RF BUILDING BLOCK MODELING

OPTIMIZATION AND SYNTHESIS

Wei Cheng

The graduation committee consists of:

Chairman and Secretary:

Prof. dr. ir. A. J. Mouthaan	Universiteit Twente

Promotor:

Prof. dr. ir. B. Nauta	Universiteit Twente

Assistant Promotor:

Dr. ir. A. J. Annema	Universiteit Twente

Referee:

Dr. J. A. Croon	NXP Semiconductors

Members:

Prof. dr. ir. F. E. van Vliet	Universiteit Twente
Prof. dr. ir. C. H. Slump	Universiteit Twente
Prof. dr. ir. D. M. W. Leenaerts	Technische Universiteit Eindhoven

This research was funded by NXP semiconductors, the Netherlands and was performed in the Integrated Circuit Design (ICD) group, Center for Telematics and Information Technology (CTIT), University of Twente.

CTIT Ph.D. Thesis Series No. 12-220
Center for Telematics and Information Technology
P.O. Box 217, 7500 AE
Enschede, The Netherlands

Title: RF building block modeling: optimization and synthesis
Author: Wei Cheng
ISSN: 1381-3617 (CTIT Ph.D. thesis series No. 12-220)
ISBN: 978-90-365-3353-9
DOI: 10.3990./1.9789036533539

RF BUILDING BLOCK MODELING
OPTIMIZATION AND SYNTHESIS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
Prof. dr. H. Brinksma,
on account of the decision of the graduation committee
to be publicly defended
on Thursday 10th May 2012 at 12:45

by

Wei Cheng

born on 5th September 1980
in Wuhan, Hubei, China

This dissertation has been approved by

Promotor: Prof. dr. ir. Bram Nauta

Assistant promotor: Dr. ir. Anne Johan Annema

To my parents and Julie

Table of Contents

Summary

For circuit designers it is desirable to have relatively simple RF circuit models that do give decent estimation accuracy and provide sufficient understanding of circuits. Chapter 2 in this thesis shows a general weak nonlinearity model that meets these demands. Using a method that is related to harmonic balance, this model yields closed-form expressions that are a linear combination of technology dependent transistor nonlinearity parameters and topology-dependent AC transfer functions only. Using this, time-invariant weakly nonlinear analyses can be accomplished using time-invariant linear analyses.

This general distortion model is used in this thesis to derive design insights and novel methods to cancel distortion in attenuator and in cascoded LNAs. Chapter 3 presents a proof-of-concept resistive feedback LNA fabricated in a standard 0.16µm CMOS process, for 0.1GHz to 1GHz, showing improvements of 6.3dB to 10dB for IIP3 and 0.2dB to 1dB for gain without noise degradation and with a very modest power and area penalty. Chapter 4 shows distortion cancellation for CMOS attenuators: in the demonstration Π-attenuator system with 4 discrete attenuation settings, for DC-5GHz, >3dBm input P_{1dB} and >26dBm IIP3 are achieved in measurements, while the active area is 0.0054mm^2. In the demonstration T-attenuator system with 4 discrete attenuation settings, for DC-5.6GHz, >13dBm input P_{1dB} and >27dBm IIP3 are achieved in measurements, while the active area is 0.0067mm^2. Both

simulation and measurement results demonstrate good robustness against PVT variations.

Chapter 5 shows a full analysis of distortion and noise for Gilbert mixers. Using the introduced analysis method, the noise and distortion of this time-varying system are estimated by a limited number of time-invariant AC calculation. The analyses show that the decreasing transistor output resistance together with the low supply voltage in deep submicron technologies contributes significantly to flicker-noise leakage. The analyses also show that the slope of the LO signal has significant effect on IIP2 while little effect on IIP3. Design insights for low flicker noise are then presented.

Using the design insights in chapter 5, chapter 6 presents techniques to simultaneously cancel flicker noise and IM3 in Gilbert-type mixers. Two proof-of-concept double-balanced mixers in 0.16µm CMOS were fabricated using these techniques. One chip is designed for full-IM3/partial-flicker-noise cancellation at 0.9GHz, which achieves 9dB flicker noise suppression, improvements of 10dB for IIP3, 5dB for conversion gain, and 1dB for input P_{1dB} while the thermal noise increased by 0.1dB at the cost of a small power and area penalty. The other chip is designed for full-flicker-noise/partial-IM3 cancellation at 0.9GHz with a low supply voltage ($0.67 \times V_{DD}$), which shows >10dB flicker noise suppression within ±200% variation of the negative impedance bias current. The overall conclusion and suggestion for future work are summarized in chapter 7.

Samenvatting

Het is nuttig voor circuitontwerpers om relatief eenvoudige RF-circuitmodellen te hebben die voldoende mate van nauwkeurigheid hebben en die inzicht geven in de werking van het circuit. Hoofdstuk 2 in dit proefschrift beschrijft een model voor zwak-niet-lineaire circuits dat voldoet aan deze eisen. Gebruikmakend van een harmonic-balance-gerelateerde methode geeft dit model gesloten uitdrukkingen die slechts een lineaire combinatie zijn van technologieafhankelijke niet-lineaire transistoreigenschappen en van topologieafhankelijke kleinsignaaloverdrachten. Hiermee kunnen tijdsafhankelijke zwak-niet-lineaire analyses gedaan worden met behulp van tijdsinvariante lineaire analyseresultaten.

Dit model wordt verder in dit proefschrift gebruikt om inzichten en nieuwe distorsiecompensatiemethoden voor verzwakkers en voor gecascodeerde LNA's te verkrijgen. Hoofdstuk 3 toont als proof-of-concept een resistief teruggekoppelde LNA in een standaard 0,16µm CMOS proces, werkend tussen 100MHz en 1GHz, met tussen 6,3dB en 10dB verbetering in IIP3 met een gelijktijdige versterkingstoenamen tussen 0,2dB en 1dB zonder toename van ruis en tegen een zeer geringe toename van het vermogensgebruik en van de afmetingen op de chip. Hoofdstuk 4 toont vervormingscancellatie voor CMOS verzwakkers: metingen aan het demonstratieontwerp met Π-verzwakkers en 4 discrete verzwakkingsstanden laat een ingangsgerefereerde P_{1dB}>3dBm zien met IIP3 groter dan 26dBm voor frequenties van DC tot 5Ghz met een chipoppervlak van 0,0054mm^2. Het demonstratieontwerp met

T-verzwakkers en 4 discrete verzwakkingsstanden bereikt in metingen in het frequentiegebied van DC tot 5,6GHz een ingangsgerefereerde P_{1dB}>13dBm met IIP3>27dBm met een chipoppervlak van 0,0067mm^2. Zowel simulaties als metingen laten zien dat deze systemen voldoende robuust zijn voor PVT-variaties.

Hoofdstuk 5 presenteert een volledige ruis- en vervormingsanalyse van Gilbert-mixers. Gebruikmakend van de methode uit hoofdstuk 1 wordt zowel ruis als vervorming bepaald met behulp van een klein aantal tijdsinvariante kleinsignaalberekeningen. Verdere analyse hiervan laat zien dat de afnemende uitgangsweerstand van transistoren, samen met lage voedingsspanningen, in diep-sub-micron technologieën sterk bijdragen aan 1/f-ruislek. Tevens wordt getoond dat de steilheid van de flanken van het LO-signaal een groot effect heeft op IIP2 terwijl het slecht een klein effect heeft op IIP3. Hieruit volgen meerdere ontwerpinzichten om lage 1/f-ruis te krijgen.

Gebruikmakend van de inzichten van hoofdstuk 5 toont hoofdstuk 6 technieken om gelijktijdig 1/f-ruis en IM3 te elimineren in Gilbert-achtige mixers. Twee proof-of-concept dubbelgebalanceerde mixers in 0,16µm CMOS zijn gemaakt met deze technieken. Eén van deze chips is ontworpen voor volledige IM3-eliminatie met gedeeltelijke 1/f-eliminatie bij 0,9 GHz; deze chip bereikt 9dB onderdrukking van 1/f-ruis, 10dB verbetering in IM3, 5dB verbetering van de conversie-gain en 1dB P_{1dB} verbetering terwijl de thermische ruis 0,1dB toeneemt. De prijs die hiervoor betaald wordt is slecht een klein beetje vermogensconsumptie en een kleine toename in het chipoppervlak. De andere schakeling is ontworpen voor volledige onderdrukking van 1/f-ruis met een gelijktijdige gedeeltelijke onderdrukking van de IM3 bij 0,9Ghz, met een lage voedingsspanning ($0,67xV_{DD}$); deze schakeling geeft >10dB 1/f-ruisonderdrukking binnen 200% variatie van de instelstroom van de negatieve impedantie.

Hoofdstuk 7 geeft een samenvatting van de belangrijkste conclusies en geeft aanbevelingen voor verder onderzoek.

Chapter 1

Introduction

Nowadays integrated circuits are hidden everywhere in our life. They are making the internet, PCs, mobile phones and the world, go round. Designing the integrated circuits requires lots of manpower. Fortunately, the design of digital integrated circuits can be highly automated. For the circuit function description defined by designers, the computer can generate the according layout mask set that is sent to the foundry for manufacturing this integrated circuit. However, for RF/analog integrated circuits, this design process is mainly performed by designers manually, which motivates the research of RF/analog design automation.

1.1 **P-cell for RF circuit block**

The design automation of the transistor is very well developed so far. The designer only needs to input very few parameters of the transistor such as width and length rather than delving deeply into the physics. The transistor P-cell (parameterized-cell) takes care of the rest job such as linking to process library model, giving circuit simulation results and making mask layout. The

Section 1.1 was part of the paper published in IEEE International Symposium on Circuits and Systems (ISCAS) 2008 [4]. The full paper is in the supplementary materials section in the end of this thesis, and shows some more detail on using a multi-step P-cell approach for LNA design automation.

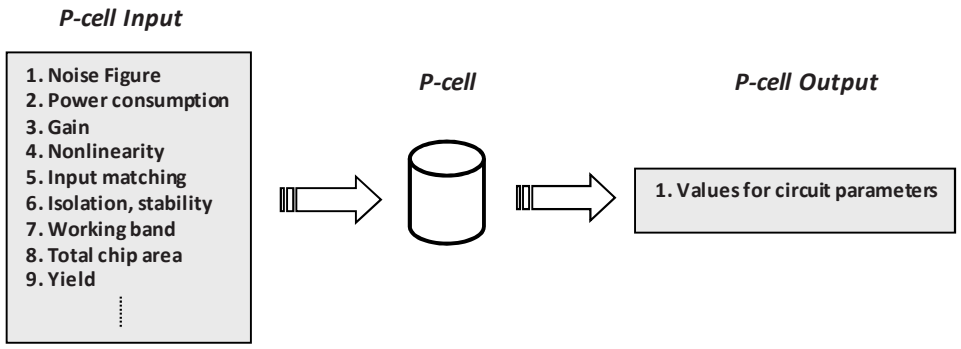


Fig. 1.1. The function of the P-cell

similar P-cell (parameterized-cell) is also available for passives such as inductors recently. Such automatic instances are less error-prone and reduce the design-to-market time, which bring the economic benefits.

As a result of the mature transistor P-cell, strong interest now moves into building P-cell for the RF block design automation. As the market for wireless communication expands, the need for RF IC with demanding performance specifications is increasing. However, the design of RF ICs is a highly changeling task, which relies on the experience of RF designers to shorten the design cycle. The P-cell for the RF block on the other hands provides the potential to reduce the design-to-market time. The ultimate goal of the P-cell is illustrated in Fig. 1.1: for any specified target circuit performance, the P-cell would calculate the optimal values for each circuit component. In other words, the P-cell determines the circuit parameters such as transistor size, bias condition and passive value automatically for any given circuit performance specifications, which is a reverse-direction approach compared with the commonly-used approach in the synthesis [1-3] (calculating the circuit performance metrics from chosen circuit parameters).

Available RF block design automation is mainly based on circuit synthesis [1-3]. Starting at some initial circuit component values such as transistor size, bias condition and passive values, the circuit performance metrics are calculated and then a cost function is evaluated. Typically, the calculation of circuit performance is done using conventional analog circuit simulators.

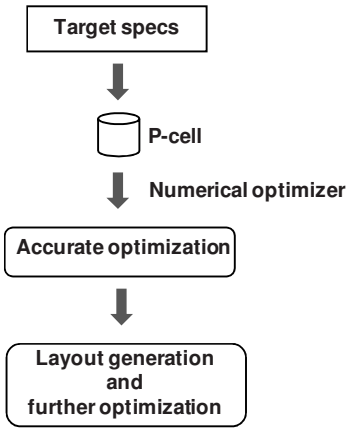


Fig. 1.2. Block diagram of the multi-step approach for RF block design automation [4].

Adaptation of circuit component values is done using e.g., a gradient descent algorithm that optimizes the pre-specified cost function. Already for small sized circuits, the number of parameters is large and numerical optimization costs lots of time, while there is usually no guarantee that the algorithm finds a solution. In mathematical terminology the main problems are due to the large search space and sticking in local minima.

To speed up the RF design automation process, we proposed a multi-step approach in [4] aiming to solve the traditional drawbacks of optimization. As illustrated in Fig. 1.2 the steps are:

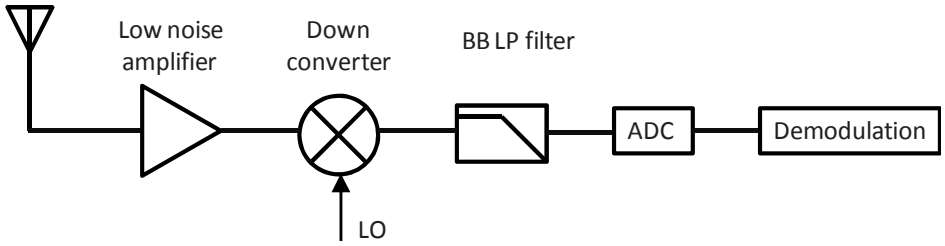
- The user selects a circuit topology and specifies target performance metrics, which are entered into the P-cell. A straight-forward extension is that also the best performing (optimized) circuit out of a set of circuit topologies is selected automatically by the P-cell based on defined cost function evaluation.
- The P-cell first makes a coarse optimization of circuit component based on the input specifications. This first step uses circuit analyses for noise and distortion modeling. A built-in interface with state-of-the-art MOS models such as PSP [5] extracts the information of in-

trinsic noise and nonlinearities of the transistor. As a result, this first step has moderate accuracy but is very fast.

- The result of the coarse optimization is used as starting point for a numerical optimizer, wrapped around conventional simulators such as Spectre [6]. Because of the relatively good starting point for the numerical optimizer, this second optimization is very fast.
- The results of the second optimization can be used as settings for the layout generation. Extraction and optimization on the extracted circuit, taking into account layout parasitics (including those of passive components) and variability, can increase accuracy at a significant calculation time penalty.

1.2 RF circuit block modeling

The most important element of the P-cell is the circuit model, which provides a mathematical link between the target performance metrics with the circuit parameters such as bias and component values [4]. Therefore, this thesis focuses on the circuit modeling of two major RF circuit blocks, low noise amplifiers (LNA) and mixers. As illustrated in Fig. 1.3, an integrated receiver usually starts with an LNA that provides suitable impedance to the antenna and amplifies the weak antenna signals. A down converter (usually a mixer) then translates the received high-frequency signal to a lower frequency. The typical LNA characteristics are noise figure (NF), gain, input matching and intermodulation distortion (IIP3 and IIP2) [7]. In a time-invariant system, time-invariant linear analyses straight-forwardly provide the model for NF, gain, input matching. In contrast, the intermodulation distortion (IIP3 and IIP2), due to inevitable curvatures of device characteristics, requires time-invariant weakly nonlinear analyses. The design specifications for the mixer are NF, gain and intermodulation distortion (IIP3 and IIP2) [7]. Different than for the LNA, the mixer not only has a RF signal input but also a periodic signal input (LO). Thus, the model for NF and gain and intermodulation distortion (IIP3 and IIP2) require analyses for time-varying system. Table 1.1 summarizes the analysis methods for the design specifications of LNAs and mixers.



1.3. Simplified block diagram of a receiver for wireless communication [7].

Table 1.1. Circuit analysis methods for LNAs and Mixers.

	Analysis method for NF, gain and input matching (only for LNA)	Analysis method for IIP3 and IIP2
LNA	Linear analysis for time-invariant system	Weakly nonlinear analysis for time-invariant system
Mixer	Linear analysis for time-varying system	Weakly nonlinear analysis for time-varying system

The time-invariant linear analysis is easy enough to be taught in popular IC design textbooks [7-8]. Nevertheless, the other three analysis methods involve complex calculations, and thus have attracted much research devotion. Firstly, the Volterra series approach is utilized for the time-invariant weakly nonlinear systems such as amplifiers [9-10]. In order to avoid the complex calculation of the Volterra series approach, an alternative approach using conventional algebra is developed for the harmonic distortion calculation of analog amplifiers [11]. For pin-pointing the transistors that contribute to the distortion dominantly, the per-nonlinearity analysis is proposed [12]. However, no information can be provided about which nonlinearity of the drain current within one transistor has more effects. The approach of [13-15] decomposes the circuit distortion output into the very basic contributor. Each nonlinearity within every transistor of one circuit is differentiated. The intention of [12-15] is to generate a compact high-level model by keeping only the dominant nonlinearity contributors.

The mixer can be considered as a linear system for noise and gain calculation or a weakly nonlinear system for the distortion calculation. The bias condition of this linear or weakly nonlinear system is modulated by the LO signal. A symbolic modeling approach is developed for the linear time-varying system using harmonic transfer matrices [16]. The algorithm based on this approach can be used to generate the linear transfer functions for the mixer. For the calculation of the mixer distortion, the time-varying Volterra series is used [17].

Not surprisingly, the research results in [9-17] altogether are able to provide sufficient tools to build the P-cell for the LNAs and mixers. After all, these two building blocks have been widely used for many years, and people ought to know them adequately well. The only limitation of the approaches of [9-17] is that they devote themselves mostly to the circuit simulation and synthesis. Complex calculations and numerical algorithm are obligatory in those approaches, which doesn't lead to enough design insights for the circuit designers to improve the circuit performance, or even to invent new circuits. A remedy to this is the focus of our thesis. The direction that this thesis tries to follow includes:

- Simplify the complex calculations in [9-17] so that circuit designers can carry out by hand-calculations, while the estimation accuracy is sufficient. The approaches proposed in this thesis enable the estimation of the time-invariant weakly nonlinear analysis by time-invariant linear analysis, and the estimation of the time-varying weakly nonlinear analysis by time-invariant nonlinear analysis. Therefore, all the performance specifications (NF, gain, impedance matching, IIP3 and IIP2) of both LNA and mixers can be estimated by time-invariant linear analysis.
- Obtain the design insights for improving the circuit performance as well as circuit innovations. Based on the design insights provided by our model, in this thesis, we propose a flicker noise/IM3 cancellation technique for active mixer, a wideband IM3 cancellation technique for CMOS attenuators and a wideband IM3 cancellation technique for LNAs with cascode topology. All three novel circuits have been implemented in silicon.

1.3 Thesis outline

Chapter 2 focus on the distortion modeling of the LNAs as the model for NF, gain and impedance is straight-forward. A general weak nonlinearity model [18] for various LNA topologies is presented. This model enables the estimation of the time-invariant weakly nonlinear analysis by time-invariant linear analysis for various LNA topologies. It's shown that, in deep-submicrometer CMOS technologies, the distortion caused by the cascode transistor cannot be neglected, due to low supply voltage and small output resistance. The general distortion model is then used to provide insights on the linearity optimization for the cascode common source amplifier and a common gate LNA, taking into account the effect of the cascode transistor.

Due to the good understanding of the distortion behavior of time-invariant circuits e.g., LNAs and attenuators, which is provided by our general distortion model, novel IM3 distortion cancellation techniques, are proposed for LNAs and attenuators.

In chapter 3, a wideband IM3 cancellation technique for LNAs with a cascode topology is discussed. A negative impedance is used to enable distortion current cancellation between the transconductor and the cascode transistor. The measurement results of a resistive feedback LNA using this IM3 cancellation technique fabricated in a standard 0.16 µm CMOS process prove the concept. Robustness of this technique with respect to process spread and bias current variations were confirmed in measurements.

In chapter 4 the wideband IM3 cancellation technique for CMOS attenuators is presented. Analytical models and simulation results show that this cancellation technique is robust against PVT (process, voltage and temperature) variations. For proof of concept, a Π-attenuator system and a T-attenuator system using this wideband IM3 cancellation technique are fabricated in a 0.16µm standard bulk CMOS process. Measurements show that very high linearity can be achieved for CMOS attenuators by using small active area in a wideband with good PVT-robustness.

Chapter 5 proposes a noise and nonlinearity model of the Gilbert mixer for fast and accurate estimation of the circuit's noise and distortion behavior. Based on closed-form expressions, this model estimates NF, IIP3 and IIP2 of the time-varying mixer by a limited number of time-invariant circuit calcula-

tions. The model shows that the decreasing transistor output resistance together with the low supply voltage in deep-submicrometer technologies contributes significantly to flicker-noise leakage. Design insights for low flicker noise are then presented. The model also shows that the slope of the LO signal has significant effect on IIP2 while little effect on IIP3. A new IP2 calibration technique using slope tuning is presented. Based on the noise and distortion analysis of Gilbert mixer, chapter 6 presents an approach to simultaneously cancel flicker noise and IM3 in Gilbert mixers, utilizing negative impedances. For proof of concept, two prototype double-balanced mixers in 0.16μm CMOS are fabricated. Measurements and simulations prove this new circuit technique.

Final conclusions and future recommendations are given in Chapter 7.

1.4 Reference

[1] G. Tulunay and S. Balkir, "Automatic synthesis of CMOS RF front-ends," *IEEE International Symp. Circuits and Systems(ISCAS)*, pp. 4, May 2006

[2] N. Roy, M. Najmabadi, R. Raut and V. Devabhaktuni, "A systematic approach towards the implementation of a low-noise amplifier in sub-micron CMOS technology," Canadian conference on Electrical and computer engineering, pp. 1909-1913, May 2006

[3] A. Nieuwoudt, T. Ragheb and Y. Massoud, "SOC-LNA: synthesis and optimization for fully integrated narrow-band CMOS low noise amplifiers," *ACM/IEEE Design Automation Conference (DAC)*, pp.879-884, July 2006

[4] W. Cheng, A. J. Annema and B. Nauta, "A multi-step P-cell for LNA design automation," *IEEE International Symp. Circuits and Systems(ISCAS),* May 2008, pp.2550-2553.

[5] http://www.nxp.com/Philips_Models/ mos_models/index.html.

[6] *Spectre Circuit Simulator User Guide*, Cadence Product Documentation.

[7] B. Razavi, *RF Microelectronics* Pearson Education, Inc., 1998.

[8] T. Lee, The Design of CMOS Radio-Frequency Integrated Circuits, 2nd ed. New York: Cambridge Univ. Press, 2004.

[9] P. Wambacq and W.S ansen, *Distortion Analysis of Analog Integrated Circuits*, Dordrecht, The Netherlands: Kluwer, 1998.

[10] P. Wambacq, G. Gielen, P. Kinget, and W. Sansen, "High-frequency distortion analysis of analog integrated circuits," *IEEE Trans. Circuits and Syst. II*, vol. 46, pp. 335–344, Mar. 1999.

[11] G. Palumbo and S. Pennisi, "High-frequency harmonic distortion in feedback amplifiers: Analysis and applications," *IEEE Trans. Circuit Syst. I, Reg. Papers,* vol. 50, no. 3, pp. 328-340, mar. 2003.

[12] P. Li and L. T. Pileggi, "Efficient per-nonlinearity distortion analysis for analog and RF circuits," *IEEE Trans. CAD Des. Integr. Circuits Syst.*, vol. 22, pp. 1297-1309, 2003.

[13] P. Dobrovolny, G. Vandersteen, P. Wambacq and S. Donnay, "Analysis and compact behavioral modeling of nonlinear distortion in analog communication circuits," *IEEE Trans. Computer-Aided Design,* vol.22, pp.1215–1227no.9, Sept. 2003.
[14] P. Wambacq, P. Dobrovolny, S. Donnay, M. Engels and I. Bolsens, "Compact modeling of nonlinear distortion in analog communication circuits," *IEEE Design, Automation & Test in Europe (DATE)* , pp. 350-354, 2000.
[15] J. Borremans, L. D. Locht, P. Wambacq and Y. Rolain, "Nonlinearity Analysis of Analog/RF Circuits using Combined Multisine and Volterra Analysis," *IEEE Design, Automation & Test in Europe (DATE)*, April 16-20, 2007.
[16] P. V anassche, G. Gielen, and W. Sansen, "Symbolic modeling of periodically time-varying systems using harmonic transfer matrices," *IEEE Trans. Computer-Aided Design,* vol.21, pp.1011–1024, no.9, Sept. 2002.
[17] P. Dobrovolny, G. Vandersteen, P. Wambacq and S. Donnay, "Analysis and white-box modeling of weakly nonlinear time-varying circuits," *IEEE Design, Automation & Test in Europe (DATE)*, pp. 624-629, 2003.
[18] W. Cheng, A. J. Annema, J. A. Croon, D. B. M. Klaasen and B. Nauta, "A general weak nonlinearity model for LNAs," *IEEE Custom Integrated Circuits Conference (CICC)*, pp.221–224, Sept. 2008.

Chapter 2

Distortion modeling for LNAs

The low noise amplifier (LNA) is a critical building block in the RF front-end. The important design specifications of the LNA are its distortion performance, typically specified in terms of IIP2 and IIP3, and linear small-signal parameters such as noise figure, input matching and gain. As discussed in section 1.2, the distortion analysis of the LNA is rather complex, while the analysis of the linear small-signal parameters is straight-forward. Thus, this chapter focuses on the distortion modeling of the LNA, which then is used to provide design insights for the LNA linearity optimization. Instead of using brute force numerical optimizers, we apply a generalized weak nonlinearity model that only involves AC transfer functions to derive simple equations for obtaining design insights. This generalized weak nonlinearity model is applied to two known RF circuits: a cascode common source amplifier and a common gate LNA. It is shown that in deep submicron CMOS technologies

This chapter is accepted for publication in IEEE Transaction on Circuits and Systems I [40]. The last section of [40] that is about attenuator linearization isn't included in this chapter, since more detailed discussion about attenuator linearization is covered in chapter 4.

the cascode transistor in both the common source amplifier and in the common gate amplifier significantly contributes IM3 distortion. Some design insights are presented for reducing the cascode transistor related distortion, among which moderate inversion biasing that improves IIP3 by 10 dB up to 5 GHz in a 90nm CMOS process. More detail of the accuracy benchmarking of this distortion model is presented in [6] and the full paper is in the supplementary materials section in the end of this thesis.

2.1 Introduction

In recent years, the need for RF ICs with demanding performance specifications has been increasing significantly. Low intermodulation distortion is one of the most desirable design targets for the current wireless front-ends. Optimizing RF front-end circuits may be done using brute force numerical optimizers with a proper set of optimization constraints, or can (partly) be done by hand if sufficient design insight is present. Circuit distortion analyses such as Volterra series have been used to either provide design insights on the RF circuit linearity [1] or to get numerical/symbolic solutions for the behavioral modeling of the front-end [2-4]. To reduce the complexity of Volterra kernels, [5] uses nonlinear system order reduction algorithms to produce compact macromodels based on Volterra series.

As alternative for the Volterra series, in [6] we presented a general weak nonlinearity model that was applied to relatively small RF circuits: the low noise amplifier (LNA). This model can easily be used to derive e.g. the circuit's intermodulation distortion in a compact closed-form expression. Due to the nature of the method, this closed-form expression is a linear combination of a number of nonlinearity coefficients of each MOS transistor and of terminal AC transfer functions. Since the AC transfer functions involve no complex calculations, it is straightforward to utilize the general distortion model for various topologies. Nevertheless, [6] only shows the accuracy benchmarking of this general model for different LNAs while no further circuit design insights are provided.

