. No severe degradation in linearity has been found in all these scaled devices. Thus the power-efficient low noise amplification can be easily extended to 60 GHz and beyond with high linearity as the CMOS transistors are scaled down to near the gate length limit. Such a high cut-off frequency with low noise figures shown in Table 6.1 suggests that CMOS can provide comparable RF performance to III-V technologies at mm-wave frequency range with additional benefits of lower cost and higher level of integration.

The text of this chapter, in part, is a reprint of the material that has been submitted for publication: Cai, Ming; Liu, Minjian; Taur, Yuan. “A design study of nanometer-gate low-noise amplifier near the limit of CMOS scaling”, Solid-State Electronics. The dissertation author was the primary author and the co-authors listed in these publications directed and supervised the research which forms the basis for this chapter.

REFERENCES

1. M. Khare et al., “A high performance 90nm SOI technology with 0.992 μm2 6T-SRAM cell”, *IEDM Tech. Dig.*, pp. 407-410, Dec. 2002.

2. M. Horstmann, et al., “Integration and optimization of embedded-SiGe, compressive and tensile stressed liner films, and stress memorization in advanced SOI CMOS technologies”, *IEDM Tech. Dig.*, pp. 243-246, 2005.

3. Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*, Cambridge University Press, New York, 1998.

4. V. Dimitrov et al., “High performance, sub-50 nm MOSFETs for mixed signal applications”, *IEDM Tech. Dig.*, pp. 213-216, Dec. 2005.

5. E. Morifuji et al., “Future perspective and scaling down roadmap for RF CMOS”, *Symp. VLSI Tech. Dig.*, pp. 163-164, 1999.

6. L. F. Tiemeijer et al., “Record RF performance of standard 90 nm CMOS technology”, *IEDM Tech. Dig.*, pp. 441-444, Dec. 2004.

7. N. Zamdmer et al., “A 243-GHz Ft and 208-GHz Fmax, 90-nm SOI CMOS SoC Technology with Low-Power Millimeter-Wave Digital and RF Circuit Capability”, *Symp. VLSI Tech. Dig.*, pp. 98-99, 2004.

8. S. Lee et al., “Record RF performance of sub-46 nm Lgate NFETs in Microprocessor SOI CMOS technologies”, *IEDM Tech. Dig.*, pp. 251-254, Dec. 2005.

CHAPTER 7

SUMMARY OF DISSERTATION AND FUTURE WORK

SUMMARY

The main focus of this dissertation is the design study of power-efficient CMOS LNAs at 60 GHz based on state-of-the-art nMOSFETs and the projection of performance trends of CMOS LNAs near the scaling limits.

Starting with an introduction of millimeter-wave (mm-wave) communication systems, a survey of current status of research and development is conducted on various technologies that enable the mm-wave applications, with the emphasis on the Si-based CMOS technologies. It also serves as the basis for various design parameters and provides checkpoints for the simulation results in this work. The scaling effect of CMOS devices near its limit is studied by using a 40-nm bulk nMOSFET, 20-nm bulk and double-gate (DG) nMOSFET, and a 10-nm DG nMOSFET with a 2-D device simulation tool - SYNOPSYS TCAD. Special considerations are given to control the short-channel effect while scaling the MOSFET to its minimum size.

In order to facilitate the RF small-signal design using two-port parameters, a complete solution, including important relationships between circuit stability, power gains, and noise parameters are provided in a single work frame. It features a new methodology developed to use the noise measure as a proper figure of merit for two-port overall noise performance with arbitrary circuit stability.

A detailed power-efficient LNA design at 60 GHz has been completed based

on state-of-the-art nMOSFETs near the scaling limits to facilitate the wireless applications at millimeter-wave frequency range. The optimal gate bias for the power-constrained LNA is found to be 0.2-0.3 V above the threshold voltage. The 40-nm nMOS LNA has a minimum noise measure figure of 1.8 dB at 60 GHz with a power consumption of 0.2 mW/μm. The degradation in the overall noise performance due to gate resistance in current silicide technology is shown to be insignificant as long as the device width per gate finger is constrained. The performance trend of CMOS LNAs near the scaling limits has been investigated including the 20-nm bulk and DG MOSFETs, and 10-nm DG MOSFETs. Further improvement of the LNA noise performance can be achieved with higher cut-off frequency and lower noise figures in such emerging devices, and CMOS LNAs near the scaling limit are found to compete directly against III-V technologies at mm-wave frequency range with a lower cost and higher level of integration.

In terms of suggestion for future work, it would be appropriate to update the CMOS device to include the recent technology advances including halo doping and source-drain extension. A calibrated model with drift-diffusion characteristic of material properties is also useful for more accurate modeling. Also the inclusion of many parasitic parameters encountered in the hardware manufacturing, including source/drain resistance and body to substrate resistance, would also help to improve the predictability of this work.