This chapter extends the general weak nonlinearity analysis method in [6] to a number of small RF circuits with four-terminal transistors; the method is

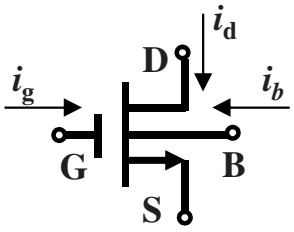


Fig. 2.1. The MOSFET as a three-input-three-output network

applied to explore the design space to optimize RF circuits and to provide design insights. Section 2.2 presents the closed-form expressions for the general nonlinearity model. Using this model, we introduce a nonlinearity cutoff frequency that indicates the relative significance of capacitive nonlinearities with respect to resistive terms for MOS transistors. This is used to simplify the general model by removing many insignificant terms from the weakly nonlinear circuit model. Section 2.3 and 2.4 discuss insights on the linearity optimization for the cascode common source RF amplifier and common gate LNA. It is shown that the distortion generated by the cascode transistor easily become dominant in the amplifier's overall distortion behavior due to the relatively large output conductance and its associated large nonlinearities. The analytical expressions indicate an IM3 cancellation scheme for amplifiers biased in the moderate inversion region. The overall conclusions are summarized in section 2.5.

2.2 The general weak nonlinearity model

2.2.1 The MOS transistor nonlinearity model

The dominant source of nonlinearity in RF circuits is usually the transistors' nonlinearity. A MOS transistor is a four-terminal device, in which all currents into the terminals and charges attributed to the terminals are nonlinear functions of the voltages across any two terminals. Mathematically, the transistor can be modeled as a three-port network with the gate-source, drain-source and bulk-source voltage as the input ports and gate current, drain current and

bulk current as outputs for any given DC bias, see Fig. 2.1. For analytical weakly nonlinear distortion analyses, Taylor series have been dominantly used to describe MOS transistor nonlinearity, where typically only the resistive nonlinearity is modeled [1-4, 7-11]. Here we present a complete weak nonlinearity model of the MOS transistor taking into account both the resistive and capacitive nonlinearity, which is given by

$$i_k(t) = \sum_K \left[G^{ks}_{nml} v^n_{gs} v^m_{ds} v^l_{bs} + C^{ks}_{nml} \frac{d(v^n_{gs} v^m_{ds} v^l_{bs})}{dt} \right] \tag{2.1}$$

$K = \{(n,m,l) | n,m,l \in \mathbb{N}; n+m+l \in (1,2,3)\}$ and $k \in \{g,d,b\}$

where $C^{ks}_{nml} = \frac{1}{n!}\frac{1}{m!}\frac{1}{l!}\frac{\partial^{(n+m+l)} Q_k}{\partial V^n_{gs} \partial V^m_{ds} \partial V^l_{bs}}\Big|_{\substack{V_{gs}=V_{GS} \\ V_{ds}=V_{DS} \\ V_{bs}=V_{BS}}}$ and $G^{ks}_{nml} = \frac{1}{n!}\frac{1}{m!}\frac{1}{l!}\frac{\partial^{(n+m+l)} I_k}{\partial V^n_{gs} \partial V^m_{ds} \partial V^l_{bs}}\Big|_{\substack{V_{gs}=V_{GS} \\ V_{ds}=V_{DS} \\ V_{bs}=V_{BS}}}$,

are respectively the capacitive and resistive coefficients. Q_k is the charge attributed to the terminal k (gate, drain or bulk) and I_k is the current into terminal k. For the first-order Taylor series terms, we have $(n+m+l) = 1$, which implies that a first derivative is taken with respect to just one port voltage. For the second-order terms, $(n+m+l) = 2$, which means that either one second derivative is used or that two first order derivative are taken with respect to port voltages. In this chapter we use only the first, second and the third-order terms, for the latter of which $(n+m+l) = 3$. For the drain terminal, the first-order coefficients G^{ds}_{100}, G^{ds}_{010} and G^{ds}_{001} correspond to the linear small signal parameters g_m, g_{ds} and g_{mb}, while C^{ds}_{100}, C^{ds}_{010} and C^{ds}_{001} are their capacitive counterparts. The higher order resistive coefficients $(G^{ds}_{200}, G^{ds}_{300})$, $(G^{ds}_{020}, G^{ds}_{030})$, and $(G^{ds}_{002}, G^{ds}_{003})$ describe second-order and third-order dependency of the resistive drain-source current respectively on V_{GS}, V_{DS} and V_{BS}, while $(C^{ds}_{200}, C^{ds}_{300})$, $(C^{ds}_{020}, C^{ds}_{030})$, and $(C^{ds}_{002}, C^{ds}_{003})$ are their capacitive counterparts. The other coefficients are the cross-modulation conductive and capacitive terms describing the dependency of drain-source current on either any two terminal or three terminal voltages. These cross-modulation terms are significant in deep sub micron CMOS technologies.

2.2.2 Generalized weakly nonlinear analysis

In the circuit example we analyzed in [6], the transistors are assumed to be three-terminal devices with interconnected bulk and source terminals. Here, we assume four-terminal transistors obeying the weakly nonlinear model given in (2.1). It is assumed that these transistors are dominant in the nonlinear behavior of the circuit with *N* transistors. We assume a two-tone input voltage $V_{\mathrm{IN}}(\mathrm{e}^{j\omega_1 t}+\mathrm{e}^{j\omega_2 t})$ with sufficiently small amplitude (V_{IN}) to ensure circuit operation in the weakly nonlinear region. The voltage swing at each port (v_{gs}, v_{ds} and v_{bs}) of each transistor results in distortion currents ($i_{gs,D}$, $i_{ds,D}$ and $i_{bs,D}$) by that transistor as described by (2.1). These distortion currents in turn generate a voltage at the ports of all transistors:

$$v_{ks,j}=\sum_{j,p=1}^{N}[E_{gs,p}^{ks,j}\cdot i_{p,D}^{gs}+E_{ds,p}^{ks,j}\cdot i_{p,D}^{ds}+E_{bs,p}^{ks,j}\cdot i_{p,D}^{bs}+F^{ks,j}(\omega_1)\cdot V_{\mathrm{IN}}e^{j\omega_1 t}+F^{ks,j}(\omega_2)\cdot V_{\mathrm{IN}}e^{j\omega_2 t}] \tag{2.2}$$

where *N* is the number of transistors in the circuit, $E_{xs,p}^{ks,j}$ is the transfer function from the current in port (x,s) of transistor p to the terminal voltage v_{ks} of transistor j, and $F^{ks,j}$ is the transfer function from voltage input to port (k,s) of transistor j, with $k,x\in\{g,d,b\}$. Since (2.2) is carried out in the frequency domain, (2.1) is rewritten into an admittance notation, $i^{ks}=\sum_K[Y_{nml}^{ks}v_{gs}^{n}v_{ds}^{m}v_{bs}^{l}]$with $Y_{nml}^{ks}(\omega)=G_{nml}^{ks}+j\omega C_{nml}^{ks}$. The generated distortion voltages result in additional distortion currents. The recursive dependency of (2.1) and (2.2) can be numerically solved by the harmonic balancing technique [7], which is often implemented in simulators. A known issue with harmonic balancing is that oversampling is required to prevent significant aliasing of higher harmonics. For the weakly nonlinear analyses done in this chapter, we assume a maximum mixing order of 3: all terms higher than third order are truncated. For weakly nonlinear systems this does not introduce significant errors, while by truncating the terms higher than third order, the number of terms remains finite and the set of equations can be analytically solved.

After truncation of higher order terms, only the terms with fundamental tones contribute to the second-order distortion, while the second-order distortion is proportional to V_{IN}^2; similarly, only the terms with fundamental tones and second-order distortion components tones contribute to the third-

order distortion components resulting in the third-order distortion proportional to V_{IN}^3. Now, a next step in the reduction of computational effort is the selection of only the frequency components leading towards the output signal component at the desired frequency (denoted as ω_D). As a result, the set of equations consisting of (2.1) and (2.2) can be analytically solved; the distortion at the circuit output is now a linear combination of the distortion contributions of each individual transistor.

$$v_{\text{out}} = \sum_{p=1}^{N} [H_p^{gs} \cdot i_{p,D}^{gs} + H_p^{ds} \cdot i_{p,D}^{ds} + H_p^{bs} \cdot i_{p,D}^{bs}] \tag{2.3}$$

where $i_{p,D}^{ks}(\omega_D) = \sum_K \beta_{nml}^{\omega_D} \cdot Y_{nml,p}^{ks}(\omega_D)$ with $\beta_{nml}^{\omega_D}$ the function that selects only the ω_D components from the product of voltages: $\beta_{nml}^{\omega_D} \triangleq (v_{gs}^n v_{ds}^m v_{bs}^l)(\omega_D)$. For IM2 calculations the function $\beta_{nml}^{\omega_D}$ thus is (with $n+m+l=2$):

$$\begin{aligned} \beta_{nml}^{\omega_{IM2}} = 0.5 \times \big[& n v_{gs}^*(\omega_2) v_{gs}^{n-1}(\omega_1) v_{ds}^m(\omega_1) v_{bs}^l(\omega_1) \\ & + m v_{ds}^*(\omega_2) v_{ds}^{m-1}(\omega_1) v_{gs}^n(\omega_1) v_{bs}^l(\omega_1) \\ & + l v_{bs}^*(\omega_2) v_{bs}^{l-1}(\omega_1) v_{gs}^n(\omega_1) v_{ds}^m \omega_1 \big] \end{aligned} \tag{2.4}$$

where v^* denotes the complex conjugate of v. The corresponding $\beta_{nml}^{\omega_D}$ for IM3 calculations is somewhat more elaborate and is given in appendix 2.1. Note that H_p^{ks} is the AC transfer function from the current in port *(k,s)* of transistor p to the output voltage and can be easily obtained by small signal analysis.

In summary, since no topology information is involved in deriving (2.3), the analysis result can be reused in any topology by deriving the topology-dependent AC transfer functions. The presented model transforms the circuit distortion calculation that usually is done by Volterra-series into rather simple AC transfer function calculations using a topology-independent transformation. When this model is used for automatic symbolic analysis of time-invariant weakly nonlinear circuits, only AC symbolic analysis is required.

2.2.3 Simplifying the transistor nonlinearity model

The nonlinearity model provides the possibility to further simplify the MOS transistor nonlinearity model given by (2.1). In (2.3) the resistive and capacitive coefficients (G_{nml}^{ks} and C_{nml}^{ks}) are combined into $Y_{nml}^{ks} = G_{nml}^{ks} + j\omega C_{nml}^{ks}$. Then a cutoff frequency $f_{nml}^{ks} = G_{nml}^{ks}/(2\pi C_{nml}^{ks})$ can be used to estimate whether

the resistive nonlinearity or the capacitive counterpart is dominant for certain operating frequencies in a specific technology for specific biasing conditions and transistor sizes. However, for transistors with fixed length (e.g. minimum length) and width that is not close to minimum width, f_{nml}^{ks} is fairly independent of transistor size (This is usually true for RF applications). This effectively leaves only bias dependent f_{nml}^{ks} factors in a certain technology.

When f_{nml}^{ks} is very high, Y_{nml}^{ks} can be reduced to a purely resistive component. Similarly, for Y_{nml}^{ks} having f_{nml}^{ks} much lower than any signal frequency, the Y_{nml}^{ks} can be seen as purely capacitive. As a result, evaluating f_{nml}^{ks} provides an approach to simplify the MOS transistor nonlinearity model for all of the weakly nonlinear RF circuits in the same transistor technology. This is different from previous work, as [3, 16] do not take the capacitive nonlinearities into account, while [17] takes a transistor as a black box and does not distinguish between the resistive and capacitive nonlinearities. Moreover, [3] removes insignificant resistive nonlinearities based on a system-level circuit model, which makes it topology-dependent, while our $f_{nml}^{ks}$is mainly dependent on bias conditions and technology parameters, and therefore largely topology-independent (appendix 2.2 shows an example for f_{210}^{ds} for an NMOS transistor). Hence, the nonlinearity parameters only need to be estimated once for a certain technology, and can then be (re)used in calculations or simulations using e.g. a look-up table.

In this chapter, a commercial 90nm CMOS process is used for demonstration purposes. All simulations are done in Spectre circuit simulator, using the PSP compact MOSFET model [18] fitted to our 90nm CMOS process.

Equation (2.3) shows that the relative importance of the nonlinearity between different terminals in one transistor can be determined by evaluating $H_p^{gs}(\omega_D) \cdot Y_{nml,p}^{gs}(\omega_D)/[H_p^{ds}(\omega_D) \cdot Y_{nml,p}^{ds}(\omega_D)]$ and $H_p^{bs}(\omega_D) \cdot Y_{nml,p}^{bs}(\omega_D)/[H_p^{ds}(\omega_D) \cdot Y_{nml,p}^{ds}(\omega_D)]$. Since $H_p^{ks}(\omega_D)$ are linear transfer functions that depend on the actual circuit topology, the evaluation of the relative impact of the nonlinearities between ports can only be used to simplify the MOS transistor nonlinearity model for individual circuits, similar to the situation in [3].

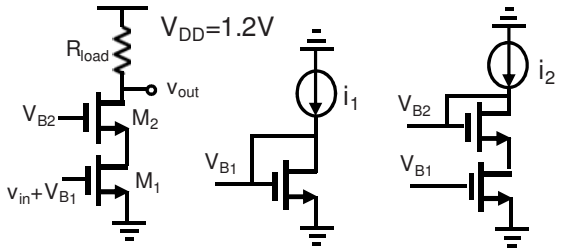


Fig. 2.2. Circuit schematic of the cascode amplifier.

2.3 **Cascode Amplifier Linearity Optimization**

The cascode amplifier topology shown in Fig. 2.2 is widely used because of its superior properties over the common-source amplifier [19-23]. Typically, the distortion contribution from cascode transistor M_2 is neglected [24-28], which is valid for sufficiently large output impedance levels of M_1. However, CMOS technology scaling yields relatively low output resistance for short transistors [29]. The distortion contribution of M_2 then can no longer be neglected.

In [30-31] only the effect of the transconductance nonlinearity G_{300}^{ds}in M_2 is analyzed, while the other nonlinearities related to the output conductance of M_2 (e.g. the third order output conductance nonlinearity G_{030}^{ds}, and the cross terms G_{210}^{ds} and G_{120}^{ds}) are neglected. In this section, we take into account all nonlinearities up to the third order and demonstrate that for low supply voltage and large gain, G_{030}^{ds}, G_{210}^{ds} and G_{120}^{ds} may be dominant in the total distortion. Note that we focus on the distortion due to the cascode transistor, therefore, we ignore the input matching for the CS amplifier and do not focus on good noise figure (NF). For simulation purpose we put a 50Ω resistor at the gate of M_1 for input matching.

Analysis results for output IM3 and a description are given below. The analysis described in the previous section shows that capacitive nonlinearities are not significant for this type of circuits in the low GHz range and can be

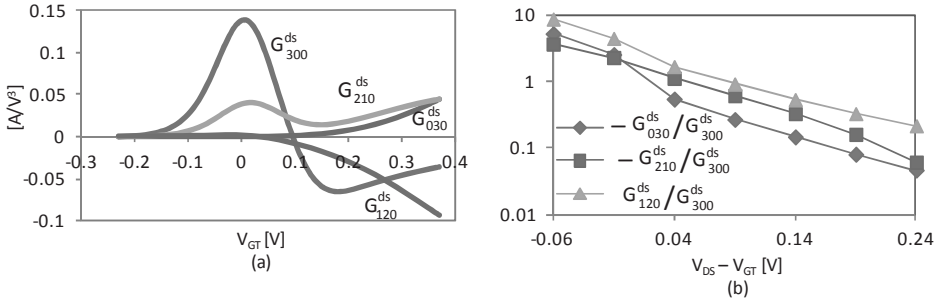


Fig. 2.3. (a) Simulated third-order nonlinearity of an NMOS as a function of the overdrive voltage V_{GT}. W/L=50/0.1µm, and $V_{DS} = 0.3V$. (b) Simulated third-order nonlinearity ratio ($-G^{ds}_{030}/G^{ds}_{300}$, $-G^{ds}_{210}/G^{ds}_{300}$ and $G^{ds}_{120}/G^{ds}_{300}$) of an NMOS as a function of the drain-source voltage V_{DS}- V_{GT}. W/L=50/0.1 µm, and $V_{GT} = 0.2V$.

neglected. The first-order approximation (see appendix 2.3 for the derivation) of the output IM3 of the cascode amplifier is

$$\begin{aligned} v_{out}^{\omega_{IM3}} \approx \frac{-3\times {V_{IN}}^3 R_{load}}{4\left(g_{ds}^{M_1}+g_m^{M_2}\right)} \times \Bigg[g_m^{M_2} \times \Bigg\{ G_{300,M_1}^{ds} - \left(\frac{g_m^{M_1}}{g_{ds}^{M_1}+g_m^{M_2}}\right)^3 G_{030,M_1}^{ds} \\ + \left(\frac{g_m^{M_1}}{g_{ds}^{M_1}+g_m^{M_2}}\right)^2 G_{120,M_1}^{ds} - \left(\frac{g_m^{M_1}}{g_{ds}^{M_1}+g_m^{M_2}}\right) G_{210,M_1}^{ds} \Bigg\} \\ + g_{ds}^{M_1} \left(\frac{g_m^{M_1}}{g_{ds}^{M_1}+g_m^{M_2}}\right)^3 \times \Big\{ G_{300,M_2}^{ds} + \left(g_m^{M_2} R_{load}-1\right)^3 G_{030,M_2}^{ds} \\ + \left(g_m^{M_2} R_{load}-1\right)^2 G_{120,M_2}^{ds} - \left(g_m^{M_2} R_{load}-1\right) G_{210,M_2}^{ds} \Big\} \Bigg] \end{aligned} \tag{2.5}$$

Assuming $g_m^{M_1} \approx g_m^{M_2} \gg g_{ds}^{M_1}$ and $g_m^{M_2} R_{load} \gg 1$, (2.5) can be further simplified to

$$\begin{aligned} v_{out}^{\omega_{IM3}} \approx \frac{-3\times {V_{IN}}^3 R_{load}}{4\left(1+g_m^{M_2} r_{ds}^{M_1}\right)} \times \Big[g_m^{M_2} r_{ds}^{M_1} \left(G_{300,M_1}^{ds} + G_{120,M_1}^{ds} - G_{030,M_1}^{ds} - G_{210,M_1}^{ds}\right) \\ - \left(g_m^{M_1} R_{load}\right)^3 \cdot G_{030,M_2}^{ds} - g_m^{M_1} R_{load} \cdot G_{210,M_2}^{ds} \\ + G_{300,M_2}^{ds} + \left(g_m^{M_1} R_{load}\right)^2 \cdot G_{120,M_2}^{ds} \Big] \end{aligned} \tag{2.6}$$

where transistor nonlinearitis G_{300}^{ds}, G_{030}^{ds}, G_{120}^{ds} and G_{210}^{ds} are extracted from simulation as shown in Fig. 2.3.

Fig. 2.3(a) shows that in the strong inversion region the nonlinearities G_{300}^{ds} and G_{120}^{ds} have the same sign (negative) while G_{030}^{ds} and G_{210}^{ds} have the opposite

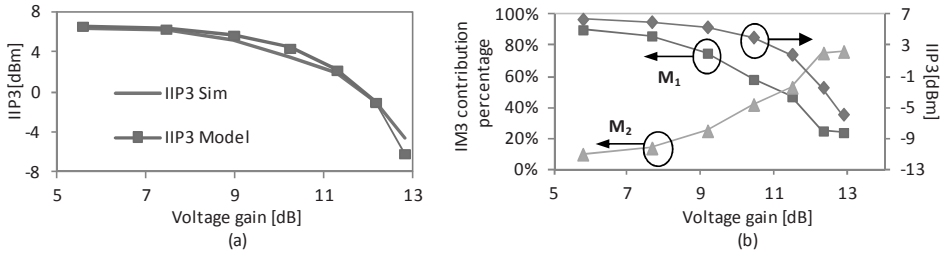


Fig. 2.4. (a) The simulated IIP3 and the calculated IIP3 by the model which only includes the transistor third-order nonlinearity. For M_1, W/L=50/0.1µm, V_{B1}=0.6 V, V_{th}=0.42 V; for M_2, W/L=50/0.1µm, V_{B2}=1.2 V. (b) The calculated IM3 contribution from M_1 and M_2 to the circuit output and simulated IIP3 in different voltage gain settings.

sign (positive). It follows from equation (2.6) that the contribution from each third-order nonlinearity adds up in the circuit output. Equation (2.6) also shows that a large output resistance of M_1($r_{ds}^{M_1}$) results in negligibly small distortion contributions of M_2 for the total IM3 compared to the contributions of M_1. However, in deep-submicrometer CMOS technologies, typically the output resistance of M_1 is relatively low: the IM3 distortion contribution from M_2 then may become dominant. On top of that, the low supply voltage together with high gain operation tends to push M_2 out of the deep saturation region, resulting in a very significant increase of the third-order output conductance nonlinearity term G_{030,M_2}^{ds} and the cross-modulation nonlinearities G_{120,M_2}^{ds} and G_{210,M_2}^{ds}, see Fig. 2.3(b).

To demonstrate the increasing IM3 contribution from M_2 as the gain increases, Fig. 2.4 shows simulation results for the cascode amplifier in Fig. 2.2 for different gain settings by only changing R_{load} (from 100Ω to 250Ω) for a constant bias current. The two-tone signals are at 1GHz and 1.01GHz and the IIP3 is extrapolated by sweeping the input power from -35 to -25dBm. Fig. 2.4(a) shows that the model given by (2.5), including only the third-order transistor nonlinearity terms, provides an accurate IIP3 estimation for different gain settings. Fig. 2.4(b) shows that the voltage gain increases, while the IIP3 decreases with increasing R_{load}. For larger R_{load}, the drain voltage of M_2 decreases and M_2 comes out of deep saturation. As a result, the third-order nonlinearity terms G_{030,M_2}^{ds}, G_{120,M_2}^{ds} and G_{210,M_2}^{ds} increase significantly, which

results in a significant increase of IM3 distortion. Fig. 2.4(b) illustrates that the IM3 contribution of the cascode transistor M_2 then can be higher than that of M_1. Therefore, in deep-submicrometer CMOS technologies, linearity optimization must take into account not only G_{300}^{ds} and G_{030}^{ds}, but also cross-modulation terms G_{210}^{ds} and G_{120}^{ds}. These cross modulation-terms are usually neglected [30-31].

Expression (2.6) suggests that optimal bias for linearity should prevent M_2 from approaching the triode region, where the third-order output conductance nonlinearity $G_{030,M_2}^{ds}$and the cross-modulation nonlinearity (G_{120,M_2}^{ds} and G_{210,M_2}^{ds}) are maximum. This leads to the following three linearity optimization approaches:

- optimiz the gate bias voltage of cascode transitor
- use components to bypass part of the DC bias current
- use distortion cancellation for the transistors.

For demonstration purposes, the linearity optimization is performed for the cascode amplifier in Fig. 2.2 with R_{load} = 250Ω, where transistor M_1 and M_2 are biased in strong inversion. The dimension and bias condition (g_{m1}=g_{m2}=20mS, W_1/L_1=W_2/L_2=50/0.1µm, I_{DC}=2.23mA, V_{B1}=0.6V, V_{th}=0.42V and V_{B2}= V_{DD}=1.2V) provide 12.8dB voltage gain, 7.4dB NF, -4.5dBm IIP3 and -14dBm $P_{1\text{-dB}}$. The IIP3 is extrapolated by sweeping the input power from -35 to -25dBm with a two-tone signal at 1GHz and 1.01GHz. This design serves as reference for comparison with the optimized designs.

2.3.1 Optimizing the cascode transistor gate bias voltage

One approach for linearization is to adjust the gate bias of cascode transistor M_2. Usually the gate voltage is equal to the supply voltage V_{DD}. In this section, it is shown that other (DC-) voltages may result in better performance; we do not address applying AC-variations (e.g. gain boosting) for simplicity reasons.

It can be derived from (2.6) and the relations between the transistor nonlinearities and biasing conditions that by adjusting the gate bias of M_2 (V_{B2}) the overall circuit linearity can be optimized. For low cascode gate bias levels, M_1 is biased between the saturation region and triode region where its

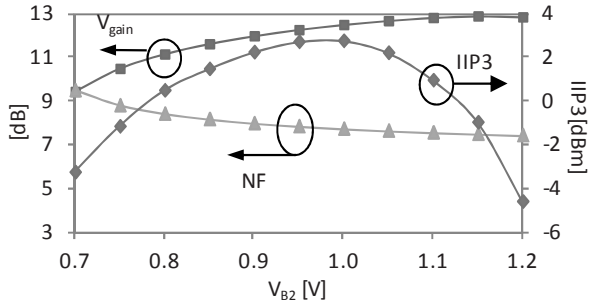


Fig. 2.5. The simulated noise figure, voltage gain and IIP3 of the cascode amplifier shown in Fig. 2.2 as a function of the gate bias V_{B2} of M_2 for a constant power consumption. For M_1, W/L=50/0.1µm, V_{B1}=0.6 V; for M_2, W/L=50/0.1µm.

output conductance nonlinearity G_{030,M_1}^{ds} and the cross-modulation nonlinearity (G_{120,M_1}^{ds} and G_{210,M_1}^{ds}) are high, resulting in rather low IIP3. At high cascode gate bias voltage levels the cascode transistor M_2 may go out of saturation which increases its nonlinearities G_{030,M_2}^{ds}, G_{120,M_2}^{ds} and G_{210,M_2}^{ds}. In between these two extremes, the total distortion of the two transistors is minimum, and typically dominated by the third-order transconductance nonlinearity G_{030}^{ds} of M_1. Fig. 2.5 shows that for the reference LNA design a cascode transistor gate bias in the range of 1 V to 1.05V yields maximum IIP3 with slightly degraded NF and voltage gain.

2.3.2 Usage of bypass components

One of the dominant effects with respect to distortion is the limited voltage headroom for either M_1 or M_2, which is among others limited by the DC-voltage drop across the resistor. Using components to bypass part of the DC-current increases the headroom and hence decreases distortion.

One way to implement this is to add a pMOS load or an (on-chip) inductor in parallel to R_{load}. A parallel pMOS load (M_3) conducts a part of the DC current and lifts up the drain voltage of M_2. As a result, the output conductance nonlinearity G_{030,M_2}^{ds} and the cross-modulation nonlinearity terms (G_{120,M_2}^{ds} and G_{210,M_2}^{ds}) of M_2 decrease. For the first-order approximation the output IM3 of the cascode amplifier given by (2.6) changes to

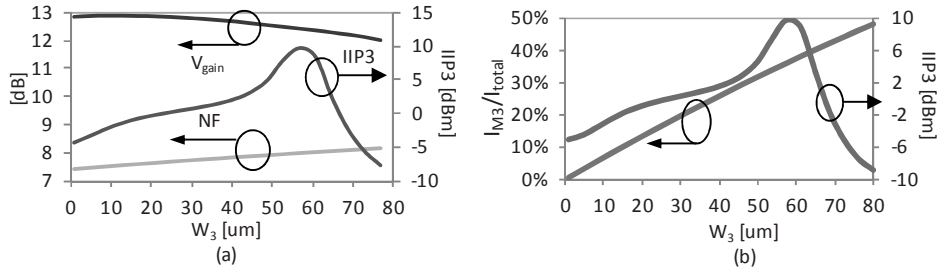


Fig. 2.6. Simulation results of the cascode amplifier with the pMOS load as a function of the width M_3. (a) NF, voltage gain and IIP3. (b) IIP3 and the dc current supplied by the pMOS load M_3 divided by the total dc current.

$$v_{out}^{\omega_{IM3}} \approx \frac{-3 \times {V_{IN}}^3 R_{load}}{4\left(1+g_m^{M_2} r_{ds}^{M_1}\right)} \times \Big[g_m^{M_2} r_{ds}^{M_1} \left(G_{300,M_1}^{ds} + G_{120,M_1}^{ds} - G_{030,M_1}^{ds} - G_{210,M_1}^{ds}\right) - \left(g_m^{M_1} R_{load}\right)^3 \cdot G_{030,M_2}^{ds} - g_m^{M_1} R_{load} \cdot G_{210,M_2}^{ds} + G_{300,M_2}^{ds} + \left(g_m^{M_1} R_{load}\right)^2 \cdot G_{120,M_2}^{ds} - \left(1+g_m^{M_2} r_{ds}^{M_1}\right) \cdot \left(g_m^{M_1} R_{load}\right)^3 \cdot G_{030,M_3}^{ds} \tag{2.7}$$

where the last term represents the distortion contribution from M_3 via its output conductance nonlinearity G_{030,M_3}^{ds}. Although M_3 contributes additional distortion, the circuit linearity can still be improved with a proper design. Fig. 2.6 shows the simulation result for the cascode amplifier with pMOS load M_3 in parallel to R_{load} by sweeping the width of M_3 (W_3). A channel length three times the minimum length is used to increase the output resistance of M_3 for keeping the voltage gain almost unchanged. As W_3 increases, the drain voltage of M_2 increases since less dc current passes through R_{load}. The IIP3 increases as M_2 enters further into the saturation region. More DC current through M_3 further increases the drain voltage of M_2. This pushes M_3 out of deep saturation and causes more distortion and noise from M_3. The IIP3 is optimum at the region where both the cascode transistor M_2 and the M_3 are in saturation. Then the output conductance nonlinearity G_{030,M_2}^{ds}, the cross-modulation nonlinearity (G_{120,M_2}^{ds} and G_{210,M_2}^{ds}) of M_2 and the output conductance nonlinearity G_{030,M_3}^{ds} of M_3 are less significant than the third-order transconductance G_{030,M_1}^{ds} of M_1. Alternatively, an on-chip stacked inductor load can also be used to increase the drain voltage of M_2 [32]. However, for

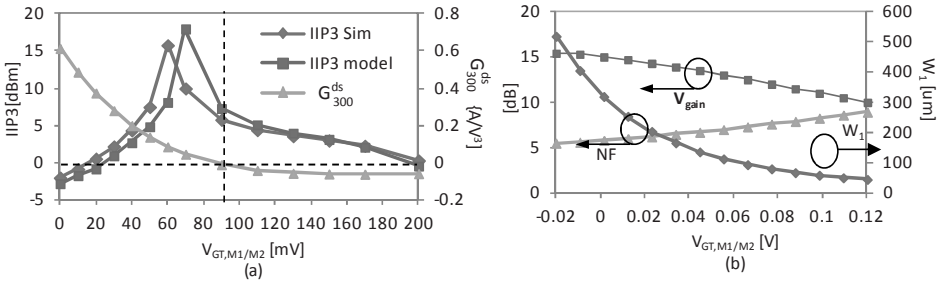


Fig. 2.7. (a) The simulated IIP3 and the calculated IIP3 modeling only third-order nonlinearity. (b) NF, voltage gain and transistor width as a function of V_{GT} for a constant 1.17 mA current.

frequencies in the lower GHz range, the low quality factor introduces rather small shunt parasitic resistance that limits the amplifier gain. Moreover, on-chip stacked inductors typically consume much more area than a pMOS load [32-33].

2.3.3 Optimal bias in moderate inversion region

Assuming that the main nonlinearity of a MOS transistor arises from transconductance nonlinearity G^{ds}_{300}, the IIP3 sweet spot of the single transistor amplifier coincides with the setting at which G^{ds}_{300} is zero [9]. Due to increasingly nonlinear output conductance and cross terms in submicron CMOS technologies, the actual IIP3 sweet spot of a single transistor amplifier however does not coincide with zero-G^{ds}_{300} [9, 11]. As the cascode transistor may contribute significant distortion, the effect of the cascode transistor on the IIP3 sweet spot needs to be included.

The simplified model in (2.6) is used to estimate the IIP3 sweet spot of the cascode amplifier. Fig. 2.3(a) shows that in moderate inversion the nonlinearities G^{ds}_{300}, G^{ds}_{030} and G^{ds}_{210} are positive and G^{ds}_{120} is negative. Thus the distortion generated by G^{ds}_{300} of M_1 and M_2 cancels the distortion of all the other nonlinearities within M_1 and M_2 as suggested by (2.6). As illustration for this, Fig. 2.7 shows the simulation and calculation result for the cascode amplifier where M_1 and M_2 are set to have a constant g_m of 20mS at 1GHz, which is the same as in the reference design. Firstly, Fig. 2.7(a) shows that

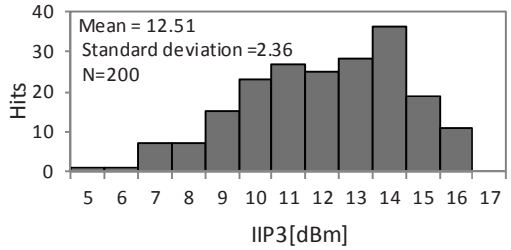


Fig. 2.8 Simulated IIP3 of the cascode amplifier optimized in the moderate inversion region in Monte Carlo simulation (200run) for mismatches and process corner at 1GHz.

the model given by (2.6) including only the third-order transistor nonlinearity provides an accurate IIP3 estimation for the moderate inversion bias region. As shown in (2.6) and Fig 2.4(a), for very low V_{GS}, G_{300,M_1}^{ds} and G_{300,M_2}^{ds} are large and dominantly contribute to the output distortion. As V_{GS} increases, G_{300,M_1}^{ds} and G_{300,M_2}^{ds} start to decrease and their distortion cancels the distortion generated by the other transistor nonlinearities; this enables a high-IIP3 region around V_{GT}=70mV, which is about 20mV away from the zero-G_{300}^{ds} setting illustrated by the dashed line in Fig. 2.7(a). For large V_{GS} when the transistors enter strong inversion, G_{300,M_1}^{ds} and G_{300,M_2}^{ds} get negative and there is no distortion cancellation. Based on Fig. 2.7 we choose one optimal design ($W_1/L_1=W_2/L_2$=104/0.1μm, V_{GT}=70mV, I_{DC}=1.17mA). Compared to the reference cascode amplifier design, the transistor width is doubled while the DC current is about halved.

Fig. 2.8 shows that for a set of 200-time Monte Carlo simulation with mismatch and process corner spread the moderate inversion optimal region enables mean IIP3 of 12.5dBm at 1GHz, which is an improvement of about 16dB compared to the reference design operating in strong inversion. To

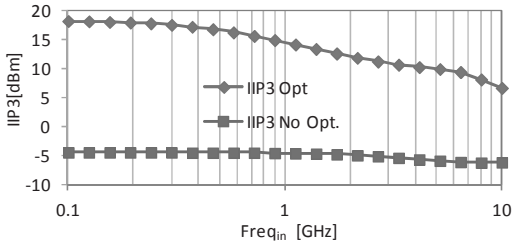


Fig. 2. 9 Simulated IIP3 of the cascode amplifier optimized in the moderate inversion region over input frequency.

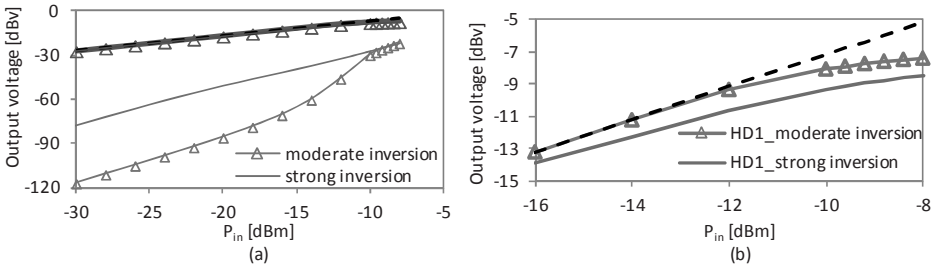


Fig. 2.10. (a) Simulated HD1 and IM3 for varying input power. (b) Simulated HD1 for varying input power denoting the 1dB compression.

illustrate frequency-dependencies, Fig. 2.9 shows the simulated results of this optimal design for input signal frequency from 0.1GHz to 10GHz. Fig. 2.9 shows that optimal bias in the moderate inversion improves IIP3 by more than 10dB for frequencies up to 10GHz. The cancellation degrades at higher frequencies because of increasing phase shifts between the distortion components generated by G_{300}^{ds} and by the other nonlinearity components. The simulated IM3 and HD1 for varying input power in Fig.2.10 shows that the IM3 cancellation in the moderate inversion region becomes less effective for input signals larger than -15dBm. This is due to higher-order transistor nonlinearities. Since the voltage drop across R_{load} is halved in the moderate inversion region, there is more headroom for the output swing and hence increases $P_{1\text{-}dB}$ from -14dBm to -10dBm.

2.3.4 Summary

The general weak nonlinearity model used for the cascode amplifier topology shows that the cascode transistor M_2 may contribute significantly to

TABLE 2.1

Comparision of IIP3 Optimization approaches

	I_{DC} [mA]	IIP3 [dBm]	Gain [dB]	NF [dB]	Active area [μm^2]	P_{1dB} [dBm]
Ref. Design	2.23	-4.5	12.8	7.4	10	-14
Opt. A	2.23	2.7	12.5	7.7	10	-11.7
Opt. B	2.23	9.7	12.5	8	27.1	-13.8
Opt. C	1.17	14.5	12.8	7.1	20.8	-10

Opt.A: cascode transistor gate bias adjustment.
Opt.B: pMOS load.
Opt.C: moderate inversion biasing.

the overall distortion, especially in high gain settings in deep-submicrometer CMOS. A number of ways to minimize distortion were discussed, among which optimum gate biasing of the cascode transistor, using DC-current bypass components, and enabling distortion cancellation in moderate inversion operation.

Table 2.1 lists the simulation results of optimal linearity designs using the three optimization approaches discussed in section 2.2. The optimal designs are obtained using the data in Fig. 2.5, Fig. 2.6 and Fig. 2.7 respectively. Table 2.1 shows that only adjusting the cascode transistor gate bias (Opt.A) increases IIP3 by 6dB, while gain and noise are slightly affected. This approach takes no extra active area. For higher IIP3 either the pMOS load (Opt.B) should be used or the cascode amplifier should be biased in the moderate inversion (Opt.C). Both Opt.B and Opt.C need extra active area while gain and noise are slightly affected. However, biasing the amplifier in the moderate inversion (Opt.C) uses about 50% less current and achieves thelargest IIP3 improvement due to the distortion cancellation between M_1 and M_2 while little effect on gain and NF. Less dc bias current provides more headroom for the output swing and increases P_{1dB}. As shown in section 2.2.3, for all process corners and for frequencies up to 10 GHz biasing in moderate inversion appears to be optimum.

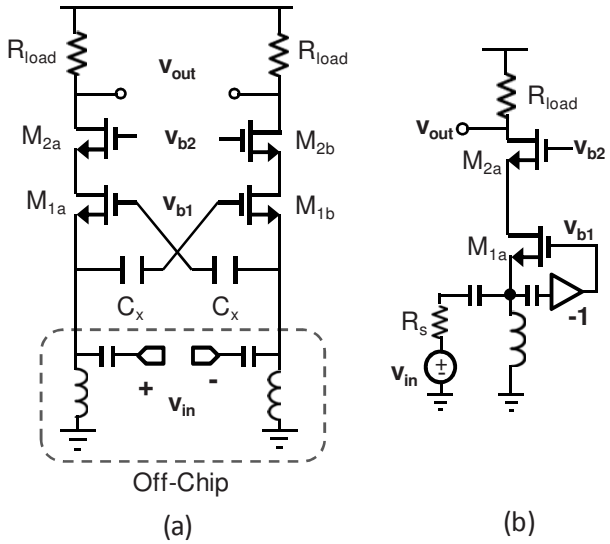


Fig. 2.11 Schematic of the capacitive cross-coupled common-gate LNA. (a) Differential schematic and (b) half-circuit model.

2.4 **Common-Gate LNA Linearity Optimization**

Due to the strict demands on input matching, the transconductance (g_m) for a common-gate (CG) LNA is fixed, resulting in difficulties in simultaneously providing NF<3dB and high gain [34-39]. The cross-coupled capacitors shown in Fig. 2.11 are frequently used for g_m boosting in order to achieve high gain and low NF [34-36, 38-39]. For 50Ω input matching, M_{1a} and M_{1b} are dimensioned for a fixed transconductance ($g_m = 1/2R_s$): for high gain and a low NF then a large R_{load} is required. Similar to the discussions in section 2.2 for the cascode common source amplifier, the large R_{load} decreases the drain voltage of M_{2a}/M_{2b} and tends to push M_{2a}/M_{2b} out of deep-saturation. As a result, the third-order output conductance nonlinearity G^{ds}_{030,M_2} and the cross-modulation nonlinearity (G^{ds}_{120,M_2} and G^{ds}_{210,M_2}) increases dramatically and contributes more IM3 distortion. For demonstration, we simulate the CG LNA shown in Fig. 2.11(a) in different gain settings by sweeping R_{load} (from

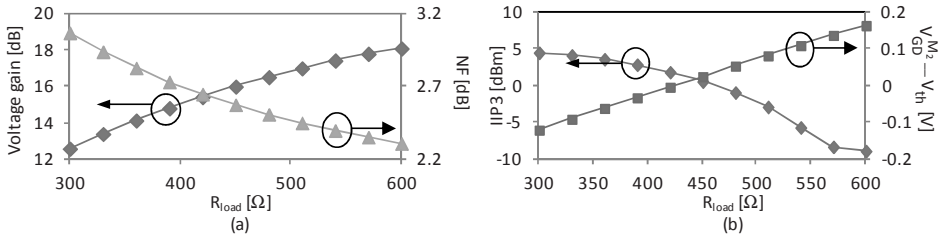


Fig. 2.12 The simulated (a) NF and voltage gain (b) $V_{GD}^{M_2} - V_{th}$ of M_{2a}/ M_{2b} and IIP3 of the CG LNA shown in Fig. 2.10a as a function R_{load} for a constant power consumption. For M_1 and M_2, W/L=26/0.1µm, V_{GTM1}=0.16 V, V_{b2}=V_{dd}=1.2 V.

300Ω to 600Ω) for a constant 2mA DC current, while keeping S_{11}<-10dB. The two-tone signals are at 1 GHz and 1.01 GHz and the IIP3 is extrapolated by sweeping the input power from -35 to -25 dBm. Fig. 2.12 shows that larger R_{load} improves NF and gain at the price of decreasing IIP3 since M_{2a}/M_{2b} are pushed out of the deep saturation region. The optimal bias region can be estimated using the expression for the output IM3 of the CG LNA half-circuit model shown in Fig. 2.11(b) (see appendix 2.4 for the derivation):

$$\begin{aligned} v_{out}^{\omega_{IM3}} \approx \frac{-3\times {V_{IN}}^3 R_{load}}{32\left(g_{ds}^{M_1}+g_m^{M_2}+g_{ds}^{M_1}g_m^{M_2}R_s+2g_m^{M_1}g_m^{M_2}R_s\right)}\times\Big\{g_m^{M_2}\times\Big[-8G_{300,M_1}^{ds} \\ +\left(\frac{1}{g_m^{M_2}R_s}-1\right)^3 G_{030,M_1}^{ds}-2\left(\frac{1}{g_m^{M_2}R_s}-1\right)^2 G_{120,M_1}^{ds}+4\left(\frac{1}{g_m^{M_2}R_s}-1\right)G_{210,M_1}^{ds}\Big] \\ +\frac{g_{ds}^{M_2}}{\left(g_m^{M_2}R_s\right)^3}\times\Big[-G_{300,M_2}^{ds}+\left(g_m^{M_2}R_{load}-1\right)^3 G_{030,M_2}^{ds} \\ -\left(g_m^{M_2}R_{load}-1\right)^2 G_{120,M_2}^{ds}+\left(g_m^{M_2}R_{load}-1\right)G_{210,M_2}^{ds}\Big]\Big\} \end{aligned} \tag{2.8}$$

Assuming that $g_m^{M_1} \approx g_m^{M_2} \approx 1/2R_s \gg g_{ds}^{M_1}$ and $g_m^{M_2}R_{load} \gg 1$, (2.8) can be simplified to

$$\begin{aligned} v_{out}^{\omega_{IM3}} \approx \frac{-3\times {V_{IN}}^3 R_{load}R_s}{32}\times\Big\{g_m^{M_2}\times\left(-8G_{300,M_1}^{ds}+G_{030,M_1}^{ds}-2G_{120,M_1}^{ds}+4G_{210,M_1}^{ds}\right) \\ +8g_{ds}^{M_2}\Big[-G_{300,M_2}^{ds}+\left(g_m^{M_2}R_{load}\right)^3 G_{030,M_2}^{ds}-\left(g_m^{M_2}R_{load}\right)^2 G_{120,M_2}^{ds} \\ +\left(g_m^{M_2}R_{load}\right)G_{210,M_2}^{ds}\Big]\Big\} \end{aligned} \tag{2.9}$$

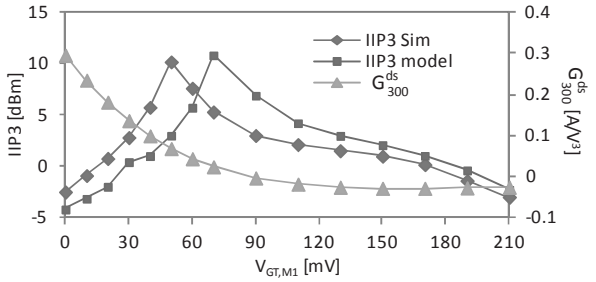


Fig. 2.13 The simulated and the calculated IIP3 and G_{300}^{ds} as a function of overdrive voltage of M_1 and M_2.

For the CG LNA, (2.9) suggests a similar IM3 cancellation scheme as for the cascode CS amplifier discussed in section 2.2. In the moderate inversion region, the transconductance nonlinearities G_{300}^{ds} turn into positive values. Thus, the distortion generated by G_{300}^{ds} of M_1 and M_2 cancels the distortion of all the other nonlinearities within M_1 and M_2. Fig. 2.13 shows the simulated IIP3 of the CG LNA as well as the calculated IIP3 using (2.8). Both M_1 and M_2 have the same dimension and a constant transconductance (g_m=9mS for S_{11}<-25dB) for different overdrive voltage V_{GT}. The load R_{load} is set to 600Ω to achieve 18dB voltage gain and 2.3dB NF. The two-tone signals are at 1 GHz and 1.01 GHz and, the IIP3 is extrapolated by sweeping the input power from -35 to -25 dBm. Fig. 2.13 shows that the model given by (2.8) provides a good prediction on the IIP3 changing trend. For very low V_{GT}, G_{300,M_1}^{ds} and G_{300,M_2}^{ds} are large and dominantly contribute to the output distortion. As V_{GT} increases, G_{300,M_1}^{ds} and G_{300,M_2}^{ds} start to decrease and their distortion cancels the distortion generated by the other transistor nonlinearity terms. This cancellation enables a high-IIP3 region around V_{GT} =50mV. For large V_{GT} the transistors enter the strong inversion region, and G_{300,M_1}^{ds} and G_{300,M_2}^{ds} become negative and as a result no distortion cancellation can take place between M_1 and M_2. For comparison we simulate two LNA designs at 1GHz. The load R_{load} is set to 600Ω and a 100nH inductor with Q=80 is used to model the off-chip inductor. In both designs M_1/M_2 are set to g_m=9mS, while in LNA1 the transistors are biased in strong inversion region and in LNA2 the transistors are biased in moderate inversion. Table 2.2 shows that

TABLE 2.2

Comparision of CG LNA in different bias regions

	V_{GT} [mV]	I_{dc} [mA]	IIP3 [dBm]	Gain [dB]	NF [dB]	S_{11} [dB]	$P_{1\text{-}dB}$ dBm]	$W_{M1/M2}$ [um]
LNA1	170	1.92	-5	18.2	2.35	-29	-15.5	26
LNA2	48	1	10	18	2.32	-26	-13.5	56

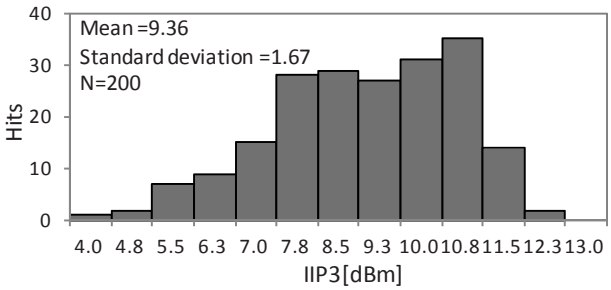


Fig. 2.14 Simulated IIP3 of the cascode amplifier optimized in the moderate inversion region in Monte Carlo simulation (200run) for mismatches and process corner.

in the optimal moderate inversion region, IIP3 is improved by 15dB; the DC current is decreased by 50% ,while NF, gain and input matching stay the same. However, the price to be paid is about 3dB smaller bandwidth since the transistor width increases by about two times in the optimal moderate inversion region.

A 200-sample Monte Carlo simulation with mismatch and corner spread shows in Fig. 2.14 that moderate inversion biasing yields a mean IIP3 of 9.4dBm, which is about 14dB higher than biasing in saturation. Fig. 2.15 shows the IIP3 for the designs as a function of frequency from 0.1GHz to 10 GHz. It is shown that the optimal bias in the moderate inversion improves IIP3 by more than 10dB up to 5GHz. The distortion cancellation degrades towards higher frequencies because of phase shifts between the distortion components due to G_{300}^{ds} and due to the other nonlinearity terms. Fig. 2.16 shows the simulated IM3 and HD1 as a function of input power; the IM3 cancellation in moderate inversion becomes less effective for input signals

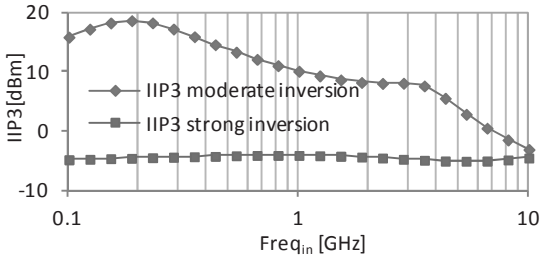


Fig. 2.15 Simulated IIP3 of the CG LNA optimized in the moderate inversion region over input frequency.

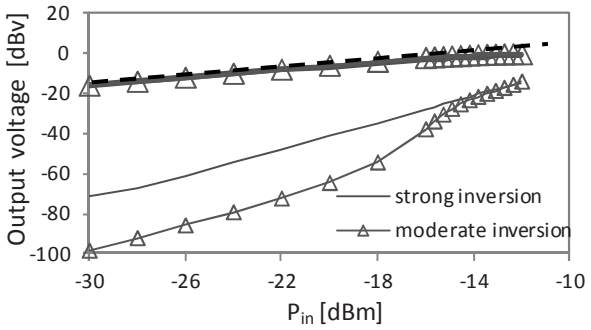


Fig. 2.16 Simulated HD1 and IM3 for varying input power.

larger than -18dBm. This is due to transistor nonlinearities higher than third-order. Since the voltage drop across R_{load} is halved in the moderate inversion region, there is more headroom for the output swing, and hence this increases $P_{1\text{-}dB}$ from -15.5dBm to -13.5dBm.

2.5 **Conclusion**

We introduced a generalized weak nonlinearity analysis method, which is somewhat related to harmonic balance analyses. It can obtain closed-form expressions for circuit distortion. Due to the nature of the method, the obtained expressions consist of technology dependent transistor nonlinearity parameters and topology-dependent AC transfer functions only. Simple

techniques were introduced to maximally decrease computational effort, such as limiting calculations in such a way that only signals leading to the targeted distortion component are included in the calculations. Secondly a nonlinearity cutoff frequency f_{nml}^{ks} was used to determine the relative importance between the resistive nonlinearity and its capacitive counterpart and to allow for omission of nonlinearity terms. The characterization results of f_{nml}^{ks} is topology-independent and can be (re)used for all the circuit designs in the same process, which improves the efficiency of numerical circuit optimization.

The general weak nonlinearity model is applied to two RF circuits to explore the design space for linearity optimization insights that is usually available in today's deep-submicrometer CMOS technologies. We show that in a standard cascode LNA circuit, the cascode transistor can significantly contribute to distortion in deep-submicrometer CMOS technologies. This is due to the low supply voltage and the decreasing output resistance. A number of ways to decrease the distortion with (almost) unchanged NF and gain are discussed, including DC-current bypass components and biasing the transistor in the moderate inversion region to get distortion cancellation. For both common source amplifier and common gate LNA, this IM3 cancellation scheme provides robustly more than 10dB IIP3 improvement for signal frequencies up to 5GHz in a 90nm CMOS process.

Appendix 2.1

For calculating the output IM3, the following function takes into account only the terms leading to the signals at $\omega_{IM3} = 2\omega_2 - \omega_1$

$$\begin{aligned}\beta_{nml}^{\omega_{IM3}}\big|_{n+m+l=3} = \frac{{V_{IN}}^3}{4} \cdot [\, & n \cdot {v_{gs}}^*(\omega_1){v_{gs}}^{n-1}(\omega_2){v_{ds}}^m(\omega_2){v_{bs}}^l(\omega_1) \; + m \\ & \cdot {v_{gs}}^n(\omega_2){v_{ds}}^*(\omega_1){v_{ds}}^{m-1}(\omega_2){v_{bs}}^l(\omega_2) + l \\ & \cdot {v_{gs}}^n(\omega_2){v_{ds}}^m(\omega_2){v_{bs}}^*(\omega_1){v_{bs}}^{l-1}(\omega_2)\,]\end{aligned} \tag{2.10}$$

$$\beta_{nml}^{\omega_{IM3}}\Big|_{n+m+l=2} = \frac{{V_{IN}}^3}{4}\cdot[n\cdot {v_{gs}}^*(\omega_1-\omega_2){v_{gs}}^{n-1}(\omega_2){v_{ds}}^m(\omega_2){v_{bs}}^l(\omega_1)$$
$$+m\cdot {v_{gs}}^n(\omega_2){v_{ds}}^*(\omega_1-\omega_2){v_{ds}}^{m-1}(\omega_2){v_{bs}}^l(\omega_2)$$
$$+l\cdot {v_{gs}}^n(\omega_2){v_{ds}}^m(\omega_2){v_{bs}}^*(\omega_1-\omega_2){v_{bs}}^{l-1}(\omega_2) \qquad (2.11)$$
$$+n\cdot {v_{gs}}^*(\omega_1){v_{gs}}^{n-1}(2\omega_2){v_{ds}}^m(2\omega_2){v_{bs}}^l(2\omega_2)$$
$$+m\cdot {v_{gs}}^n(2\omega_2){v_{ds}}^*(\omega_1){v_{ds}}^{m-1}(2\omega_2){v_{bs}}^l(2\omega_2)$$
$$+l\cdot {v_{gs}}^n(2\omega_2){v_{ds}}^m(2\omega_2){v_{bs}}^*(\omega_1){v_{bs}}^{l-1}(2\omega_2)]$$

In this, $*$ denotes the complex conjugate function.

Appendix 2.2

The MOS transistor resistive nonlinearity can be extracted in many ways in time domain or in frequency domain. In this work we derived the nonlinearity coefficients from simulations using Spectre and a well fitted PSP model. The PSP model is known to be able to correctly fit at least up to the third derivative [12, 13, 18]. Using a PSP model has advantages over getting derivatives from measurements mainly because measurement noise is largely eliminated: the PSP model can be used to accurately smoothen measurement results.

The resulting nonlinearity parameters scale (as a good approximation) linearly with transistor width which allows normalization with respect to transistor width. Furthermore, transistor length is assumed to be minimum. Then the nonlinearity parameters are mainly functions of port voltages, and need to be determined just once for each technology. Storing them in e.g. a look-up table then allows for computational efficient use in e.g. calculations.
As an example, the f_{210}^{ds} is extracted from simulations, as a function of V_{GS} and V_{DS} for a minimum length transistor, for the 90nm CMOS process used throughout this chapter. The resulting contour plot of f_{210}^{ds} is shown in Fig. 2.17; for readability, $\lg(f_{210}^{ds}/1GHz)$ is plotted; the minimum value of 2 in the plot hence corresponds to f_{210}^{ds} =100GHz. The plot indicates that at frequencies lower than 10 GHz, G_{210}^{ds} is dominant compared to C_{210}^{ds} and hence C_{210}^{ds} can be neglected.

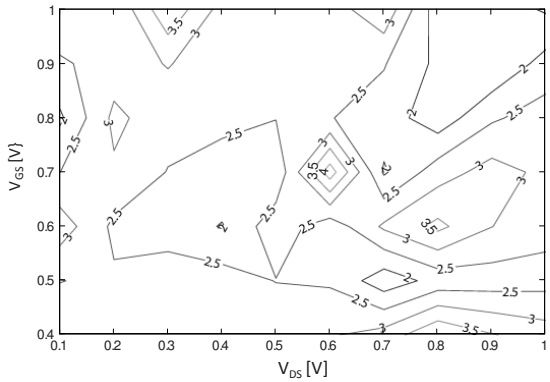


Fig. 2.17. Contour plot of $\lg(f_{210}^{ds}/1GHz)$ with different gate and drain bias.

Appendix 2.3

The drain-source resistive nonlinearity of M_1, the drain-source resistive nonlinearity of M_2, the gate-source capacitive nonlinearity of M_2 and the bulk-source capacitive nonlinearity of M2 contribute to the IM_3 of the cascode amplifier shown in Fig. 2.2. Applying the general weak nonlinearity model given in (2.3) to the cascode amplifier yields

$$v_{out}^{\omega_{IM3}} = \sum_K H_{M_1}^{ds}(\omega_{IM3}) \cdot \beta_{nml,M_1}^{\omega_D} G_{nml,M_1}^{ds} \tag{2.12}$$

$$+ \sum_K \beta_{nml,M_2}^{\omega_D} \cdot [\, H_{M2}^{ds}(\omega_{IM3}) \cdot G_{nml,M_2}^{ds} + H_{M_2}^{gs}(\omega_{IM3}) \cdot j\omega_{IM3} C_{nml,M_2}^{gs}$$

$$+ H_{M_2}^{bs}(\omega_{IM3}) \cdot j\omega_{IM3} C_{nml,M_2}^{bs}].$$

Using the model shown in Fig. 2.18a to calculate the H function, (2.12) can be rewritten into

$$v_{out}^{\omega_{IM3}} = \sum_K \frac{-g_m^{M_2} r_{ds}^{M_1} \cdot R_{load}}{1 + g_m^{M_2} r_{ds}^{M_1}} \cdot \beta_{nml,M_1}^{\omega_D} G_{nml,M_1}^{ds} \tag{2.13}$$

$$+ \sum_K \frac{-R_{load}}{1 + g_m^{M_2} r_{ds}^{M_1}} \cdot \beta_{nml,M_2}^{\omega_D} \cdot [\, G_{nml,M_2}^{ds} - g_m^{M_2} r_{ds}^{M_1} \cdot j\omega_{IM3} C_{nml,M_2}^{gs} - g_m^{M_2} r_{ds}^{M_1} \cdot j\omega_{IM3} C_{nml,M_2}^{bs}].$$

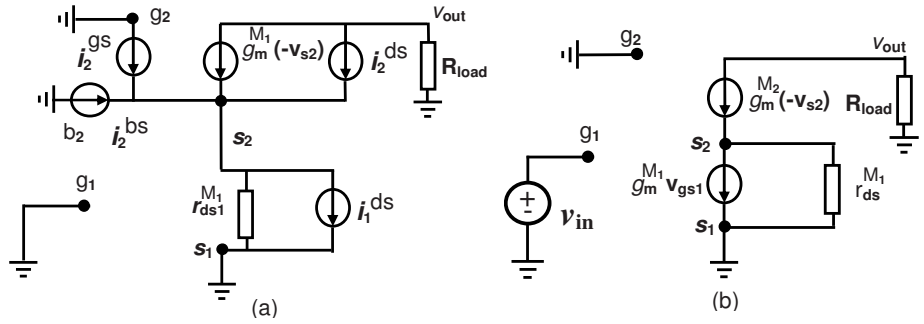


Fig. 2.18. (a) the equivalent model for calculating the H function. (b) the equivalent model for calculating the β functions.

For e.g. M_2, the relative importance between the gate-source capacitive nonlinearity, the bulk-source capacitive nonlinearity and the drain-source resistive nonlinearity can be determined by

$$\begin{aligned} f_{nml.M2}^{\frac{ds}{gs,bs}} &= G_{nml,M_2}^{ds} / [g_m^{M_2} r_{ds}^{M_1} \cdot \omega_{IM3} (C_{nml,M_2}^{gs} + C_{nml,M_2}^{bs})] \\ &\approx G_{nml,M_2}^{ds} / [10\omega_{IM3} \cdot (C_{nml,M_2}^{gs} + C_{nml,M_2}^{bs})] \end{aligned} \tag{2.14}$$

Characterization of $f_{nml.M2}^{ds/(gs,bs)}$ shows that for M_2 in the cascode amplifier the dominant nonlinearity is the drain-source resistive nonlinearity. Then (2.13) can be simplified to

$$v_{out}^{\omega_{IM3}} = \sum_K \frac{-g_m^{M_2} r_{ds}^{M_1} \cdot R_{load}}{1+g_m^{M_2} r_{ds}^{M_1}} \cdot \beta_{nml,M1}^{\omega_D} G_{nml,M1}^{ds} + \sum_K \frac{-R_{load}}{1+g_m^{M_2} r_{ds}^{M_1}} \cdot \beta_{nml,M2}^{\omega_D} \cdot G_{nml,M2}^{ds} \tag{2.15}$$

Firstly assuming that the drain-source resistive nonlinearity related to the bulk-source voltage swing can be neglected, and secondly only including the third-order nonlinearity, we use the model shown in Fig. 2.18(b) to calculate the β functions, yielding equation (2.5) in section 2.3.

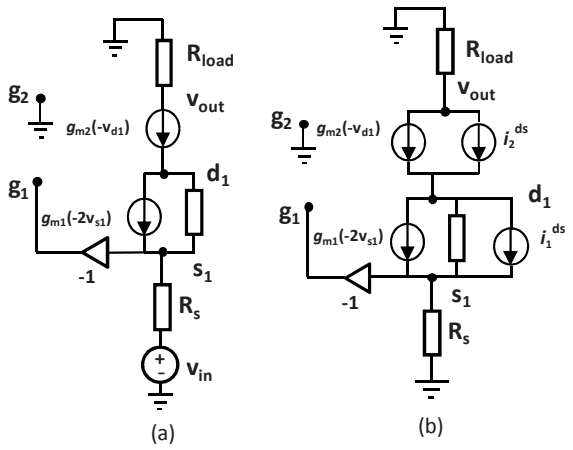


Fig. 2.19. (a) the equivalent model for calculating the β function and (b) the H functions in the CG LNA.

Appendix 2.4

For the CG LNA shown in Fig. 2.11(b), assuming firstly that the resistive nonlinearity is dominant between the drain-source terminal, and secondly including only the third-order nonlinearities, the general weak nonlinearity model given in (2.3) can be rewritten as

$$v_{out}^{\omega_{IM3}} = \sum_K H_{M_1}^{ds}(\omega_{IM3}) \cdot \beta_{nml,M_1}^{\omega_D} G_{nml,M_1}^{ds} + \sum_K H_{M_2}^{ds}(\omega_{IM3}) \cdot \beta_{nml,M_2}^{\omega_D} G_{nml,M_2}^{ds} \quad (2.16).$$

Assuming perfect input matching ($R_s = (g_{ds}^{M_1} + g_m^{M_2})/[g_m^{M_2}(g_{ds}^{M_1} + 2g_m^{M_1})]$), we use the model shown in Fig. 2.19 to calculate the H and β functions. Then (2.16) can be rewritten as equation (2.8) in section 2.4.

2.6 References

[1] H. Zhang, E. Sanchez-Sinencio "Linearization techniques for CMOS low noise amplifiers: a tutorial," *IEEE Trans. Circuits and Syst. I*, vol.58, no.1, pp.1215–1227, Jan. 2011.

[2] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*, Dordrecht, The Netherlands: Kluwer, 1998.

[3] P. Wambacq, G. Gielen, P. Kinget, and W. Sansen, "High-frequency distortion analysis of analog integrated circuits," *IEEE Trans. Circuits and Syst. II*, vol. 46, pp. 335–344, Mar. 1999.

[4] P. Dobrovolny, G. Vandersteen, P. Wambacq and S. Donnay, "Analysis and compact behavioral modeling of nonlinear distortion in analog communication circuits," *IEEE Trans. Computer-aided Design of Integrated Circuits and Systems*, vol.22, pp.1215–1227no.9, Sept. 2003.

[5] P. Li and L. T. Pileggi, "Compact reduced-order modeling of weakly nonlinear analog and RF circuits," *IEEE Trans. CAD Des. Integr. Circuits Syst.*, vol. 23, pp. no. 2, 184-203, 2005.

[6] W. Cheng, A. J. Annema, J. A. Croon, D. B. M. Klaasen and B. Nauta, "A general weak nonlinearity model for LNAs," *IEEE Custom Integrated Circuits Conference*, pp.221–224, Sept. 2008.

[7] S. El-Rabaie, V. F. Fusco and C. Stewart, "Harmonic balance evaluation of nonlinear microwave circuits-a Tutorial approach" *IEEE Trans. Education*, vol. 31, No.3, pp. 181–192, 1988.

[8] K. Sanghoon, B. Choi and B. Kim, "Linearity analysis of CMOS for RF application," *IEEE Trans. Microwave Theory and Techniques*, vol. 51, Issue 3, pp. 972–977, Mar. 2003.

[9] B. Toole, C. Plett, and M. Cloutier, "RF circuit implications of moderate inversion enhanced linear region in MOSFETs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 2, pp. 319–328, Feb. 2004.

[10] H. Khatri, P. S. Gudemand L.E.Larson, "Distortion in current commutating passive CMOS downconversion mixers," *IEEE Trans. Microwave Theory and Techniques*, vol. 57, no.11, pp. 2671–2681, Nov. 2009.

[11] X. Wei, G. Niu, Y. Li, M. Yang and S. S Taylor, "Modeling and characterization of intermodulation linearity on a 90-nm RF CMOS technology," *IEEE Trans. Microwave Theory and Techniques*, vol. 57, no.4, pp. 965–971, Apr. 2009.

[12] A. J. Scholten, G. D. J. Smit, B. A. de Vries, L. F. Tiemeijer, J. A. Croon, D. B. M. Klaassen, R. van Langevelde, X. Li, W. Wu and G. Gildenblat, "The New CMC Stand-

ard Compact MOS Model PSP: Advantages for RF Applications", *IEEE J. Solid State Circuits*, vol. 44, no. 5, may 2009, pp. 1415-1424

[13] X. Li, W. Wu, A. Jha, G. Gildenblat, R. van Langevelde, G.D.J. Smit, A.J. Scholten, D.B.M. Klaassen, C.C. McAndrew, J. Watts, C.M. Olsen, G.J. Coram, S. Chaudhry and J. Victory, "Benchmark Tests for MOSFET Compact Models With Application to the PSP Model", *IEEE tr. Electron Devices*, vol. 56, no. 2, Feb. 2009, pp. 243-251

[14] P. Bendix, P. Rakers, P. Wagh, L. Lemaitre, W. grabinski, C.C. McAndrew, X. Gu and G. Gildenblat, "RF Distortion Analysis with Compact MOSFET Models", *IEEE Custom Integrated Circuits Conference,* pp. 9-12， 2004

[15] C.C. McAndrew, "Validation of MOSFET model Source-Drain Symmetry", *IEEE tr. Electron Devices*, vol. 53, no. 9, Sept. 2009, pp. 2202-2206

[16] J. Borremans, L.D.Locht, P.Wambacq and Y.Rolain, "Nonlinearity Analysis of Analog/RF Circuits using Combined Multisine and Volterra Analysis," *IEEE DATE*, France, April 16-20, 2007.

[17] P. Li and L. T.Pileggi, "Efficient per-nonlinearity distortion analysis for analog and RF circuits," *IEEE Trans. CAD Des. Integr. Circuits Syst*., vol. 22, pp. 1297-1309, 2003.

[18] [Online]. Available: http://www.nxp.com/models/mos_models/psp/

[19] J. Borremans, P. Wambacq, C. Soens, Y. Rolain and M. Kuijk, "Low-area active-feedback low-noise amplifier design in scaled digital CMOS," *IEEE J. Solid-State Circuits,* vol. 43, no.11, pp. 2422–2433, Oct. 2008.

[20] B. Perumana, J. C. Zhan, S. S. Taylor, B. R. Carlton and J. Laskar, "Resistive-feedback CMOS low-noise amplifiers for multiband applications," *IEEE Trans. Microwave Theory and Techniques*, vol. 56, no.5, pp. 1218–1225, May 2008.

[21] J. Yoon, H. Kim, C. Park, J. Yang, H. Song, S. Lee and B. Kim, "A new RF CMOS Gilbert mixer with improved noise figure and linearity," *IEEE Trans. Microwave Theory and Techniques*, vol. 56, no.3, pp. 626–631, Mar. 2008.

[22] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. New York: Cambridge Univ. Press, 2004.

[23] B. Razavi, *Design of Analog CMOS Integrated Circuits*, New York: McGraw-Hill Companies Inc., 2001.

[24] N. Kim, V. Aparin, K. Barnett and C.Persico, "A cellular-Band CDMA 0.25-um CMOS LNA linearized using active post-distortion," *IEEE J. Solid-State Circuits*, vol. 41, No. 7, pp. 1530–1534, July 2006.

[25] S. Ganesan, E. Sanchez-Sinencio and J. Silva-Martinez, "A highly linear low noise amplifier," *IEEE Trans. Microwave Theory and Techniques*, vol. 54, no.12, pp. 4079–4085, Dec. 2006.

[26] H. Zhang, X. Fan and E. Sanchez-Sinencio, "A low-power, linearized, ultra-wideband LNA design technique," *IEEE J. Solid-State Circuits*, vol. 44, No. 2, pp. 320–330, Feb. 2009.

[27] T. Kim and B. Kim, "Linearization of differential CMOS low noise amplifier using cross-coupled post distortion canceller," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, pp.83–86, 2008.*

[28] J. Kuo, Z. Tsai, K. Lin and H. Wang, "Design and analysis of novel linearization technique of cascode cell in a 60-GHz CMOS demodulator," *IEEE Trans. Microwave Theory and Techniques*, vol. 59, no.2, pp. 456– 465, Feb. 2011.

[29] B. Razavi, "Design considerations for future RF circuits", *IEEE Internationl Symposium on Circuits and Systems*, pp.747–744, 2007.

[30] X. Fan, H. Zhang and E.Sanchez-Sinencio, "A noise reduction and linearity improvement technique for a differential cascode LNA," *IEEE J. Solid-State Circuits,* vol. 43, no.3, pp. 588–599, Mar. 2008.

[31] C. Cui, T. S. Kim, S. K. Kim, J. K. Cho, S. T. Kim and B.S. Kim, "Effects of the nonlinearity of the common-gate stage on the linearity of CMOS cascode low noise amplifier," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, 2011.*

[32] J. Borremans, P. Wambacq, C. Soens, Y. Rolain and M. Kuijk, "Low-area active-feedback low-noise amplifier design in scaled digital CMOS," *IEEE J. Solid-State Circuits,* vol. 43, no.11, pp. 2422–2433, Nov. 2008.

[33] C. Tang, C. Wu and S. Liu, "Miniature 3-D inductors in standard CMOS process," *IEEE J. Solid-State Circuits,* vol. 37, no.4, pp. 471–480, Apr. 2002.

[34] W. Zhuo, S. H. K. Embabi, J. Pineda de Gyvez, and E. Sanchez-Sinencio, "Using capacitive cross-cupling technique in RF low-noise amplifiers and down-conversion mixer design," *IEEE Eur. Solid-State Circuits Conf. (ESSCIRC),* Sep.2000, pp.116-119.

[35] W. Zhuo, X. Li, S. Shekhar, S. H. K. Embabi, J. Pineda de Gyvez, D.J.Allstot and E. Sanchez-Sinencio, "A capacitor cross-coupled common-gate low noise amplifier," *IEEE Trans. Circuits and Syst. II*, vol. 52, no. 12, pp. 875–879, Dec. 2005.

[36] S. Woo, W. Kim, Chang-Ho Lee and K.Lim, "A 3.6mW differential common-gate CMOS LNA with positive-negative feedack," *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers,* Feb. 2009, pp. 218-219.

[37] J. Kim, S. Hoyos and J. Silva-Martinez, "Wideband common-gate CMOS LNA employing dual negative feedback with simultaneous noise, gain and bandwidth optimization," *IEEE Trans. Microwave Theory and Techniques*, vol. 58, no.9, pp. 2340–2351, Sept. 2010.

[38] J. Borremans, G. Mandal, B. Debaillie, V. Giannini and J. Craninckx, "A sub-3dB NF voltage-sampling front-end with +18dBm IIP3 and +2dBm blocker compression point," in *Proc. IEEE Eur. Solid-State Circuits Conf. (ESSCIRC),* Sep.2010, pp.402-405.

[39] J. Borremans, G. Mandal, V. Giannini, T. Sano, M. Ingels, B. Verbruggen and J.Craninckx, "A 40nm CMOS highly linear 0.4-to-6GHz receiver resilient to 0dBm out-of-band blockers," *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers,* Feb. 2011, pp. 62-63.

[40] W. Cheng, M. S. Oude Alink, A. J. Annema, J. A. Croon and B. Nauta, "RF circuit linearity optimization using a general weak nonlinearity model," accepted to *IEEE Trans. Circuits and Systems I*, 2012.

Chapter 3

IM3 cancellation technique for LNAs with cascode topology

As discussed in chapter 2, the cascode transistor may significantly generate distortion in modern CMOS. Biasing transistors in the moderate inversion region provides one solution with the price of less bandwidth. This chapter presents a novel technique that using a negative impedance to enable distortion cancellation between the transconductor and the cascode transistor for LNAs with a cascode topology. As a proof of concept, a resistive feedback LNA using this IM3 cancellation technique in a standard 0.16μm CMOS process shows that for 0.1GHz to 1GHz, improvements of 6.3dB to 10dB for IIP3 and 0.2dB to 1dB for gain are achieved without noise degradation. The power consumption for the LNA is increased by 2%, and the die area by about 700μm^2.

3.1 **Introduction**

Much effort has been put in improving the linearity of LNAs. As a commonly used technique, an auxiliary path replicating the distortion of the

This chapter was accepted to IEEE Radio Frequency Integrated Circuits Symposium (RFIC) in 2012 [9].

main path is combined with the main LNA in a subtracting node. The auxiliary path can be transistors biased in weak inversion [1-2], transistors biased in saturation region as a nonlinear resistor [3] or current sources that injects IM2 to suppress IM3 [4]. The limitations of most reported linearization techniques mainly are due to:

- only focusing on transconductance nonlinearity, neglecting nonlinearity related to output conductance [5].
- neglecting the distortion generated by the cascode transistor.

Neglecting output conductance related distortion and neglecting the distortion contribution of cascode transistors is valid for older technologies with long-channel devices and high supply voltage. However, in deep submicron CMOS technologies, typically the output resistance of the transconductor stage is relatively low: the IM3 distortion contribution from the cascode transistor then may become dominant. On top of that, the low supply voltage together with high gain operation tends to push the cascode transistor out of the deep saturation region. This results in a very significant increase of the third-order output conductance nonlinearity term and the cross-modulation nonlinearities, which causes the increase of distortion generated by the cascode transistor [7].

In this chapter, we present a wideband IM3 cancellation technique that takes into account the distortion of the cascode transistor and all the third-order nonlinearity contributions related to the transconductance and the output conductance. A negative impedance is used to enable cancellation between the distortion current of the transconductor and the cascode transistor. Without loss of generality, we apply this IM3 cancellation technique to a resistive feedback wideband LNA. Section 3.2 presents an analysis to explain the proposed IM3 cancellation and presents a short discussion on the effect on gain and noise. Section 3.3 shows both simulation and measurement results to verify the theory.

3.2 Theory of IM3 Cancellation using Negative Impedance

In the resistive feedback LNA shown in Fig. 3.1, M_{1a}/M_{1b} is the transconductor stage while the cascode transistor M_{2a}/M_{2b} increases the output

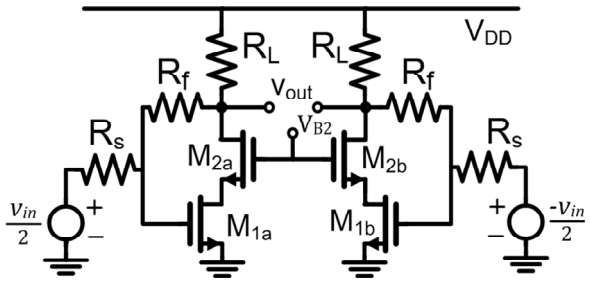


Fig. 3.1. Schematic of a conventional differential resistive feedback LNA.

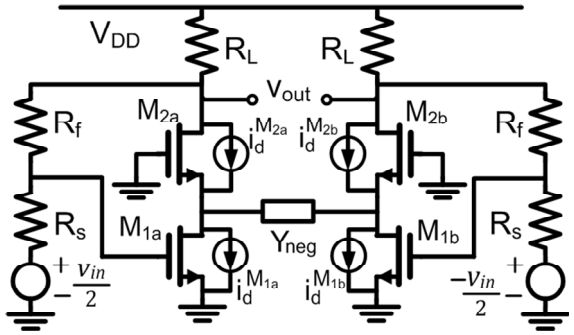


Fig. 3.2. Circuit model for the distortion and gain analysis .

impedance and improves the isolation between input and output. The shunt feedback resistor R_f is used to match to the source resistance R_s.

To improve the linearity, we apply a negative impedance ($Y_{neg} = G_{neg} + j\omega C_{neg}$) between the drain of M_{1a}/M_{1b} as shown in Fig. 3.2. The IM3 distortion current of transistor M_x is due to the voltage swings at the transistor terminals, where we neglect the effect of the source-bulk signal for simplicity reasons only. This results in a 3rd order current $i_d^{Mx} = g_{m3}v_{gs}^3 + g_{x21}v_{gs}^2 v_{ds} + g_{x12}v_{gs}v_{ds}^2 + g_{d3}v_{ds}^3$, for which a short hand notation $i_d^{Mx} = f_d^{Mx}(v_{gs}^{Mx}, v_{ds}^{Mx})$ will be used in the remainder of this chapter. The exact function that describes the transistor nonlinearity is

not relevant in this chapter. Given the differential circuit topology, we assume $i_d^{M_{1a}} = -i_d^{M_{1b}} = i_d^{M_1}$ and $i_d^{M_{2a}} = -i_d^{M_{2b}} = i_d^{M_2}$.

Using the general weakly nonlinearity model presented in Chapter 2, the closed-form solution for LNA's IM3 is given by

$$v_{out}^{IM3} = -\frac{2(R_f + R_s)R_L}{D_v}\left[g_m^{M_2} i_d^{M_1} + \left(Y_{ds}^{M_1} + 2Y_{neg}\right) i_d^{M_2}\right] \quad (3.1).$$

$$\text{where } D_v = g_m^{M_2}\left(R_f + R_L + R_s + g_m^{M_1} R_L R_s\right) + \left(R_f + R_L + R_s\right)\left(Y_{ds}^{M_1} + 2Y_{neg}\right)$$

In (3.1), the 3rd order distortion currents $i_d^{M_1}$ and $i_d^{M_2}$ are according to the previously introduced $i_d^{M_x} = f_d^{M_x}(v_{gs}^{M_x}, v_{ds}^{M_x})$, which in this circuit expands to:

$$i_d^{M_1} = f_d^{M_1}\left(v_{gs}^{M_{1a}}, v_{ds}^{M_{1a}}\right) = f_d^{M_1}\left[\frac{\left(R_f + R_L\right)\left(g_m^{M_2} + Y_{ds}^{M_1} + 2Y_{neg}\right)v_{in}}{2D_v}, \frac{-g_m^{M_1}\left(R_f + R_L\right)v_{in}}{2D_v}\right] \quad (3.2)$$

$$i_d^{M_2} = f_d^{M_2}\left(v_{gs}^{M_{2a}}, v_{ds}^{M_{2a}}\right) = f_d^{M_2}\left[\frac{g_m^{M_1}\left(R_f + R_L\right)v_{in}}{2D_v}, \frac{g_m^{M_1}\left(R_f + R_L - g_m^{M_2} R_f R_L\right) + R_L\left(g_m^{M_2} + Y_{ds}^{M_1} + 2Y_{neg}\right)}{2D_v}\right] \quad (3.3)$$

It now follows from (3.2-3.3) that for $Y_{neg} > -0.5(g_m^{M_2} + Y_{ds}^{M_1})$ the polarity of the voltage swings for M_{1a}/M_{1b} and M_{2a}/M_{2b} are unchanged, and hence so is the phase of the IM3 currents of the transistors. At the same time, (3.1) shows that for $Y_{neg} < -0.5Y_{ds}^{M_1}$ the distortion components of the transistors are subtracted. Combining these results yields a region defined by :

$$G_{neg} \in \left[\frac{-g_{ds}^{M_1} - g_m^{M_2}}{2}, \frac{-g_{ds}^{M_1}}{2}\right] \text{ and } C_{neg} \approx -0.5C_{ds}^{M_1} \quad (3.4)$$

in which IM3 contributions of M_1 and M_2 (partially or fully) cancel. Note that doing so, the distortion of the transconductor transistors is canceled using the distortion contribution of the cascode transistors.

The small signal voltage gain can be derived from the model in Fig. 3.2, resulting in

$$A_V = \frac{R_L\left[g_m^{M_2}(1 - g_m^{M_1} R_f) + Y_{ds}^{M_1} + 2Y_{neg}\right]}{D_v} \approx \frac{R_L(-g_m^{M_1} g_m^{M_2} R_f + Y_{ds}^{M_1} + 2Y_{neg})}{D_v} \quad (3.5).$$

Since Y_{neg} increases the overall output impedance of the transconductor M_{1a}/M_{1b}, a higher gain is resulted. The noise figure (NF) is calculated by the

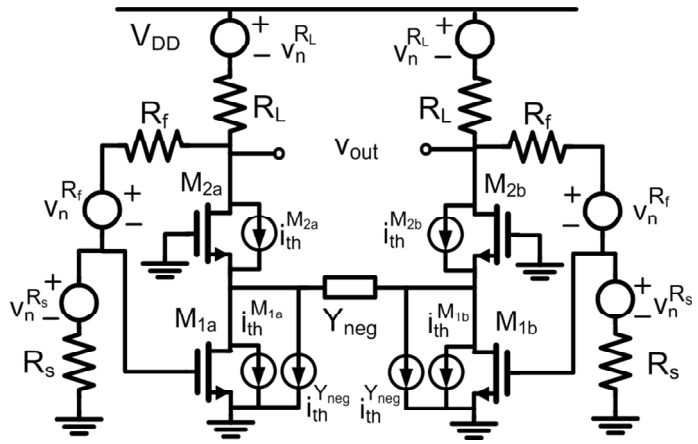


Fig. 3.3. Circuit model for calculating noise.

model shown in Fig. 3.3, where we include the thermal noise current of M_{1a}/M_{1b}, M_{2a}/M_{2b} and Y_{neg} ($i_{th}^{Y_{neg}} = |4kTG_{neg}|$) and the thermal noise voltage of R_f, R_s and R_L. The calculated NF is given by

$$NF = 1 + \frac{\left[g_m^{M_2} R_L (R_f + R_s)\right]^2 \gamma g_m^{M_1}}{R_s D_n^2} + \frac{\left[R_L\left(g_m^{M_2} + g_m^{M_1} g_m^{M_2} R_s + Y_{ds}^{M_1} + 2Y_{neg}\right)\right]^2 R_f}{R_s D_n^2} + \frac{\left[(R_f + R_s)\left(g_m^{M_2} + Y_{ds}^{M_1} + 2Y_{neg}\right)\right]^2 R_L}{R_s D_n^2} + \frac{\left[R_L (R_f + R_s)\left(Y_{ds}^{M_1} + 2Y_{neg}\right)\right]^2 \gamma g_m^{M_2}}{R_s D_n^2} + \frac{\left[g_m^{M_2} R_L (R_f + R_s)\right]^2 |G_{neg}|}{R_s D_n^2} \quad (3.6),$$

where $D_n = g_m^{M_2}(1 - g_m^{M_1} R_f) + Y_{ds}^{M_1} + 2Y_{neg} \approx -g_m^{M_1} g_m^{M_2} R_f + Y_{ds}^{M_1} + 2Y_{neg}$

where the five terms account for the thermal noise from respectively M_{1a}/M_{1b}, R_f, R_L, M_{2a}/M_{2b} and Y_{neg}. Compared to the situation without an Y_{neg}, when the IM3 cancellation condition (3.4) is met, the denominator D_n^2 is increased and the nominator of the second term and the third term are decreased. This results in decreases of NF contribution from M_{1a}/M_{1b}, R_f and R_L. Since Y_{neg} increases the overall output impedance of the transconductor M_{1a}/M_{1b}, less noise contribution can come from the cascode transistor M_{2a}/M_{2b}. Therefore, it can be concluded that although a Y_{neg} circuit introduces extra noise (last term in (3.6)), it also reduces the NF contribution from M_{1a}/M_{1b}, M_{2a}/M_{2b}, R_f and R_L. As a result, the effect on NF by a Y_{neg} circuit can be small.

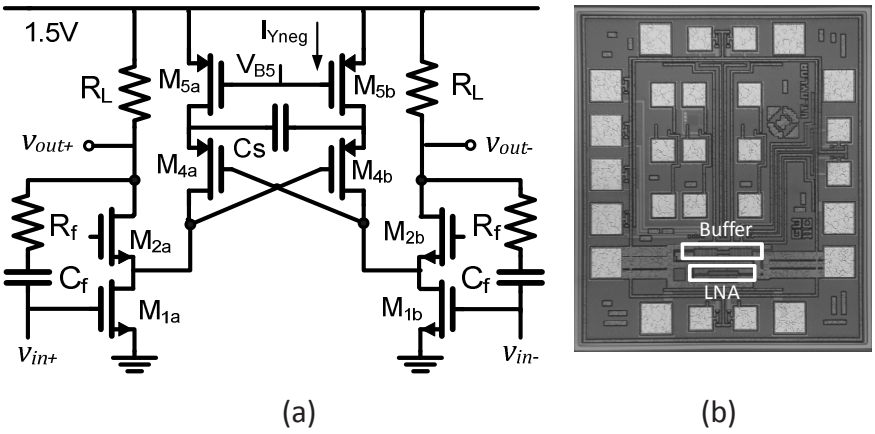


Fig. 3.4. LNA using Y_{neg} for IM3 cancellation. (a) Schematic. (b) Microphotograph of the fabricated chip.

3.3 **LNA design and experimental results**

To prove this IM3 cancellation concept, the LNA in Fig. 3.4 is implemented in a standard 0.16µm CMOS process. The negative impedance is implemented by the cross-coupled pair M_{4a}/M_{4b} with source degeneration provided by capacitor C_s (1.6pF) and current source M_{5a}/M_{5b} [8]. The capacitor C_f (1.57pF) and R_f (370Ω) provide the shunt feedback. The Y_{neg} circuit provides an almost-constant negative resistance and a frequency dependent negative capacitance (decreasing with frequency).

We designed the Y_{neg} circuit for full IM3 cancellation at 1GHz. A buffer and a resistive attenuator are put in parallel on-chip after the LNA for noise/gain and IIP3 measurement. The IIP3 is extrapolated from -30dBm to -20dBm. The chip microphotograph is shown in Fig. 3.4b. The LNA occupies 0.00295mm^2 active area, of which 25% is taken by the Y_{neg} circuit. Packaged chips were measured on PCB boards. Two off-chip baluns were used at the input and output of the chip for single-ended-to-differential conversion. By switching on/off of the Y_{neg} circuit, we measure the effect of Y_{neg} on LNA

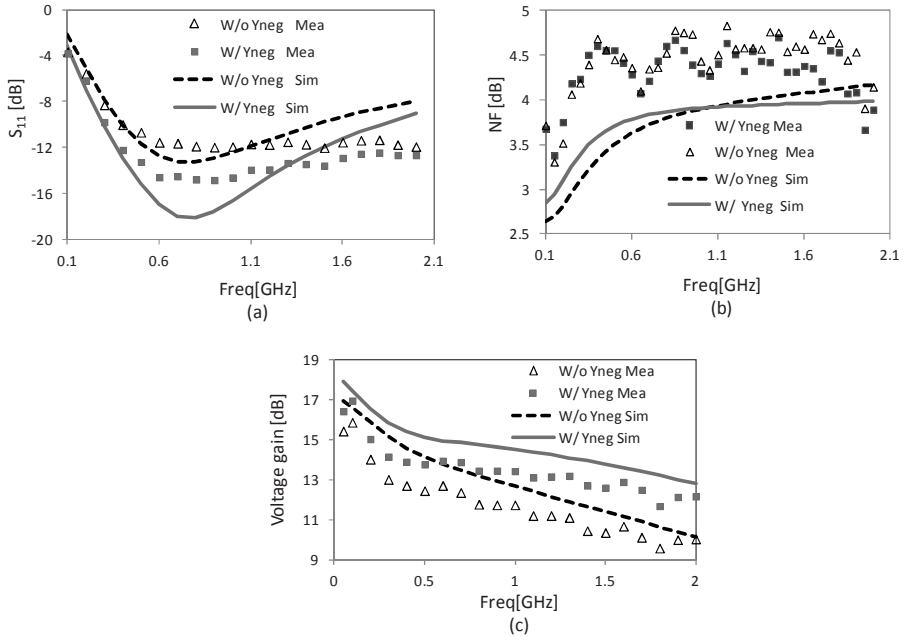


Fig. 3.5. Measured and simulated (a) S_{11}, (b) NF and (c) Voltage gain.

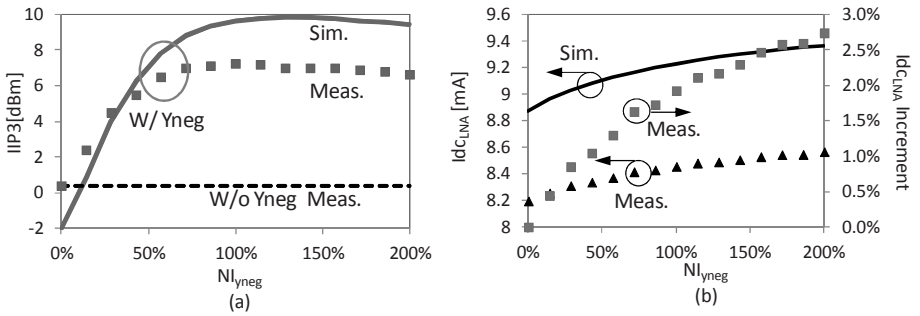


Fig. 3.6. Measured and simulated (a) IIP3 at 1GHz and (b) DC current consumption of LNA as a function of the nominated bias current of Y_{neg}.

performance. The measured and simulated S_{11}, NF and voltage gain is shown in Fig. 3.5.

For 0.1GHz to 1GHz, the Y_{neg} circuit introduces no degradation on NF, while improves S_{11} by 1-3dB and improves gain by 0.2-1dB. Below 0.3GHz S_{11} becomes >-10dB because the impedance of C_f starts to block the shunt feedback. This can be improved by using a larger C_f. Note that this LNA is

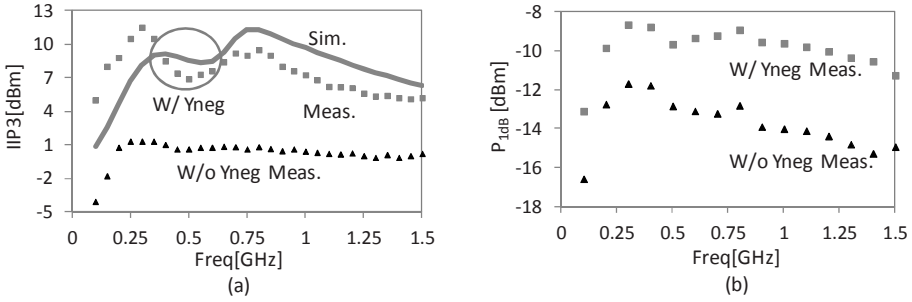


Fig. 3.7. Measured (a) IIP3 and (b) input P_{1dB} as a function of RF frequency.

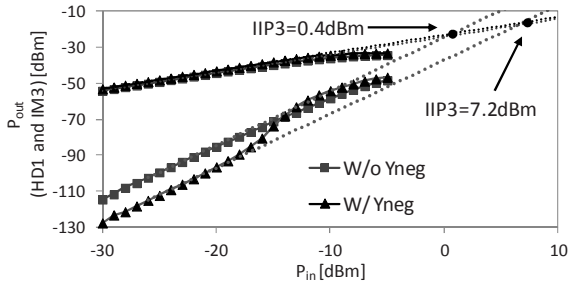


Fig. 3.8. Measured HD1 and IM3 at 1GHz as a function of input power.

not optimized for very low NF as we only focus on demonstrating the IM3 cancellation technique.

To verify the robustness against process spread, the bias current of Y_{neg} (I_{Yneg}) is swept within $\pm$100% variation of the optimal value. Fig. 3.6a shows the IIP3 improvement with respect to the circuit without Y_{neg} as a function of I_{Yneg} normalized to the optimum $I_{Yneg,opt}$; this parameter is denoted as NI_{yneg}. In the optimum setting hence NI_{yneg}=100%. Fig. 3.6a shows that for wide bias variation (NI_{yneg}=-30% to NI_{yneg}=+100%), +6dB IIP3 improvement is achieved at 1GHz. The power overhead of this technique is depicted in Fig. 3.6b.

Fig. 3.7a shows the frequency dependence of the IM3 cancellation technique, for NI_{yneg}=100%, on both IIP3 and on P_{1dB}. The measurements and simulation results in Fig. 3.7 show a weak frequency dependence in the IIP3 improvement and hence quite good robustness. Fig. 3.8 shows the measured HD1 and IM3 output at the optimal bias value of Y_{neg} (NI_{Yneg}=100%). The

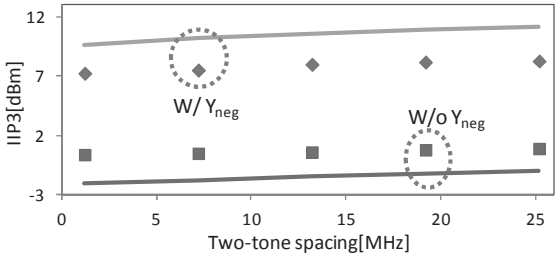


Fig. 3.9. Measured IIP3 at 1GHz as a function of two-tone spacing. Line for simulations and symbol for measurements.

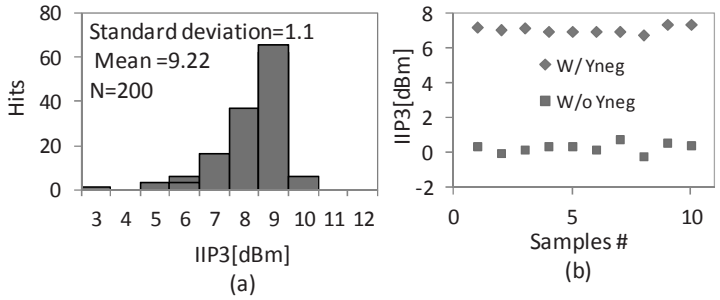


Fig. 3.10. Effect of mismatches and process spread on IIP3 at 1GHz. (a) 200-time Monte Carlo simulation results of IIP3 with Y_{neg} (NI_{yneg}=100%) and (b) measured IIP3 of ten dies for LNA with and without Y_{neg}.

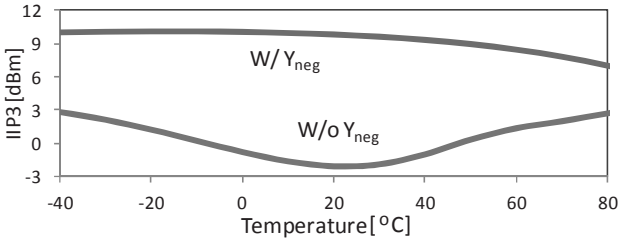


Fig. 3.11. Simulated IIP3 as a function of temperature for LNA with and without Y_{neg}.

IM3 curve starts to show 5th order behavior for Pin>-18dBm due to the transistors' higher-order nonlinearities that kick in at high input magnitudes. Fig. 3.9 presents IIP3 simulations results and IIP3 measurements as a func-

tion of the two-tone spacing, showing that the IM3 cancellation technique is not sensitive to two-tone spacing.

To estimate the overall effect of process spread and mismatch on this IM3 cancellation, a 200-time Monte-Carlo simulation is performed for an RF signal at 1GHz. Fig. 3.10a shows that the mean IIP3 is 9.2dBm at the optimal bias value of Y_{neg} (NI_{Yneg}=100%) (the nominal value is 9.6dBm), which is 9dB higher than the LNA without Y_{neg}. This shows good robustness of this IM3 cancellation technique. The measurement results of ten LNA samples shows +6.2dB IIP3 improvement at 1GHz as shown in Fig.3.10b. The simulation results in Fig. 3.11 show that the LNA with the Y_{neg} circuit provides a constant IIP3 from -40°C to 40°C and starts to decreases as the temperature higher than 40°C.

3.4 **Conclusion**

This chapter presents a wideband IM3 cancellation technique using a negative impedance, applied to a wide band cascode LNA. Using a suitable negative impedance, the distortion current generated by the cascode transistor cancels the distortion from the transconductor. The negative impedance also increases gain while its effect on NF can be minimal. For a resistive feedback LNA fabricated in a standard 0.16µm CMOS process, for 0.1GHz to 1GHz this IM3 cancellation technique improves IIP3 by 6.3dB to 10dB, gain by 0.2 to 1dB and P_{1dB} by +3dB while NF is not degraded, at a low area and power penalty. Robustness of this cancellation technique is demonstrated both in simulation and in measurements.

3.5 Reference

[1] T. W. Kim, "A common-gate amplifier with transconductance nonlinearity cancellation and its high-frequency analysis using the Volterra series," *IEEE Trans. Microw. Theory Tech.,*vol.57, no. 6, pp.1461-1469, 2009.

[2] T. -S. Kim and B. -S. Kim, "Linearization of differential CMOS low noise amplifier using cross-coupled post distortion canceller" *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium,* pp.83-86, 2008.

[3] H. Zhang, X. Fan and E. Sanchez-Sinencio, "A low-power linearized ultra-wideband LNA design technique," *IEEE J. Solid State Circuits,* no. 2, pp.320-330, 2009.

[4] S. Lou and H. C. Luong, "A linearization technique for RF receiver front-end using second-order-intermodulation injection" *IEEE J. Solid State Circuits,* vol. 43, no. 11, pp.2404-2412, Nov. 2008.

[6] H. Zhang and E. Sanchez-Sinencio, "Linearization techniques for CMOS low noise amplifiers: A Tutorial," *IEEE Trans. Circuits and Systems I,* vol. 58, no. 1, pp.22-36, Jan. 2011.

[7] W. Cheng, A.J.Annema, J.A.Croon and B.Nauta, "Noise and nonlinearity modeling of active mixers for fast and accurate estimation" *IEEE Trans. Circuits and Systems I,* no. 2, pp.276-289, Feb. 2011.

[8] C. Tilhac, S.Razafimandimby, A.Cathelin, S.Bila, V.Madrangeas and D.Belot, "A Tunable bandpass BAE-filter architecture using negative capacitance circuitry," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium,* pp.605-608, 2008.

[9] W. Cheng, A. J. Annema, G. J. M. Wienk and B. Nauta, "A wideband IM3 cancellation technique using negative impedance for LNA with cascode topology," accepted to *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, 2012.

Chapter 4

A wideband IM3 cancellation technique for CMOS Π and T attenuators

Based on the general distortion model presented in chapter 2, a novel technique that improves the linearity for both CMOS Π- and T-attenuators without using large devices is presented in this chapter. With proper transistor width ratios, the dominant distortion currents of transistor switches cancel each other. As a result, a high IIP3 robust to PVT variations can be achieved when using relative small transistors. The prototype Π-attenuator system with four discrete attenuation settings provides +26dBm IIP3 and +3dB P_{1dB} for 50MHz to 5GHz while only using 0.0054mm^2 active area in a 0.16µm standard bulk CMOS process. The prototype T-attenuator system with four discrete attenuation settings provides >+27dBm IIP3 and >13dB P_{1dB} for 50MHz to 5.6GHz while only using 0.0067mm^2 active area in a 0.16µm standard bulk CMOS process.

This chapter has been submitted to IEEE Journal of Solid-State Circuits in 2012.

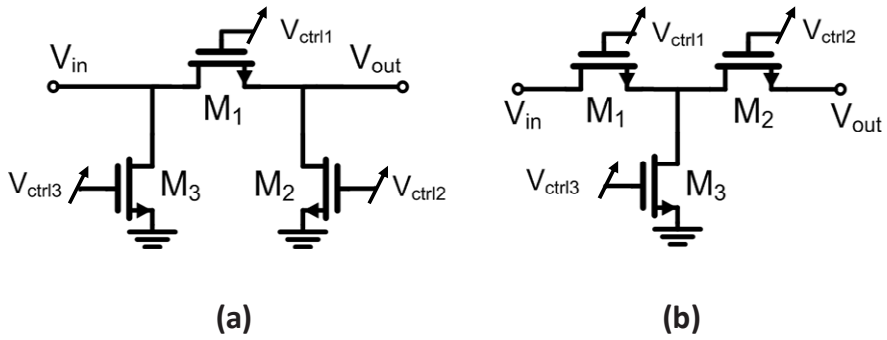


Fig. 4.1. Schematic of (a) Π-attenuator and (b) T-attenuator.

4.1 **Introduction**

In the receiver path and in spectrum analyzers [1] typically gain control blocks are used to limit the incident power to a level that the receiver circuitry can handle without degrading the linearity; in the transmitter path stringent power control is also desirable in a pre-distortion or correction loop before a power amplifier [2]. Traditionally, variable-gain amplifiers (VGAs) implement the gain control block, while attenuators based on FET transistors show superior performances for linearity, power handling capability and power consumption [2-5] when only signal attenuation is required.

The Π-attenuator and T-attenuator shown in Fig. 4.1 are widely-used gain-control elements [4-6], using three transistors as voltage-controlled resistors. By changing the gate voltages of transistors between V_{SS} to V_{DD}, continuously-controlled signal attenuation level (linear-in-dB controllability) and input/output matching are achieved [4]. In these attenuators, mainly the voltage swings across the transistor in combination with the transistors'nonlinarities generate distortion and limit the linearity of attenuators [6].

Much efforthas been devoted to improving the linearity and power handling capability of continuously-tuning attenuators [2-4] and discrete-step attenuators [5]. Adaptive bootstrapped body biasing [2] is used in a cascaded Π-attenuator to suppress the body-related parasitic effects and to improve

P_{1dB} (1dB compression point). The stacked-FET technique used in [3] distributes the signal swing among many FETs in series to reduce the drain-source voltage swing for each FET and hence reduce the IM3 distortion. However, the large transistors required by this technique bring in large parasitic capacitances, which lower the bandwidth and increase the minimum insertion loss (IL) at high frequencies. Moreover, the capacitive nonlinearities introduced by large parasitic capacitances will limit the highest achievable IIP3. Therefore, this technique is mainly effective in SOI CMOS [3]. A two-stage cascaded T-attenuator (in [6], T-attenuators are shown to be more linear than Π-attenuators) is used to improve the linearity especially at higher attenuation settings, resulting in an IIP3 of +20dBm at mid-attenuation [4].

To reduce the distortion generated by the transistors' output resistance nonlinearity, transistors can be used as switches rather than voltage-controlled resistors, with passive resistors (which are usually more linear) providing the signal attenuation. Such a system implies discrete-step attenuation settings instead of linear-in-dB controllability. Moreover, large transistors are usually desirable to reduce switch-on resistance and minimize distortion generated by the switches. This results in less bandwidth and larger active area. In [5] a Π-attenuator with parallel branches is designed for discrete attenuation steps achieves +23dBm IIP3 in the TV band

In [7] we presented a wideband IM3 cancellation technique for discrete-step Π-attenuators in bulk CMOS that alleviates the tradeoff between IIP3 and size (and thus bandwidth). This technique relies on canceling the distortion currents of series and shunt transistor switches, enabling highly linear attenuators without large transistor switches. In [8] similar IM3 cancellation was shown in a voltage divider (simulations only). In this paper, we elaborate on the wideband IM3 cancellation technique and show that it can be applied to both Π-attenuators and T-attenuators. In section 4.2, the analysis of the proposed IM3 cancellation technique is given for both the Π-attenuators and for the T-attenuators. Section 4.3 discusses the effect of parasitic, nonlinear capacitance and PVT variations on this IM3 cancellation technique. Section 4.4 shows simulation and measurement results. The conclusions are summarized in section 4.5.

4.2 Attenuator distortion analysis

For the Π-attenuator with continuous attenuation settings, shown in Fig. 4.1a, the attenuation is achieved mainly by increasing the resistance of the series device M_1. Simultaneously, the control voltage of the shunt devices M_2/M_3 adjusts the channel resistance for input/output matching. Similarly, for the T-attenuator with continuous attenuation settings, shown in Fig. 4.1b, the attenuation is achieved mainly by decreasing the resistance of the shunt device M_3 while series devices M_1/M_2 together M_3 provides input/output matching. For the Π-attenuator at high attenuation settings, the channel resistance of M_1 is large so that a large part of input signal drops across M_1. As a result, the nonlinear channel resistance of M_1 generates relatively high level of distortion. In T attenuator, the channel resistance of M_3 is set small to short the signal to ground, resulting less distortion by M_3. Consequently, T-attenuators in general are more linear than the Π-attenuators, especially at higher attenuation values [6]. Nevertheless, for discrete-step attenuators, in this section we show that by properly sizing the switches the linearity of Π-attenuator and T-attenuator can be improved to a similar level.

4.2.1 Π-Attenuator

For the Π-attenuator shown in Fig. 4.2, the input power source is modeled as a voltage source $v_s = 2v_{in}$ (V_{IN} is the magnitude of v_{in}) with source impedance R_s. Assuming perfect input matching provided by the attenuator, the input voltage for the attenuator is v_{in} and the gain (attenuation is $1/A$) is defined by $A = v_{out}/v_{in}$. Large resistors in series with gate and bulk of M_1 force the gate and bulk voltages to follow the source/drain voltage: with sufficiently large resistors these voltages are purely AC-coupled [5], resulting in $v_{gs}^{M_1} \cong v_{bs}^{M_1} \cong v_{ds}^{M_1}/2$. These bootstrapping resistors extend the bandwidth of the attenuator, and minimize the distortion caused by all nonlinearities related to $v_{gs}^{M_1}$ and $v_{bs}^{M_1}$ [5]. For M_2 and M_3, these bootstrapping resistors are not used since they have negligible impact on all relevant performance parameters (e.g. linearity and bandwidth), as indicated in simulations. All transistors are using minimum length for maximum bandwidth.

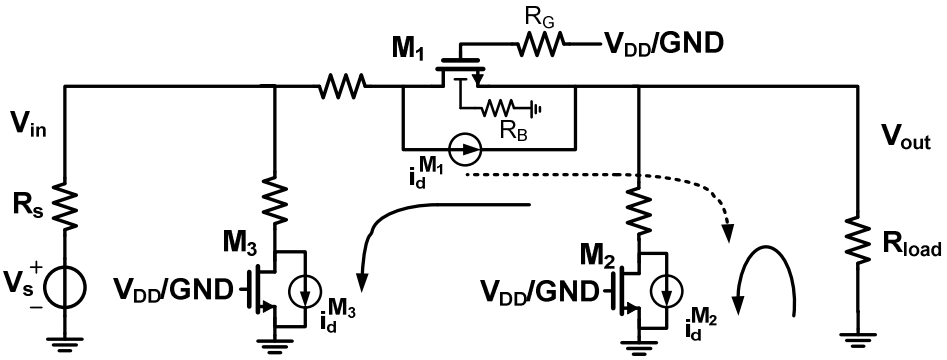


Fig. 4.2. Illustration of IM3 cancellation principle in a Π-attenuator.

As a first-order approximation, we assume the distortion current between the drain and source of transistors are dominant (direction defined from drain to source) [6]. Applying the general nonlinearity model given in [8] to the Π-attenuator shown in Fig. 4.2 and only including the third-order nonlinearity, it can be derived that (see Appendix I for the derivation):

$$v_{out}^{\omega_{IM3}} \approx \frac{3 \times {V_{IN}}^3}{8{R_s}^3(1+A)^4} \times \left[A^4\left(r_{on}^{M_1}\right)^4\left(16Y_{030,M_1}^{ds} + 8Y_{120,M_1}^{ds} + 4Y_{210,M_1}^{ds} + 2Y_{300,M_1}^{ds}\right) - (1-A)^4A^3\left(r_{on}^{M_2}\right)^4 Y_{030,M_2}^{ds} - (1-A)^4A\left(r_{on}^{M_3}\right)^4 Y_{030,M_3}^{ds}\right] \quad (4.1)$$

where Y_{nml}^{ds} is the transistor nonlinear admittance between drain and source [9], defined as $Y_{nml}^{ds} = G_{nml}^{ds} + j\omega_{IM3}C_{nml}^{ds}$

with $G_{nml}^{ds} = \frac{1}{n!}\frac{1}{m!}\frac{1}{l!}\frac{\partial^{(n+m+l)}I_{ds}}{\partial V_{gs}^n \partial V_{ds}^m \partial V_{bs}^l}\Big|_{\substack{V_{gs}=V_{GS}\\V_{ds}=V_{DS}\\V_{bs}=V_{BS}}}$ and $C_{nml}^{ds} = \frac{1}{n!}\frac{1}{m!}\frac{1}{l!}\frac{\partial^{(n+m+l)}Q_d}{\partial V_{gs}^n \partial V_{ds}^m \partial V_{bs}^l}\Big|_{\substack{V_{gs}=V_{GS}\\V_{ds}=V_{DS}\\V_{bs}=V_{BS}}}$

The $r_{on} = 1/G_{010}^{ds}$ is the small-signal switch-on resistance; G_{030}^{ds} is the third-order output conductance nonlinearity, G_{300}^{ds} is the third-order transconductance nonlinearity, and G_{210}^{ds} and G_{120}^{ds} are the cross-modulation

nonlinearities; C_{030}^{ds}, C_{300}^{ds}, C_{210}^{ds} and C_{120}^{ds} are their capacitive counterparts. As the switched-on transistors stay in the deep triode region, the third-order output admittance nonlinearity is dominant [8]. This allows for simplification of (4.1) into:

$$v_{out}^{\omega_{IM3}} \approx \frac{3 \times {V_{IN}}^3}{8{R_s}^3(1+A)^4} \times \Big[16A^4\big(r_{on}^{M_1}\big)^4\big(G_{030,M_1}^{ds} + j\omega_{IM3}C_{030,M_1}^{ds}\big) - (1-A)^4A^3\big(r_{on}^{M_2}\big)^4\big(G_{030,M_2}^{ds} + j\omega_{IM3}C_{030,M_2}^{ds}\big) - (1-A)^4A\big(r_{on}^{M_3}\big)^4\big(G_{030,M_3}^{ds} + j\omega_{IM3}C_{030,M_3}^{ds}\big)\Big] \tag{4.2}$$

To a first-order approximation, $1/r_{on}$, the third-order output conductance nonlinearity G_{030}^{ds} and output capacitiance nonlinearity C_{030}^{ds} are proportional to W (for a given process and a fixed channel length). Therefore, we define $r_{on} = K_{r_{on}}/W$, $G_{030}^{ds} = K_{G_{030}^{ds}} \cdot W$ and $C_{030}^{ds} = K_{C_{030}^{ds}} \cdot W$ which yields:

$$v_{out}^{\omega_{IM3}} \approx \frac{3 \times {V_{IN}}^3 {K_{r_{on}}}^4}{8{R_s}^3(1+A)^4} \times \Big(K_{G_{030}^{ds}} + j\omega_{IM3}K_{C_{030}^{ds}}\Big) \times \left[\frac{16A^4}{{W_{M_1}}^3} - \frac{(1-A)^4A^3}{{W_{M_2}}^3} - \frac{(1-A)^4A}{{W_{M_3}}^3}\right] \tag{4.3}$$

Equation (4.3) indicates that for a certain attenuation value 0<A<1, the distortion current (both resistive and capacitive) of switch M_1 can cancel the distortion from switch M_2 and M_3 for specific width ratio between M_1, M_2 and M_3. This is illustrated in Fig. 4.2, where the distortion current of M_1 flows into the load R_{load} and the distortion currents of M_2 and M_3 flow outwards from R_{load}.

For our 0.16µm CMOS process $K_{G_{030}^{ds}}$ is dominant over $K_{C_{030}^{ds}}$ for frequencies up to about 10GHz. This allows for simplification of (4.3) into

$$v_{out}^{\omega_{IM3}} \approx \frac{3 \times {V_{IN}}^3 {K_{Ron}}^4 K_{G_{030}^{ds}}}{8{R_s}^3(1+A)^4} \times \left[\frac{16A^4}{{W_{M_1}}^3} - \frac{(1-A)^4A^3}{{W_{M_2}}^3} - \frac{(1-A)^4A}{{W_{M_3}}^3}\right] \tag{4.4}$$

Solving for W_{M_1}, the optimum switch width f is:

$$W_{M_1,\mathrm{opt}} \approx \sqrt[3]{\frac{2}{\frac{(1-A)A^2}{{W_{M_2}}^3} + \frac{(1-A)}{{W_{M_3}}^3}}} \cdot \frac{2A}{(1-A)} \tag{4.5}$$

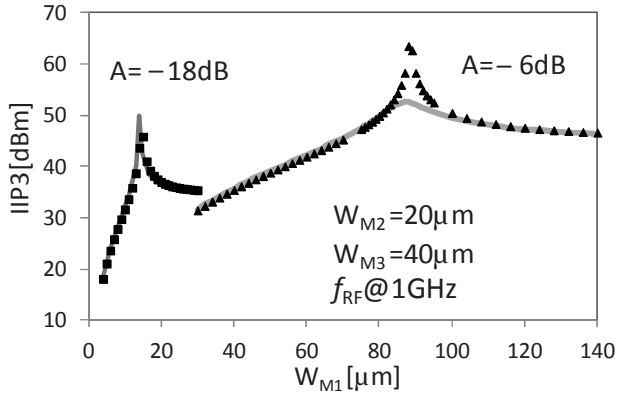


Fig. 4.3. Simulated and calculated IIP3 as a function of W_{M1} for two Π-attenuator (A=-6dB and A=-18dB) at 1GHz. Symbol for calculation and line for simulation.

Note that, the simplification from (4.3) to (4.4) doesn't affect the result of optimal width in (4.5). As can be observed from (4.4), the IM3-distortion decreases for larger transistors, since less voltage drops across transistors, as is well known, but this comes at the cost of bandwidth. Nevertheless, (4.4) indicates that the cancellation also works for attenuators with small transistors. This breaks the tradeoff between linearity and bandwidth. Since the IM3 cancellation requires different width ratios for different attenuation settings, it mandates the use of a discrete step attenuator.

Fig. 4.3 shows simulation results for two Π-attenuators (A=-6dB and A=-18dB) by sweeping W_{M1} for fixed W_{M2}=20µm and W_{M3}=40µm, which are rather small values. To keep 50Ω impedance matching and the desired attenuation, R_1, R_2 and R_3 are set accordingly (thus R_1 is swept along with W_{M1}). [1] The center frequency f_{RF} is 1GHz with tones at f_{RF}+1.6MHz (so 3.2MHz spacing) and the IIP3 is extrapolated for input power from -15dBm to -5dBm. The simulation results agree very well with the simple model of (4.4), which is plotted in Fig. 4.3. The close-to-infinite IIP3 at the optimum

[1] All simulations are performed in Spectre, using the PSP compact MOSFET model [9] fitted to our 0.16µm CMOS process. The PSP model is known to correctly fit derivatives up to the third order [10-11] and to satisfy the so-called Gummel symmetry test [12-13], which is essential for accurate simulation of distortion in the attenuator

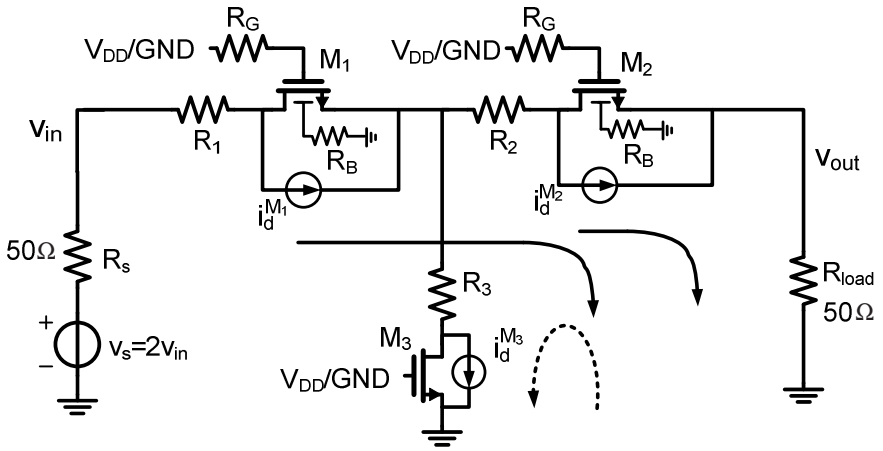


Fig. 4.4. Illustration of IM3 cancellation within T attenuator.

width in calculation is due to neglecting other nonlinearities than G_{030}^{ds} in (4.4). Nevertheless, the optimum width is well predicted. For small W_{M1}, M_1 is dominant for the IM3 output. As W_{M1} increases, its distortion decreases and hence IIP3 increases until its maximum at full IM3 cancellation. For even larger W_{M1}, the IIP3 is dominated by M_2 and M_3 yielding a saturated sub-optimum value because W_{M2} and W_{M3} are fixed. The IIP3 peaking area for A=-6dB is less sensitive to width variations than for A=-18dB since larger W_{M1} is used for A=-18dB. As suggested by (4.4), the IM3 cancellation is robust against process spread since it only relies on the ratio of transistor widths and predetermined gain settings, assuming that the spread of passive resistors in the attenuators is small. As the switches operate in very deep triode with gate connected to V_{DD}, the threshold voltage mismatches hardly play a role. The effect of devices mismatches at IIP3 peaking region can be minimized by using wide switches.

4.2.2 T-Attenuator

For the T-attenuator shown in Fig. 4.4, a similar analysis can be performed, yielding (see Appendix II for the derivation)

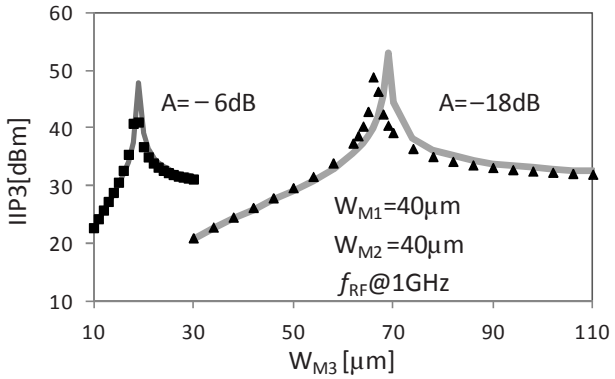


Fig. 4.5. Simulated and calculated IIP3 as a function of W_{M3} for two T-attenuatiors values (A=-6dB and A=-18dB). Symbol for calculation and line for simulation.

$$v_{out}^{\omega_{IM3}} \approx \frac{3 \times {V_{IN}}^3}{64{R_s}^3} \times \left[A\left(r_{on}^{M_1}\right)^4\left(8Y_{030,M_1}^{ds} + 4Y_{120,M_1}^{ds} + 2Y_{210,M_1}^{ds} + Y_{300,M_1}^{ds}\right)\right. \quad (4.6)$$
$$\left. + A^3\left(r_{on}^{M_2}\right)^4\left(8Y_{030,M_2}^{ds} + 4Y_{120,M_2}^{ds} + 2Y_{210,M_2}^{ds} + Y_{300,M_2}^{ds}\right) - 8(1-A)^4\left(r_{on}^{M_3}\right)^4 Y_{030,M_3}^{ds}\right]$$

Since for our 0.16µm CMOS process the third-order output conductance nonlinearity G_{030}^{ds} contributes dominantly to the attenuator output distortion for f_{RF}<10GHz, equation (4.6) can be simplified to

$$v_{out}^{\omega_{IM3}} \approx \frac{3 \times {V_{IN}}^3 \cdot {K_{R_{on}}}^4 \cdot K_{G_{030}^{ds}}}{8{R_s}^3} \times \left[\frac{A}{{W_{M_1}}^3} + \frac{A^3}{{W_{M_2}}^3} - \frac{(1-A)^4}{{W_{M_3}}^3}\right] \quad (4.7)$$

As illustrated in Fig. 4.4, the distortion currents of M_1/M_2 flow into the load R_{load} while the distortion current of M_3 flow outwards of R_{load}. The optimum switch width for full IM3 cancellation is

$$W_{M_3,opt} \approx \sqrt[3]{\frac{1-A}{\frac{A}{{W_{M_1}}^3} + \frac{A^3}{{W_{M_2}}^3}}} \cdot (1-A) \quad (4.8)$$

A similar simulation as for the T-attenuator is performed, and the results are shown in Fig. 4.5. Here we swept W_{M3} for two attenuation values (A=-6dB and A=-18dB) with W_{M1}=40µm and W_{M2}=40µm, and the resistors again set to achieve a 50Ω impedance matching and the targeted attenuation. The optimum width is predicted with good accuracy by (4.8) (for A=-6dB,

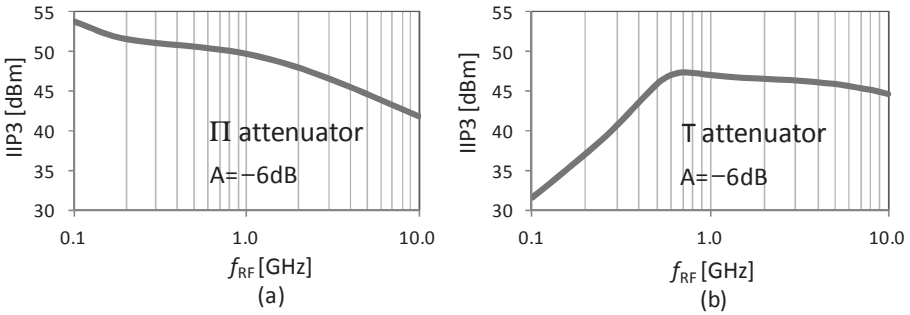


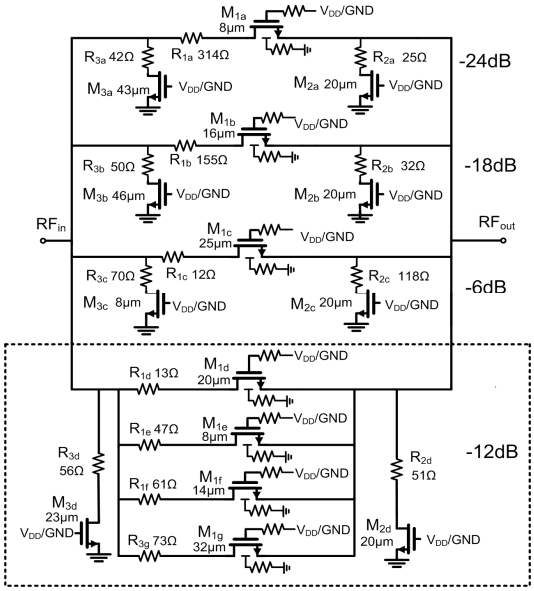
Fig. 4.6. Simulated IIP3 as a function of frequency for (a) Π-attenuator with 6dB attenuation (V M1=88μm, V M2=20μm and V M3=40μm) and (b) T-attenuator with 6dB attenuation (V M1=40μm, V M2=40μm and V M3=19μm).

calculated $W_{M3,opt}$ =19μm is equal to the simulated optimum; for A=-18dB, calculated $W_{M3,opt}$ =66μm and simulated $W_{M3,opt}$ =69μm).

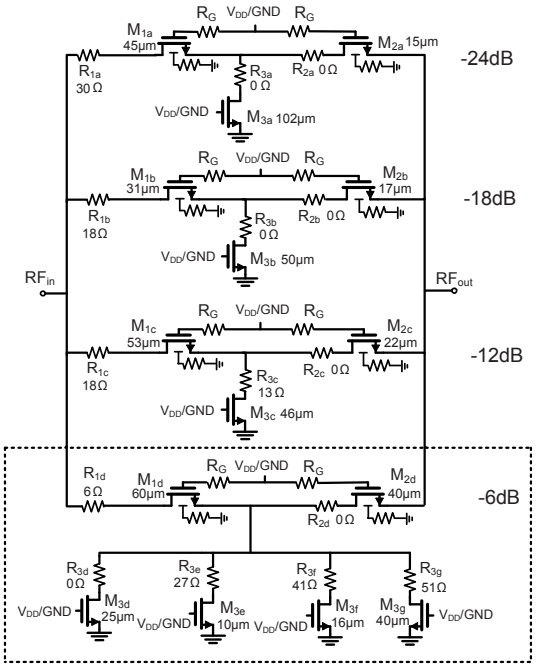
4.3 Limiting factors for IM3 cancellation

4.3.1 Effect of parasitic capacitance

In the previous analysis, it was assumed that the distortion currents of the transistors are in phase or out-of-phase, and hence they have either 0^o or 180^o phase shift with respect to each other. At low frequencies this assumption is valid; at higher frequencies the parasitic capacitances, however, violate this assumption, leading to degraded distortion cancellation. This is illustrated in Fig. 4.6 where the simulated IIP3 of the optimized A=-6dB attenuator of sections 4.2.1 and section 4.2.2 is shown as a function of frequency. Clearly the IIP3 degrades as the frequency increases. The lower IIP3 for the T-attenuator, at low frequency is because the AC-bootstrapping for the series devices becomes less effective.



(a)



(b)

Fig. 4.7. The schematic of attenuator networks implemented in a 0.16μm bulk CMOS process. (a) Π-attenuator network and (b) T-attenuator network.

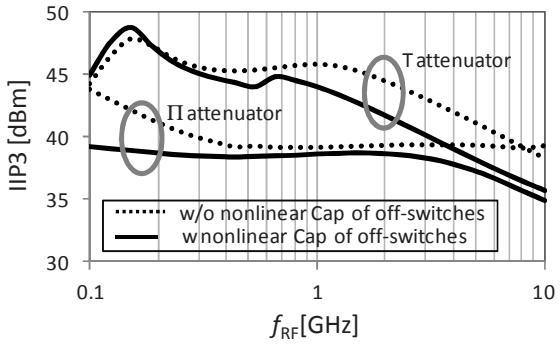


Fig. 4.8. The effect of off-state switches' nonlinear capacitances on the simulated IIP3 for the -24dB attenuation branch in both Π-attenuator network and T-attenuator network shown in Fig. 4.9.

4.3.2 Effect of nonlinear capacitance

Since the proposed IM3 cancellation requires specific switch ratios for each specific attenuation setting, we use the discrete-step attenuator network system shown in Fig. 4.7a and Fig. 4.7b. Each attenuator branch is optimized for IM3 cancellation at a specific attenuation value. During operation only one branch is enabled. The nonlinear parasitic capacitances of off-state transistor switches now usually set an upper boundary for the maximum IIP3 that can be achieved by the enabled attenuator branch. As high linearity is desirable for high attenuation settings, to demonstrate the effect of nonlinear capacitances, we simulate the -24dB attenuator branch in Fig. 4.7a and Fig. 4.7b as a function of f_{RF}, with and without the other branches connected to the system. When all other attenuator branches (disabled) are connected to the -24dB branch at the input/output port, the simulated IIP3 includes the effect of the nonlinear capacitances in off-state switches. When all other attenuator branches (disabled) are disconnected from the -24dB branch at the input/output port, no effect of the nonlinear capacitances in off-state switches is included in the simulated IIP3. As shown in Fig. 4.8, the nonlinear capacitances of the off-state switches in the disabled attenuator branches reduce the IIP3 of the enabled attenuator, especially for high frequency. For the Π-

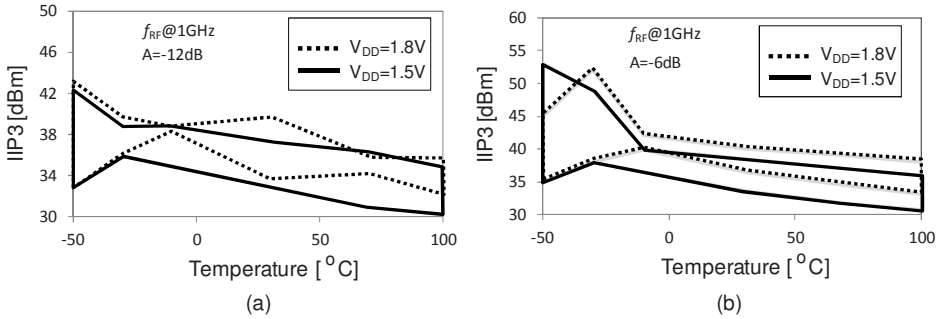


Fig. 4.9. The simulated IIP3 range of 200-time Monte Carlo simulations for mismatch and process spread at 1GHz as a function of temperature when VDD=1.8V (full supplR) and VDD=1.5V. (a) for -12dB Π-attenuator and (b) for -6dB T-attenuator

attenuator network, at f_{RF} <0.3GHz, the nonlinear capacitances also degrades the IIP3 since the bootstrapping is less effective.

4.3.3 Effect of Process, Voltage and Temperature Variations

As indicated by (4.5) and (4.6), the proposed IM3 cancellation technique relies on transistor width ratios, which makes it inherently robust against PVT variations to some extent. For a -12dB Π-attenuator in Fig. 7a (M_{1d}, M_{2d}, M_{3d}) and the -6dB T-attenuator in Fig. 7b (M_{1d}, M_{2d}, M_{3d}), we ran 200-time Monte-Carlo simulations using realistic production variations at different temperatures and different V_{DD} to check the effect of PVT variations. Fig. 4.9 shows the maximum and minimum IIP3 at 1GHz as a function of temperature for V_{DD}=1.8V (full supply) and V_{DD}=1.5V. It shows, for a wide temperature range [-50°C to 100°C], both the Π-attenuator and T-attenuator always achieves >30dBm IIP3, even when the supply voltage drops to 1.5V.

Overall, the proposed IM3 cancellation is relatively robust against PVT variations. Wide transistor switches reduces the sensitivity of full IM3 cancellation over mismatches, but limits bandwidth and introduce two factors that limit maximum achievable IIP3: capacitance nonlinearity and phase shift due the parasitic capacitances. As a result, careful optimization is necessary for the trade-off between using wide transistors for lower mismatches

sensitivity and small transistor widths for broader bandwidth and less off-state nonlinear capacitances.

4.4 **Design**

To verify the concept of IM3 cancellation, the Π-attenuator network and a T-attenuator network of Fig. 4.7 are implemented in a standard 0.16µm CMOS process. Both attenuators networks contain two blocks for two different measurement purposes: 1) an attenuator block (A =-12dB for the Π-attenuator, A =-6dB for the T-attenuator, encircled in Fig. 4.7) for demonstrating the validity of (4.5) and (4.8) and 2) a four-step attenuator system with 6dB, 12dB, 18dB and 24dB attenuation.

In the 12dB attenuation block of the Π-attenuator, each of the four branches is designed for 12dB attenuation, but has different width for M_1 to obtain maximum or only partial IM3 cancellation. Therefore, this mimics a Π-attenuator with selectable W_{M1} for fixed W_{M2} (20µm) and W_{M3} (23µm).

The four-step Π-attenuator system contains the upper three attenuator branches in Fig. 4.7a (all optimized) and the attenuator branch in the -12 dB block that is optimized for IM3 cancellation (W_{M1} = 20µm). During operation, only one branch is enabled. For isolation and bootstrapping purposes, the gate and bulk of M_1 are connected to the controlling voltage via 40k resistors; the gates and bulks of the shunt devices are connected directly to the controlling voltages to save area. For minimum signal attenuation, the transistors M_{1a}, M_{1b}, M_{1c}, M_{1d} are enabled, and the shunt transistors are disabled, yielding an additional -1.8 dB setting that sets the minimum IL of this system. Poly resistors are used for the series and shunt resistance in the attenuator because of their high linearity (IIP3 around +50dBm according to simulations).

The T-attenuator is designed in a similar way, with the minimum attenuation equal to -1.2dB. A digital decoder provides the controlling voltage (V_{DD} = 1.8V for enabling and V_{SS} = 0V for disabling), and is shared by the attenuators. Simulated nominal IIP2 for the Π-attenuator (T-attenuator) is +55dBm (+45dBm) for all settings at f_{RF} = 2.5 GHz with a two-tone spacing up to 1 GHz, and can be improved by increasing the switch size.

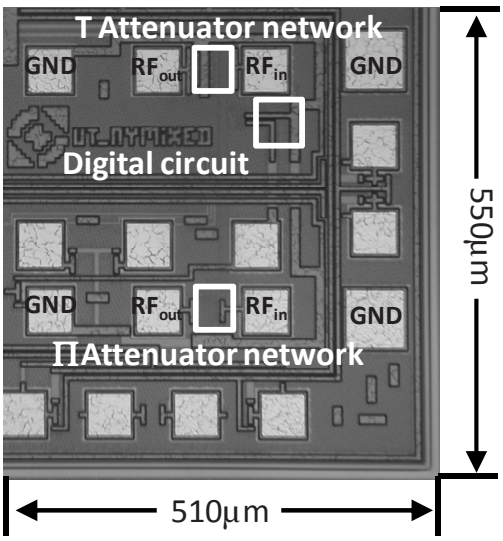


Fig. 4.10. The chip microphotograph for Π-attenuator network and T-attenuator network fabricated in a standard 0.16µm CMOS process.

The chip micrograph is shown in Fig. 4.10. The active areas of the digital decoder (not optimized), the Π-attenuator system and the T-attenuator system are 60x65µm^2, 50x30µm^2 and 54x53µm^2 respectively.

4.5 **Measurement**

The measurements were performed by on-wafer probing. The P_{1dB} is extrapolated from -20dBm, and IIP3 from -15dBm with 3.2MHz two-tone spacing. In all simulations the estimated bondpad capacitances are included at the input/output of the attenuator system. Using the -12 dB setting of the Π-attenuator (-6 dB for the T-attenuator), W_{M1} (W_{M3}) is varied, and the correspondence between (4.5) ((4.8)), simulations and measurements at 1 GHz is verified in Fig. 4.11. The IIP3-peaking is clearly visible in the measurements, and occurs at the width predicted by our model and simulations. The small difference between measured and simulated IIP3 may be due to limited accuracy of transistor modeling and bondpad parasitics. For the optimum W_{M1}

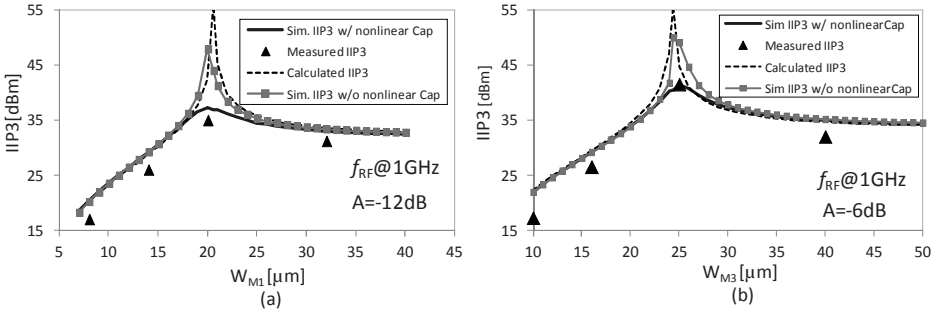


Fig. 4.11. Measured and simulated IIP3 for f_{RF}=1GHz for (a) Π-attenuator as a function of W_{M1} (b) T-attenuator as a function of W_{M3}. Line for simulation results, symbol for measurement results.

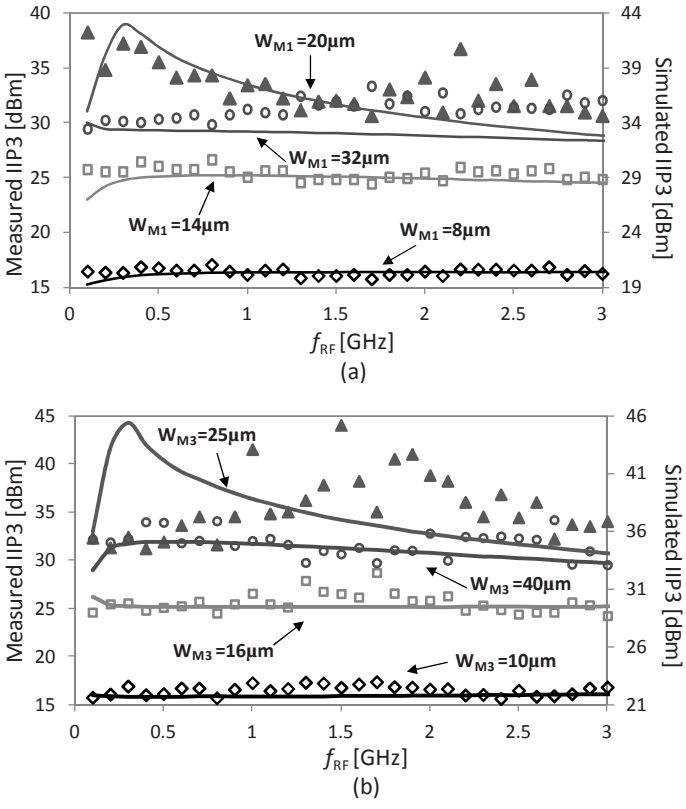


Fig. 4.12. Measured and simulated IIP3 vs f_{RF} for (a) Π-attenuator as a function of W_{M1} (b) T-attenuator as a function of W_{M3}. Line for simulation results, symbol for measurement results.

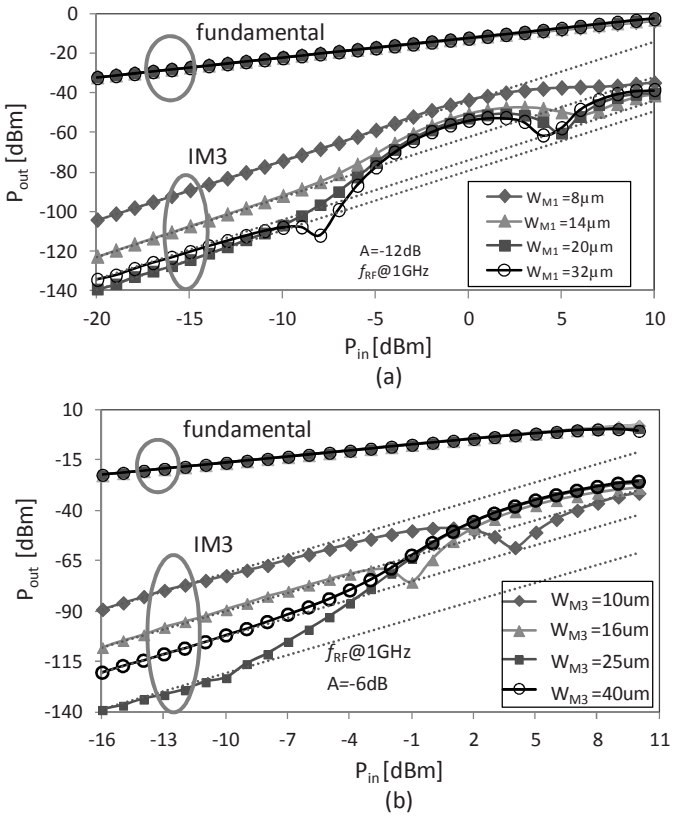


Fig. 4.13 Measured IIP3 curves vs input power at 1GHz for (a) Π-attenuator for different W_{M1} (b) T-attenuator for different W_{M3}.

(W_{M3}), the achievable maximum IIP3 is limited by the nonlinear capacitances of the off-state switches.

IIP3 as a function of f_{RF} is shown in Fig. 4.12. The measurement points match relatively well with the simulations including the nonlinear capacitors (note the slightly different scale of the measurement and simulation y-axes).

Fig. 4.13a shows the IIP3 curves for the different W_{M1} (W_{M3}) at 1GHz. It shows that the IM3 improvement becomes less effective for input powers above approximately -8dBm due to the higher-order nonlinearities. Nevertheless, the IM3-products at the optimum width remain the lowest up to even higher input powers.

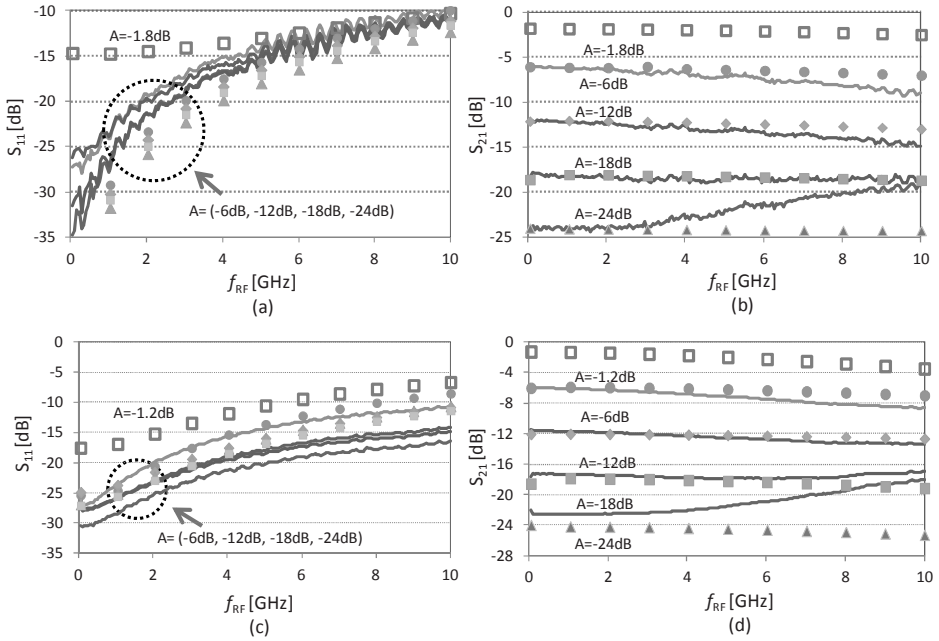


Fig. 4.14. S_{11}/S_{21} of the four-step (a,b) Π-attenuator system (c,d) T-attenuator system. Line for measurement results, symbol for simulation results.

Proceeding to the optimized four-step attenuator systems, the measured and simulated S_{11} and S_{21} (50 reference) for the different settings of the Π-attenuator (T-attenuator) are shown in Fig. 4.14. Due to a mistake in the decoder design, the minimum attenuation setting of -1.8dB (-1.2dB) cannot be enabled, so we only show their *simulated* values. Due to unaccounted parasitics, the measured S_{21} deviates > 1.6dB for f_{RF}>5GHz (f_{RF}>5.6GHz) for the -24dB setting. The T-attenuator requires wider shunt devices for the high attenuation settings, and thus wider series devices for IM3 cancellation. This results in less bandwidth (S_{11} <-10dB) and more area compared to the Π-attenuator system. The measured summation of NF and IL (NF + IL = NF- S_{21}) is below 0.5dB for all settings, which shows that negligible excess noise is introduced by the attenuator.

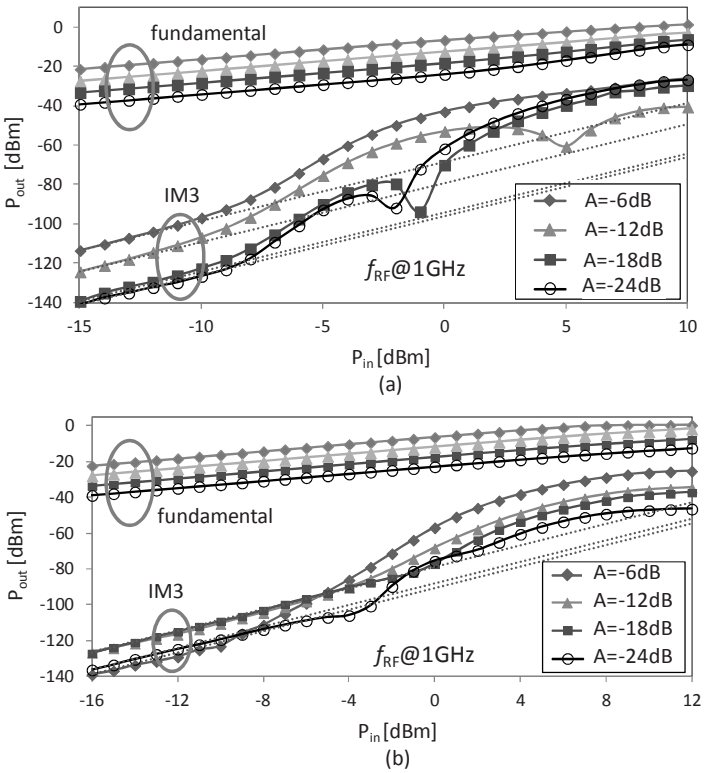


Fig. 4.15. Measured IIP3 curves vs input power at 1GHz for different settings of the (a)Π-attenuator system (b)T-attenuator.

The measured IIP3 curves as a function of input power at f_{RF}=1GHz are shown in Fig. 4.15. The IIP3 for the Π-attenuator (T-attenuator) are respectively 31dBm (40dBm), 33dBm (34dBm), 38dBm (30dBm) and 36dBm (35dBm) for attenuation settings -6dB, -12dB, -18dB and -24dB. Again, for high input powers higher order nonlinearities kick in.

Fig. 4.16 summarizes the measured IIP3 for various input frequency f_{RF}. Due to bandwidth limitations of our measurement setup, IIP3 cannot be measured below f_{RF} = 50MHz. For both attenuators, IIP3 is above 30dBm in the TV bands (0.05–1GHz). The T-attenuator obtains an IIP3 >+25dBm for

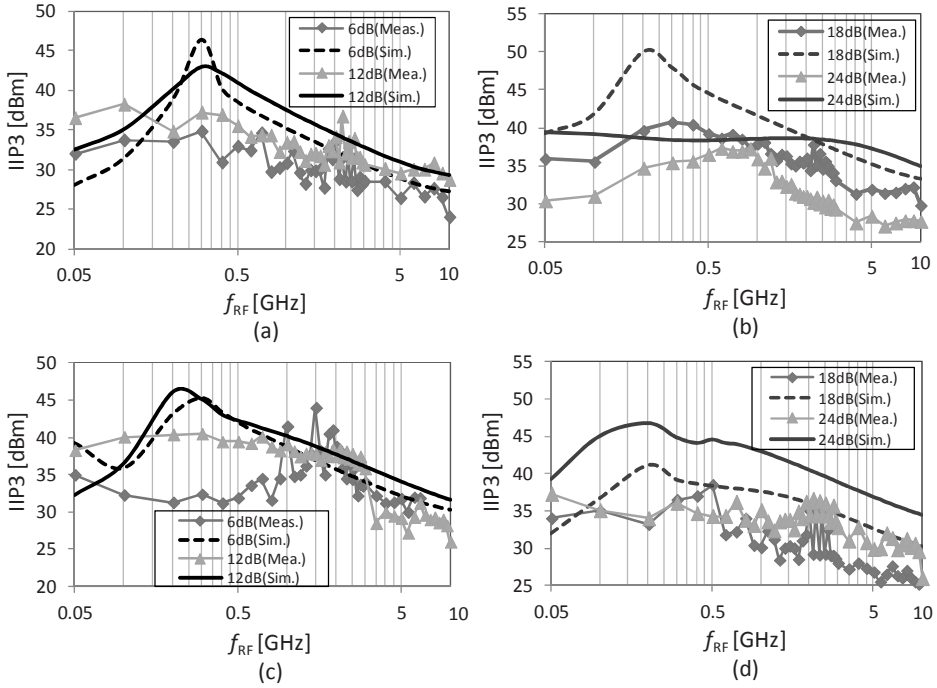


Fig. 4.16 Measured and simulated IIP3 vs f_{RF} for different settings of the (a,b)Π-attenuator system (c,d) T-attenuator system.

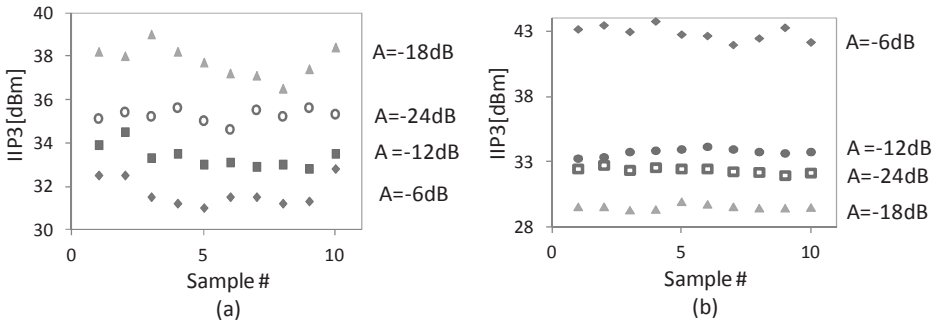


Fig. 4.17 Measured IIP3 of ten samples at 1GHz for (a) Π-attenuator system (b) T-attenuator.

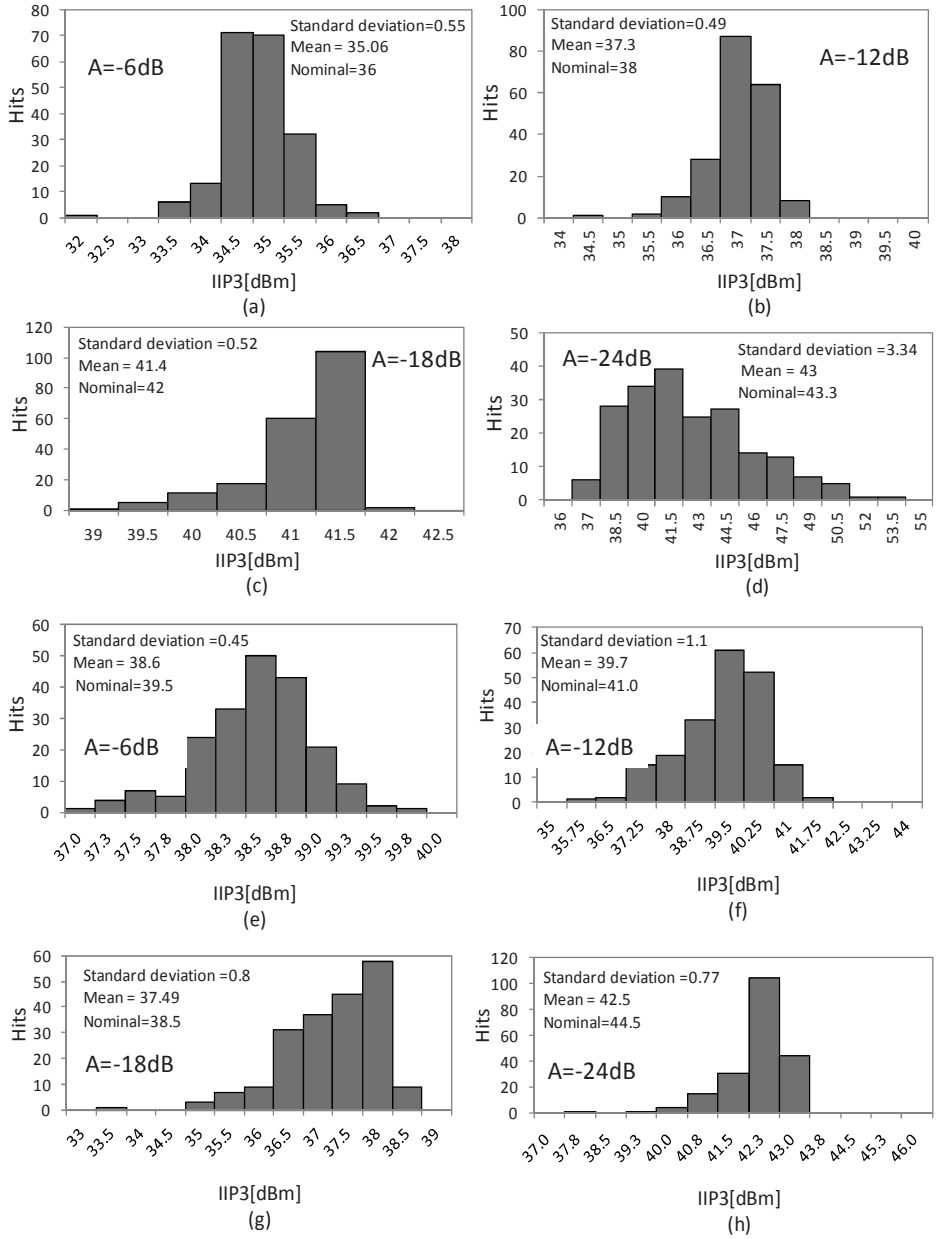


Fig. 4.18 Simulated IIP3 of 200-time Monte Carlo simulation with mismatches&process spread at 1GHz for the (a,b,c,d) Π-attenuator system and (e,f,g,h) T-attenuator.

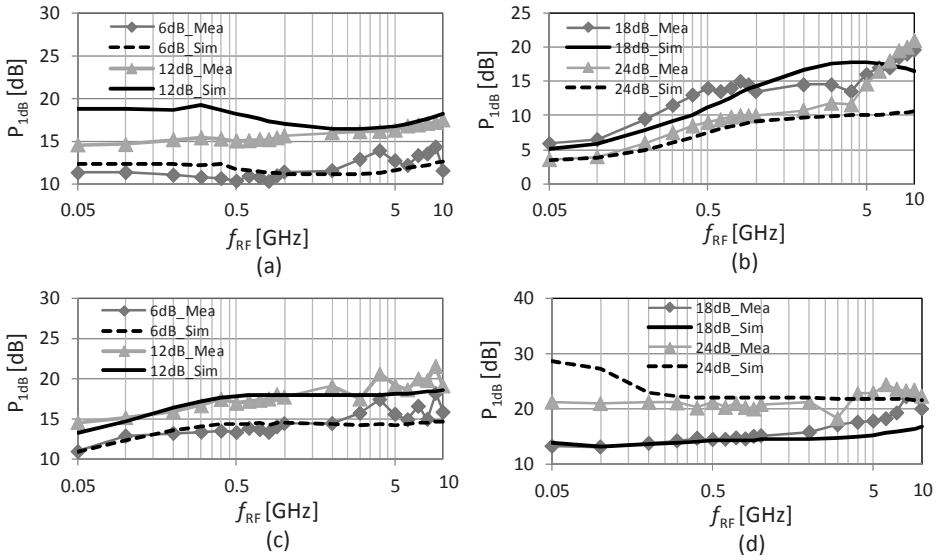


Fig. 4.19 Measured and simulated P_{1dB} as a function of f_{RF} for the T-attenuator system with four attenuation settings.

the whole range 0.05–10GHz, while the Π-attenuator obtains an IIP3>+26dBm for 0.05–5GHz and IIP3>+24dBm for 0.05–10GHz. At higher f_{RF}, extra phase shifts caused by the parasitic capacitances degrades the IM3 cancellation.

The measured IIP3 of ten dies in one wafer for f_{RF}=1GHz shows <1.5dB IIP3 variation, as shown in Fig. 4.17. The results of a 200-run Monte Carlo simulation for mismatches and process spread is shown in Fig. 4.18. The difference between mean IIP3 and minimum IIP3 is <6dB and IIP3 is >32dBm for all samples, which shows the robustness of this IM3 cancellation technique.

At f_{RF}=2.5GHz, measured curves for a two-tone spacing of 100kHz and 30MHz showed negligible difference compared to the 3.2MHz spacing, which is confirmed by simulations for a spacing from 100 kHz to 1 GHz at the same f_{RF}.

The Π-attenuator obtains P_{1dB}>+3dBm for 0.05–10GHz, see Fig. 4.19. For attenuation branches A=−18dB and A=−24dB, M_1 is quite small and

TABLE I
COMPARISON WITH STATE-OF-THE-ART ATTENUATORS.

	Huang [2]	Youssef [5]	Dogan [6]	Granger-Jones [3]	Ku [14]	This work (Π)	This work (T)
CMOS	0.18 μm	65 nm	0.13 μm	SOI	0.18 μm	0.16 μm	0.16 μm
V_{DD} [V]	1.8	1.2	1.2	5	N/A	1.8	1.8
Chip area [mm^2]	0.28	0.05	0.7	N/A	0.5	0.0054	0.0067
Bandwidth [GHz]	0.4–3.7	0.4–0.8	0.0–2.5	0.05–4.0	0.0–14	0.0–5.0	0.0–5.6
IIP3 [dBm]	+15 (0.7 GHz)	+23	+10 (10 GHz)	+47	+29 (10 GHz)	+30 (0.05–1 GHz) +27 (0.05–5 GHz)	+30 (0.05–1 GHz) +27 (0.05–5.6 GHz)
CP [dBm]	+6 (0.7 GHz)	N/A	+2.5	+30	+15 (10 GHz)	+3 (0.05–1 GHz) +10 (1.0–10 GHz)	+11 (0.05–10 GHz)
Att. flatness [dB]	2.6	N/A	2.6	3.0	0.7	1.6	1.6
Max. attenuation [dB]	33	48	42	40	31.5	24	24
Min. attenuation [dB]	0.96–2.9	N/A	0.9–3.5	2.4–4.0	3.7–10	1.8–2.4 (simulation)	1.3–2.2 (simulation)
Return loss [dB]	> 9	> 12	> 8.2	> 14	> 9	> 14	> 10
Control mode	linear-in-dB	discrete step	linear-in-dB	linear-in-dB	discrete step	discrete step	discrete step

takes a large voltage swing, causing P_{1dB}<10dBm for f_{RF}<1GHz. Using wider M_1 in these settings can increase P_{1dB}. The T-attenuator has a higher P_{1dB} of 11dBm, because the two devices in series divide the voltage between v_{in} and v_{out}, and, moreover, each devices is wider, thus takes less voltage swing and generates less distortion. At lower frequencies the AC-bootstrapping becomes less effective, increasing v_{gs} of the series transistors, thus generating more distortion and decreasing P_{1dB}. Increasing the gate series resistor can alleviate this problem, but making it too large may increase the noise.

4.6 **Benchmarking**

In Table I we compare the two optimized designs with state of-the-art attenuators. Both the Π- and T-attenuator system using the proposed IM3 cancellation technique achieve very high linearity and high bandwidth for a very low active area in standard bulk CMOS. By using the proposed IM3 cancellation technique, the linearity of the Π-attenuator can be improved to the same level of the T-attenuator for similar transistor sizes.

4.7 **Conclusion**

A wideband IM3 cancellation technique is introduced for CMOS Π-attenuators and T-attenuators. For specific transistor width ratios, the dominant distortion currents cancel at the load, which results in a high IIP3, even

for relatively small transistors. Simple equations for transistor width dimensioning were introduced, and it was proved that they have good accuracy. This technique alleviates the trade-off between bandwidth and area on the one hand and high linearity on the other hand, without introducing extra devices, and thus enables highly linear wideband CMOS attenuators with small active area. A four-step Π-attenuator system designed in 0.16 µm CMOS using this IM3 cancellation achieves >30dBm IIP3 for the TV bands (0.05–1GHz), > 26dBm IIP3 for 0.05–5GHz and >3dBm P_{1dB} 0.05–10GHz, while only occupying 0.0054mm^2 of active area. A four-step T-attenuator system design achieves similar performance. Both measurement and simulation results show good robustness of this IM3 cancellation technique.

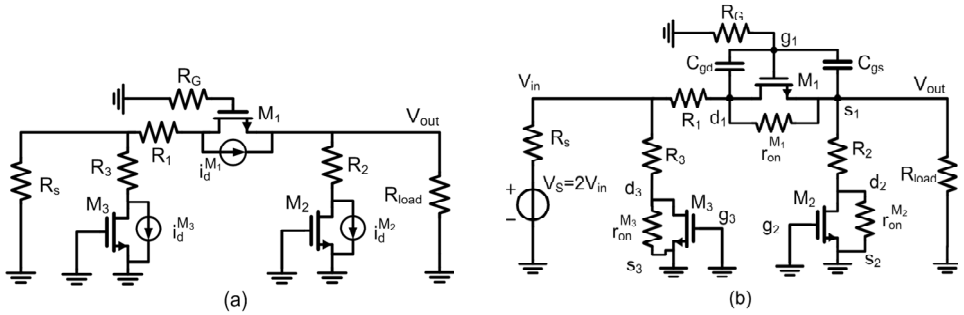


Fig. A1. The equivalent models for Π-attenuator (a) for calculating the H function. (b) for calculating the β function.

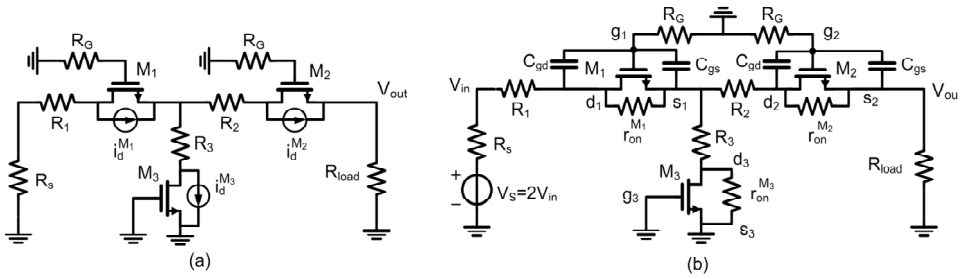


Fig. A2. The equivalent models for T-attenuator (a) for calculating the H function. (b) for calculating the β function.

4.8 **Appendix**

Applying the general nonlinearity model given in [8] to both Π-attenuator and T-attenuator shown in Fig. 4.2 and Fig.4.4 and only including the third-order nonlinearity yields

$$v_{out}^{\omega_{IM3}} \approx H_{M_1}^{ds} \cdot i_d^{M_1} + H_{M_2}^{ds} \cdot i_d^{M_2} + H_{M_3}^{ds} \cdot i_d^{M_3} \tag{A4.1}$$

$$= \frac{3V_{IN}{}^3}{4} \times [H_{M_1}^{ds} \cdot (\beta_{030,M_1} G_{030,M_1}^{ds} + \beta_{300,M_1} G_{300,M_1}^{ds} + \beta_{210,M_1} G_{210,M_1}^{ds} + \beta_{120,M_1} G_{120,M_1}^{ds}) + H_{M_2}^{ds} \cdot \beta_{030,M_2} G_{030,M_2}^{ds} + H_{M_3}^{ds} \cdot \beta_{030,M_3} G_{030,M_3}^{ds})]$$

, where $H_{M_k}^{ds} = \frac{/_{out}}{i_d^{M_k}}$ and $\beta_{nml,M_k}^{ds} = \frac{\left(v_{gs}^{M_k}\right)^n \left(v_{ds}^{M_k}\right)^m \left(v_{bs}^{M_k}\right)^l}{v_{in}^3}$ with $n, m, l \in \mathbb{N}; n +$

$m+l \in (1,2,3)$. For the Π-attenuator, H and β are calculated using the model shown in Fig. A1; For the T-attenuator, H and β are calculated usingthe model shown in Fig. A2.

For the Π attenuator, let $R_x = R_1 + r_{on}^{M_1}$, $R_y = R_2 + r_{on}^{M_2} = R_y = R_3 + r_{on}^{M_3}$, $R_{load} = R_s$. For input matching to R_s, we have $R_x = R_s(1-A)(1+A)/2A$ and $R_y = R_s(1+A)/(1-A)$. As a result, the H and β fucntions convert to

$$H_{M_1}^{ds} = \frac{v_{out}}{i_D^{M_1}} = \frac{A \cdot r_{on}^{M_1}}{1+A} \qquad H_{M_2}^{ds} = \frac{v_{out}}{i_D^{M_2}} = \frac{-(1-A)r_{on}^{M_2}}{2(1+A)}$$
$$H_{M_3}^{ds} = \frac{v_{out}}{i_D^{M_3}} = \frac{-A(1-A)r_{on}^{M_3}}{2(1+A)} \tag{A4.2}$$

$$\beta_{030,M_1}^{ds} = \left(\frac{v_{ds}^{M_1}}{v_{in}}\right)^3 = \left(\frac{R_{on}^{M_1}}{R_x + \frac{R_y R_{load}}{R_y + R_{load}}}\right)^3 \qquad \beta_{300,M_1}^{ds} = \left(\frac{v_{gs}^{M_1}}{v_{in}}\right)^3 = \frac{\beta_{030,M_1}^{ds}}{8}$$

$$\beta_{210,M_1}^{ds} = \frac{\left(v_{gs}^{M_1}\right)^2 \left(v_{ds}^{M_1}\right)}{v_{in}^3} = \frac{\beta_{030,M_1}^{ds}}{4} \qquad \beta_{120,M_1}^{ds} = \frac{\left(v_{gs}^{M_1}\right)\left(v_{ds}^{M_1}\right)^2}{v_{in}^3} = \frac{\beta_{030,M_1}^{ds}}{2} \tag{A4.3}$$

$$\beta_{030,M_2}^{ds} = \left(\frac{v_{ds}^{M_2}}{v_{in}}\right)^3 = \left(\frac{\frac{R_y R_{load}}{R_y + R_{load}}}{R_x + \frac{R_y R_{load}}{R_y + R_{load}}}\right)^3 \qquad \beta_{030,M_3}^{ds} = \left(\frac{v_{ds}^{M_3}}{v_{in}}\right)^3 = \left(\frac{r_{on}^{M_3}}{R_y}\right)^3$$

and (A1) converts to (4.1).

For the T-attenuator let $R_x = R_1 + r_{on}^{M_1}$, $R_y = R_2 + r_{on}^{M_2} = R_3 + r_{on}^{M_3}$ and $R_{load} = R_s$, for the input matching to R_s, we have $R_x = \frac{R_s(1-A)}{(1+A)}$ and $R_y = \frac{2AR_s}{(1-A)(1+A)}$. As a result, the H and β fucntions can be found as

$$H_{M_1}^{ds} = \frac{v_{out}}{i_D^{M_1}} = \frac{A r_{on}^{M_1}}{2} \qquad H_{M_2}^{ds} = \frac{v_{out}}{i_D^{M_2}} = \frac{r_{on}^{M_2}}{2} \qquad H_{M_3}^{ds} = \frac{v_{out}}{i_D^{M_3}} = \frac{(-1+A)r_{on}^{M_3}}{2} \tag{A4.4}$$

$$\beta_{030,M_1}^{ds} = (\frac{v_{ds}^{M_1}}{v_{in}})^3 = (\frac{r_{on}^{M_1}}{R_s})^3 \qquad \beta_{300,M_1}^{ds} = (\frac{v_{gs}^{M_1}}{v_{in}})^3 = \frac{1}{8}(\frac{r_{on}^{M_1}}{R_s})^3$$

$$\beta_{210,M_1}^{ds} = \frac{(v_{gs}^{M_1})^2(v_{ds}^{M_1})}{v_{in}^3} = \frac{1}{4}(\frac{r_{on}^{M_1}}{R_s})^3 \quad \beta_{120,M_1}^{ds} = \frac{(v_{gs}^{M_1})\ (v_{ds}^{M_1})^2}{v_{in}^3} = \frac{1}{2}(\frac{r_{on}^{M_1}}{R_s})^3$$

$$\beta_{030,M_2}^{ds} = (\frac{v_{ds}^{M_2}}{v_{in}})^3 = (\frac{Ar_{on}^{M_2}}{R_s})^3 \quad \beta_{300,M_2}^{ds} = (\frac{v_{gs}^{M_2}}{v_{in}})^3 = \frac{1}{8}(\frac{Ar_{on}^{M_2}}{R_s})^3 \qquad \text{(A4.5)}$$

$$\beta_{210,M_2}^{ds} = \frac{(v_{gs}^{M_2})^2(v_{ds}^{M_2})}{v_{in}^3} = \frac{1}{4}(\frac{Ar_{on}^{M_2}}{R_s})^3 \quad \beta_{120,M_2}^{ds} = \frac{(v_{gs}^{M_2})(v_{ds}^{M_2})^2}{v_{in}^3} = \frac{1}{2}(\frac{Ar_{on}^{M_2}}{R_s})^3$$

$$\beta_{030,M_3}^{ds} = (\frac{v_{ds}^{M_3}}{v_{in}})^3 = (\frac{(1-A)r_{on}^{M_3}}{R_s})^3$$

Then (A1) converts to (4.6).

4.9 **Reference**

[1] M. S. Oude Alink, E. A. M. Klumperink, A. B. J. Kokkeler, M. C. M. Soer, G. J. M. Smit and B. Nauta, "A CMOS-Compatible Spectrum Analyzer for Cognitive Radio Exploiting Crosscorrelation to Improve Linearity and Noise Performance," accepted by *IEEE Trans. Circuits and Systems I,.*

[2] Y. Y. Huang, W. Woo, Y. Yoon and C.H. Lee Y, "Highly linear RF CMOS variable attenuators with adaptive body biasing", *IEEE J. Solid-State Circuits*, Vol. 46, No. 5, May 2011.

[3] M. Granger-Jones, B. Nelson, and E. Franzwa, "A broadband high dynamic range voltage controlled attenuator MMIC with IIP3>+47 dBm over entire 30dB analog control range", *IEEE International Microwave Symp. Dig. (MTT)*, pp. 1-4, 2011.

[4] H. Dogan, R. G. Meyer and A. M. Niknejad, "Analysis and design of RF CMOS attenuators," *IEEE J. Solid-State Circuits*, Vol. 43, No. 10, Oct. 2008.

[5] A. Youssef, J. Haslett and E. Youssoufian, "Digitally-controlled RF passive attenuator in 65 nm CMOS for mobile TV tuner ICs," *Proc. IEEE Internationl Symposium on Circuits and Systems (ISCAS),* pp.1999–2002, 2010.

[6] H. Dogan and R. G. Meyer, "Intermodulation distortion in CMOS attenuators and switches," *IEEE J. Solid-State Circuits*, Vol. 42, No. 3, March 2007.

[7] W. Cheng, M. S. Oude Alink, A. J. Annema, G. Wienk and B. Nauta, "A wideband IM3 cancellation technique for CMOS attenuators," *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, San Francisco, CA,* 2012 , pp. 78 –79.

[8] W. Cheng, M. S. Oude Alink, A. J. Annema, J. A. Croon and B. Nauta, "RF circuit linearity optimization using a general weak nonlinearity model," accepted for publication in *IEEE Trans. Circuits and Systems I.*

[9] [Online]. Available: http://www.nxp.com/models/mos_models/psp/

[10] A. J. Scholten, G.D.J. Smit, B.A. de Vries, L.F. tiemeijer, J.A. Croon, D.B.M. Klaassen, R. van Langevelde, X. Li, W. Wu and G. Gildenblat, "The New CMC Standard Compact MOS Model PSP: Advantages for RF Applications", *IEEE J. Solid State Circuits*, vol. 44, no. 5, may 2009, pp. 1415-1424.

[11] X. Li, W. Wu, A. Jha, G. Gildenblat, R. van Langevelde, G.D.J. Smit, A.J. Scholten, D.B.M. Klaassen, C.C. McAndrew, J. Watts, C.M. Olsen, G.J. Coram, S. Chaudhry and J. Victory, "Benchmark Tests for MOSFET Compact Models With Application to the PSP Model", *IEEE Trans. Electron Devices*, vol. 56, no. 2, Feb. 2009, pp. 243-251.

[12] P. Bendix, P. Rakers, P. Wagh, L. Lemaitre, W. Grabinski, C.C. McAndrew, X. Gu and G. Gildenblat, "RF Distortion Analysis with Compact MOSFET Models", *IEEE Custom Integrated Circuits Conference (CICC),* pp. 9-12，2004.

[13] C. C. McAndrew, "Validation of MOSFET model Source-Drain Symmetry", *IEEE Trans. Electron Devices*, vol. 53, no. 9, Sept. 2009, pp. 2202-2206.

[14] B.-H. Ku and S. Hong, "6-bit CMOS digital attenuators with low phase variations for X-band phased-array systems," *IEEE Trans. Microw. Theory Tech.,* vol. 58, no. 7, pp. 1651-1663, 2010.

Chapter 5

Circuit modeling for Active Mixers

As active mixers face the problem of flicker noise and distortion more than passive mixers do, this chapter presents a model of active mixers for fast and accurate estimation of noise and nonlinearity. Based on closed-form expressions, this model estimates NF, IIP3 and IIP2 of the time-varying mixer by a limited number of time-invariant circuit calculations. The model shows that the decreasing transistor output resistance together with the low supply voltage in deep submicron technologies contributes significantly to flicker-noise leakage. Design insights for low flicker noise are then presented. The model also shows that the slope of the LO signal has significant effect on IIP2 while little effect on IIP3. A new IP2 calibration technique using slope tuning is presented.

5.1 **Introduction**

The active mixer is a critical building block in the RF front-end. With higher conversion gain the active mixer provides a better noise suppression of

This chapter is published in IEEE Transaction on Circuits and Systems I, vol. 58, pp. 276-289, Feb. 2011.

the subsequent stages than passive mixers. Unfortunately, the CMOS active mixer suffers more from flicker noise and nonlinearity than the passive mixer, which degrades the overall noise and linearity performance in zero-IF and low-IF receivers [1]-[5]. For circuit design insights as well as for design automation and synthesis of the RF circuits where typically iterative dimensioning loops are involved [6], a model that enables fast and accurate estimation of noise and nonlinearity is desirable. A number of papers present noise and nonlinearity analyses for the Gilbert mixer to provide design guidelines [7]-[13] or build high-level model during architectural design of RF front-ends [14]. However, for noise analyses, the transistor output resistance is typically neglected [7]–[9] or oversimplified [10]. For IIP3 calculations, [11], [13], and [14] focus on numerical calculations, while [12] neglects the periodic property of the transistor nonlinearity for IIP2. In [11]–[13], for both IIP2 and IIP3 analyses, transistor nonlinearities other than the transconductance nonlinearity are neglected. In [11], [13], and [14], the effect of the LO slope is not considered.

In this chapter, a time-varying small-signal and weakly nonlinear analysis is used, including both the output resistance and capacitances. It is shown that the output resistance effects may significantly contribute to the flicker-noise leakage and hence may make the flicker noise cancellation technique of tuning out the capacitance less effective [15], [16]. The effect of the finite LO slope on IIP3 can be neglected, while neglecting the LO rise/fall time can underestimate IIP2. Aiming for the circuit design guidelines as well as constructing an estimation model for the automatic synthesis of the active mixers, we introduce a closed-form model that properly models the output noise and nonlinearity of the active mixers. These closed-form expressions use linear interpolation between a limited number of time-invariant circuit calculations in one LO period. The noise model derived in this chapter requires data from only two ac calculations. The IIP3 model in this chapter requires one time-invariant nonlinearity calculation, while the IIP2 model requires data from a few time-invariant nonlinearity calculations. Since the time-varying mixer performance is estimated by time-invariant noise and nonlinearity calculations, this model involves no complex numerical analyses,

and it can be easily utilized by circuit designers and fast mixer design automation algorithms.

Section 5.2 introduces the fundamentals of the active mixers in deep-submicrometer technologies. Section 5.3 presents the time-varying small-signal analysis for the noise model. The impact of the transistor output resistance is investigated, and the design insights for flicker-noise leakage reduction are presented. Section 5.4 uses the time-varying weakly nonlinear analysis to derive the closed-form expressions for IIP3 and IIP2. The impact of the LO slope is analyzed for both IIP3 and IIP2, and a new IIP2 calibration technique is proposed. Section 5.5 presents the benchmarking of the accuracy for our model for the mixer operating in different bias conditions and at different frequencies. The conclusion is drawn in Section 5.6.

5.2 Active mixer in deep-submicrometer technologies

A mixer is a periodically time-varying circuit whose periodic steady state is modulated by the periodic LO signal. At any instantaneous time, the (quasi-) dc bias for the mixer is fixed, and therefore, the circuit can be linearized around this (quasi-)dc operating point. As a result, for noise analysis, the transistors within the mixer can be described by periodic small-signal parameters such as periodic transconductances, output resistances, and capacitances. For nonlinearity analyses, the transistors can be modeled by periodic weakly nonlinearities such as periodic nonlinear transconductances, output resistances, and capacitances. Note that this assumes that the transient effects are small, which is a valid simplification for mixers that operate in the low gigahertz region in modern CMOS processes. As a result of the periodic behavior, the transfer functions from the input port to the output port of the mixer can be described by periodic small-signal and weakly nonlinear properties of transistors and by time-invariant properties of passives in the circuit [17]. In the analyses in this chapter, a time-varying small-signal analysis is applied to derive a noise model, while a time-varying weakly nonlinear analysis is applied to derive a nonlinearity model. For simplicity reasons, the single-balanced Gilbert mixer shown in Fig. 5.1(a) is used for the analyses of

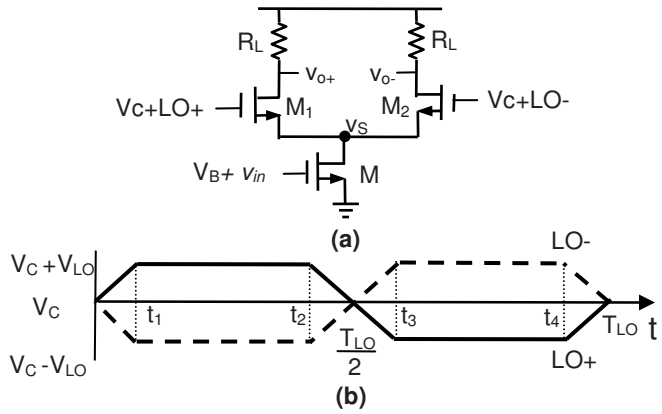


Fig. 5.1. (a) Schematic of the single-balanced Gilbert mixer. (b) Waveform of the LO signal.

noise and nonlinearity. We assume that the LO signal at the gate of the switching pair can be properly modeled by a trapezoid shown in Fig. 5.1(b).

In deep-submicrometer technologies two additional issues are considered in this chapter compared to previous work [7]-[13]:

- In deep-submicrometer CMOS, the output resistance of short transistors is relatively low so that the flicker noise contribution due to the output resistance can be as significant as that from the output capacitance. In this chapter, therefore the influence of the output resistance of M_3 on flicker-noise leakage is taken into account.
- At t=0 the gate bias of M_1 equals the common-mode voltage of the LO (V_c). Due to the low supply voltage in the submicrometer technologies, V_c can be so low that M_3 is in the triode region. Since the drain current of M_3 is small, the voltage drop across the load is small, and M_1 is generally in saturation. As the LO+ increases, M_3 gradually enters saturation region. During (0, $0.5T_{LO}$) M_1 stays in the saturation region and M_3 toggles between the triode region and the saturation region. Similarly during ($0.5T_{LO}$, T_{LO}) M_2 stays in the saturation region, and M_3 toggles between the triode region and the saturation region. In the triode region the

output conductance nonlinearity and cross-modulation nonlinearity e.g. $g_{x21} = (1/2) \times (\partial^3 I_{DS}/\partial^2 V_{GS} \partial V_{DS})$ become significant, therefore the analyses in this chapter take into account the transconductance nonlinearity as well as the output conductance nonlinearity and cross-modulation nonlinearity for IIP2 and IIP3 modeling.

5.3 Time-varying small-signal noise analysis

5.3.1 Noise model for active mixers

The flicker-noise output of the Gilbert mixer is dominantly contributed by the switch pair M_1/M_2, while transistor M_3 is causing thermal-noise folding [7]-[10]. The mixer output noise can be approximated by a stationary process [18] and therefore the output noise voltage contributed by transistor M_1, M_2 and M_3 is given by:

$$\begin{aligned} v_{n,out}^{M_k}[v_{LO}(t)] &= H^{M_k}[v_{LO}(t)] \cdot v_{n,in}^{M_k}[v_{LO}(t)] e^{j\omega_{in}t} \\ &= F_{\mathrm{n}}^{M_k}[v_{LO}(t)] e^{j\omega_{in}t}, \qquad n \in \{th, fl\} \end{aligned} \tag{5.1}$$

where $v_{n,in}^{M_k}[v_{LO}(t)]$ is the equivalent gate-referred root mean square (rms) noise voltage of transistor M_k, either flicker noise ($n = fl$) or thermal noise ($n = th$), $v_{LO}(t)$ is the LO signal and $H^{M_k}[v_{LO}(t)]$ accounts for the transfer function between the noise source to the output terminals. For an LO period T_{LO} and assuming that M_1 and M_2 are symmetric, $|H^{M_1}[v_{LO}(t)]| = \left|H^{M_2}\left[v_{LO}\left(t + (T_{LO}/2)\right)\right]\right|$. Consequently, it is sufficient to focus on $v_{n,out}^{M_1}$ and $v_{n,out}^{M_3}$ in the analyses. Because of its periodic nature, the term $F_{\mathrm{n}}^{M_k}[v_{LO}(t)]$ in (5.1) can be replaced by its Fourier series, yielding

$$v_{n,out}^{M_k} = \sum_{p=-\infty}^{+\infty} f_{p,n}^{M_k} e^{jp\omega_{LO}t} e^{j\omega_{in}t} = \sum_{p=-\infty}^{+\infty} f_{p,n}^{M_k} e^{j(p\omega_{LO}+\omega_{in})t}, \quad n \in \{th, fl\} \tag{5.2}$$

where the dc term $f_{0,n}^{M_k}$ accounts for the noise leakage (from inputs at ω_{in} to outputs at ω_{in}); the m^{th} order Fourier coefficients $f_{-m,n}^{M_k}$ account for the noise folding (from inputs at $\omega_{in} = \omega_{IF} + m\omega_{LO}$ to outputs at $\omega_{in} - m\omega_{LO} = \omega_{IF}$). As a

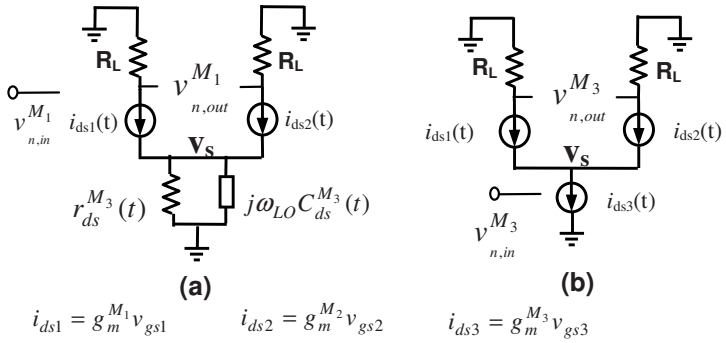


Fig. 5.2. Time-varying small signal model for calculating (a) $F_n^{M_1}[v_{LO}(t)]$ and (b) $F_n^{M_3}[v_{LO}(t)]$.

result, the output noise of the down-conversion single-balance Gilbert mixer contributed by the transistors is given by

$$S_{fl,out} = 2 \times \left|f_{0,fl}^{M_1}\right|^2 + \left|f_{0,fl}^{M_3}\right|^2 \tag{5.3}$$

$$S_{th,out} = \sum_{m=-\infty}^{+\infty} \left|f_{-\mathrm{m},th}^{M_3}\right|^2 + 2 \times \sum_{m=-\infty}^{+\infty} \left|f_{-\mathrm{m},th}^{M_1}\right|^2 \tag{5.4}$$

In these relations, $f_{0,fl}^{M_k}$ is the dc term of $F_{fl}^{M_k}$ in (5.1) for gate-referred flicker noise $V_{fl,in}^{M_k}$; $f_{-m,tn}^{M_k}$ are the Fourier series coefficients of $F_{th}^{M_k}$ in (5.1) for gate-referred thermal noise $V_{th,in}^{M_k}$.

Assuming a symmetric LO signal, for given rise/fall time, the driving signal $v_{LO}(t)$ is determined by three time instants: t_1, $0.5T_{LO}$ and t_4. The waveforms $F_n^{M_1}[v_{LO}(t)]$ and $F_n^{M_3}[v_{LO}(t)]$ at t_1, $0.5T_{LO}$ and t_4 can easily be obtained from the time-varying small signal models shown in Fig. 5.2. To avoid solving differential equations, we simplify the analysis by modeling the time-varying small signal capacitance in transistor M_k with a time-varying admittance $j\omega_{LO}C(t)$. In this section, we focus on demonstrating this time-varying noise analysis, therefore we only include the transconductance of M_1, M_2 and M_3 and the output impedance of M_3 ($Z_{\mathrm{ds}}^{M_3}(t) = r_{\mathrm{ds}}^{M_3}(t)/(1 + j\omega_{LO}C_{\mathrm{ds}}^{M_3}(t)r_{\mathrm{ds}}^{M_3}(t))$), as shown in Fig. 5.2a. Nevertheless, for highly accurate noise modeling used in the mixer design automation, all

capacitances and conductances of the transistor are taken into account. This yields:

$$F_n^{M_1}\big|_{t_1} = \left.\frac{-g_m^{M_1} R_L v_{n,in}^{M_1}}{1 + g_m^{M_1} Z_{\mathrm{ds}}^{M_3}}\right|_{t_1} \approx \left.\frac{-R_L v_{n,in}^{M_1}}{Z_{\mathrm{ds}}^{M_3}}\right|_{t_1} \tag{5.5}$$

$$F_n^{M_1}\big|_{\frac{T_{LO}}{2}} = -g_m^{M_1} R_L v_{n,in}^{M_1}\big|_{\frac{T_{LO}}{2}} \tag{5.6}$$

$$F_n^{M_1}\big|_{t_3} \approx 0 \tag{5.7}$$

$$F_n^{M_3}\big|_{t_1} = \left.\frac{-g_m^{M_1} g_m^{M_3} R_L Z_{\mathrm{ds}}^{M_3} v_{n,in}^{M_3}}{1 + g_m^{M_1} Z_{\mathrm{ds}}^{M_3}}\right|_{t_1} \approx -g_m^{M_3} R_L v_{n,in}^{M_3}\big|_{t_1} \tag{5.8}$$

$$F_n^{M_3}\big|_{\frac{T_{LO}}{2}} = 0 \tag{5.9}$$

$$F_n^{M_3}\big|_{t_3} = \left.\frac{g_m^{M_1} g_m^{M_3} R_L Z_{\mathrm{ds}}^{M_3} v_{n,in}^{M_3}}{1 + g_m^{M_1} Z_{\mathrm{ds}}^{M_3}}\right|_{t_3} \approx -F_n^{M_3}\big|_{t_1} \tag{5.10}$$

For mixers in modern deep-submicrometer CMOS and operating at frequencies of up to the lower GHz range, transient effects can be neglected. Then $F_n^{M_k}[v_{LO}(t)]$ can be sufficiently accurately approximated by interpolating between $F_n^{M_k}[v_{LO}(t_1)]$, $F_n^{M_k}[v_{LO}(0.5T_{LO})]$ and $F_n^{M_k}[v_{LO}(t_3)]$. These approximations are shown in Fig. 5.3(b) and (c).

In (t_1, t_2) M_1 and M_3 form a cascode amplifier, and in this period the drive voltage of M_1 (v_{LO}) has its maximum value. Then $F_n^{M_3}\big|_{t_1}$ is equal to the output noise voltage due to the (equivalent) input referred noise of M_3. Because of the finite output impedance of M_3 in deep-submicrometer CMOS, the noise contribution from the cascode transistor M_1, given by (5.5) cannot be neglected.

At $0.5T_{LO}$, both M_1 and M_2 are on, and they form a balanced differential pair. Then, the output impedance of M_3 has a negligibly small effect on $F_n^{M_1}$ as shown by (5.6).

In (t_3, t_4) M_2 and M_3 act as a cascode amplifier, and M_1 is off. Thus, $F_n^{M_1}$ is close to zero, and $F_n^{M_3}[v_{LO}(t)]$ is at its positive maximum. Being an odd function, $F_n^{M_3}[v_{LO}(t)]$ has no even Fourier series coefficients, and thus, for transistor M_3, the flicker noise only up-converts to sidebands around odd harmonics of the LO. The thermal noise at the sidebands around the odd

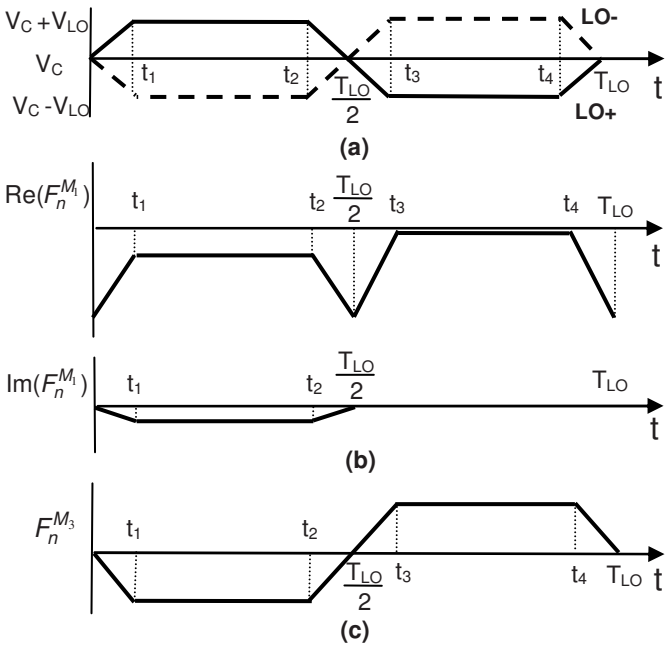


Fig. 5.3. (a) Waveform of the LO signal $v_{LO}(t)$ (b) approximation of the $F_n^{M_1}$ (c) approximation of the $F_n^{M_3}$.

harmonics of the LO frequency folds back to the IF band. As for M_1, the dc term of $F_n^{M_1}[v_{LO}(t)]$ accounts for the noise at the output without frequency translation, which corresponds to the flicker noise leakage. The thermal noise at the sidebands around harmonics of the LO frequency folds back to the IF band.

Assuming a symmetrical LO signal, with rise time and fall time equal to αT_{LO}, the time instants t_1 to t_4 can be rewritten as $t_1 = 0.5 \times \alpha T_{LO}$, $t_2 = 0.5 \times (1 - \alpha \cdot T_{LO})$, $t_3 = 0.5 \times (1 + \alpha \cdot T_{LO})$ and $t_4 = 1 - 0.5\alpha \cdot T_{LO}$. Again, under the assumption of negligibly small transient effects, this enables rewriting (5.3) and (5.4) into:

$$S_{fl,out} = 2 \times \left|f_{0,fl}^{M_1}\right|^2 = \frac{2}{T_{LO}^2} \times \left[(t_2 - t_1) \cdot F_{fl}^{M_1}\big|_{t_1} + t_1 \cdot \left(\left|F_{fl}^{M_1}\big|_{\frac{T_{LO}}{2}} - F_{fl}^{M_1}\big|_{t_1}\right|\right) + \left(t_3 - \frac{T_{LO}}{2}\right) \cdot F_{fl}^{M_1}\big|_{\frac{T_{LO}}{2}}\right]^2$$
$$= \frac{1}{2}\left(F_{fl}^{M_1}\big|_{t_1} + 2\alpha F_{fl}^{M_1}\big|_{\frac{T_{LO}}{2}} - 3\alpha F_{fl}^{M_1}\big|_{t_1}\right)^2 \quad (5.11)$$

$$S_{th,out} = \sum_{m=-\infty}^{+\infty} \left|f_{-\mathrm{m},th}^{M_3}\right|^2 + 2 \times \sum_{m=-\infty}^{+\infty} \left|f_{-\mathrm{m},th}^{M_1}\right|^2 = \frac{1}{T_{LO}} \int_{T_{LO}} \left|F_{th}^{M_3}\right|^2 dt + \frac{2}{T_{LO}} \int_{T_{LO}} \left|F_{th}^{M_1}\right|^2 dt$$
$$= \frac{3 - 4\alpha}{3}\left(F_{th}^{M_3}\big|_{t_1}\right)^2 + \frac{3 - 4\alpha}{6}\left(F_{th}^{M_1}\big|_{t_1}\right)^2 + \frac{\alpha}{6}\left[2F_{th}^{M_1}\big|_{t_1} \cdot F_{th}^{M_1}\big|_{\frac{T_{LO}}{2}} + 4\left(F_{th}^{M_1}\big|_{\frac{T_{LO}}{2}}\right)^2\right] \quad (5.12)$$

Given that the total output noise mainly consists of flicker noise from M_1 and M_2, the thermal noise from M_1, M_2 and M_3, the load R_L and input source impedance R_s, the single-side band noise figure (NF) is given by

$$NF_{SSB} = \frac{S_{fl,out} + S_{th,out} + S_{Rs,out} + S_{R_L,out}}{0.5 S_{Rs,out}} \quad (5.13)$$

where $S_{R_L,out} = 8kTR_L$. While the input source resistance R_s contributes to the output noise in the same way as the thermal noise from M_3, (5.8) and (5.12) yield $S_{R_s,out} \approx \left(1 - (4\alpha/3)\right) \cdot \left(g_m^{M_3} R_L\big|_{t_1}\right)^2 kTR_s$, with $v_{n,in}^{M_3}$ in (5.8) replaced by $\sqrt{kTR_s}$ for perfect input matching. Note that (5.11) and (5.12) use only transistor properties, bias conditions and component values at 2 distinct time instances: at t_1 and at $0.5T_{LO}$. As a result, the presented estimation for the active mixer NF can be realized by two ac calculations for trapzoid LO signals with a finite rise and fall time.

5.3.2 Impact of transistor output impedance on flicker noise

In the previous section, (5.11) indicates that the impact of the output impedance on the flicker noise leakages is described by the dc term of $F_n^{M_1}$ $(n = fl)$. By using (5.5) and (5.6), equation (5.11) is simplified to

$$S_{fl,out} \approx 0.5 \times \left[\frac{(3\alpha - 1) \cdot g_m^{M_1} R_L v_{fl,in}^{M_1}}{1 + g_m^{M_1} Z_{\mathrm{ds}}^{M_3}} \bigg|_{t_1} - 2\alpha g_m^{M_1} R_L v_{fl,in}^{M_1} \big|_{\frac{T_{\mathrm{LO}}}{2}} \right] \tag{5.14}$$

where the former term is the integral (or area) of $F_n^{M_1}$ in (t_1, t_2) shown in Fig. 5.2(b), the latter term is the integral of $F_n^{M_1}$ in (t_2, t_3), (0, t_1) and (t_4, T_{LO}). Then (5.13) can be simplified to:

1) for low IF (flicker noise dominant)

$$NF_{SSB} \approx \frac{S_{fl,out}}{0.5 S_{Rs,out}} = \frac{1}{\left(1 - \frac{4\alpha}{3}\right) KTR_s} \times \left[\frac{(3\alpha - 1) R_L v_{fl,in}^{M_1}}{g_m^{M_3} Z_{\mathrm{ds}}^{M_3}} \bigg|_{t_1} - \frac{2\alpha g_m^{M_1} v_{fl,in}^{M_1} \big|_{\frac{T_{\mathrm{LO}}}{2}}}{g_m^{M_3} \big|_{t_1}} \right]^2 \tag{5.15}$$

2) for high IF (thermal noise dominant)

$$NF_{SSB} \approx \frac{S_{th,out} + S_{Rs,out} + S_{R_L,out}}{0.5 S_{Rs,out}} = 2 \times \left[\frac{8}{\left(1 - \frac{4\alpha}{3}\right) \left(g_m^{M_3}\right)^2 R_L R_s} + \frac{4\gamma}{g_m^{M_3} R_s} \right] \bigg|_{t_1} \tag{5.16}$$

Two flicker-noise leakage mechanisms are represented by (5.14). At (0, t_1), (t_2, t_3) and (t_4, T_{LO}) the mixer acts as a differential pair. The flicker noise of the switch pair is transferred to the output just like the signal amplified by the differential amplifier. For this mechanism, the output impedance of M_3 has no effect on the flicker-noise leakage, which is shown by the second term in (5.14). At (t_1, t_2) and (t_3, t_4), one transistor in the switch pair is off, and the mixer acts as a cascode amplifier. Due to the finite output impedance of M_3, the flicker noise of the switch pair leaks to the output.

In summary, (5.14) suggests that the slope of the LO, the gain of the differential pair, and the input-referred flicker noise voltage of the switch M_1/M_2, and the output impedance of M_3 all determine the flicker-noise leakage. The following approaches can be followed to reduce the flicker noise leakage:

- Reducing the rise/fall time of the LO signal (smaller α) decreases the area of the spike in (t_2, t_3), (0, t_1) and (t_4, T_{LO})
- Choosing a wider switch pair (smaller $v_{n,fl}^{M_1}$), which comes at the price of higher LO power.
- Reducing $g_m^{M_1}\big|_{T_{\mathrm{LO}}/2}$ so that the height of the spike in (t_2, t_3) decreases. This can be realized by reducing the bias current of switch

pair at $0.5T_{LO}$ (smaller $g_m^{M_1}$) [15], [16] and [19] and by choosing a low common-mode voltage V_c for the LO. At $0.5T_{LO}$, a low V_c can force M_3 into the triode region, which reduces the DC current of M_1 and M_2, resulting in a decrease of $g_m^{M_1}|_{T_{LO/2}}$.

- Increasing the output impedance of M_3 (larger $Z_{ds}^{M_3}$)

In technologies with long-channel transistors where the output capacitance of M_3 is dominant in $Z_{ds}^{M_3}$, the output resistance $r_{ds}^{M_3}$ can be neglected [7]-[9]. To increase $Z_{ds}^{M_3}$, then inductors can be used to tune out the output capacitance for flicker noise reduction [15], [16]. However, nowadays the technology scaling offers f_T figures well above 100 GHz, while it also brings lower output resistance and lower supply voltage [20]. Neglecting the effect of $r_{ds}^{M_3}$ in deep-submicrometer technologies can yield a significant underestimation of the output flicker noise. Fig. 5.4 shows an illustration: with an ideal square-wave LO signal at 2GHz and IF@10kHz (flicker noise dominant), the calculated noise figure is compared with simulated results for different bias current. The NF is calculated using (5.13) including both $r_{ds}^{M_3}$ and $C_{ds}^{M_3}$ (cross symbol) and including only $C_{ds}^{M_3}$ (square symbol). The simulation is performed in Spectre

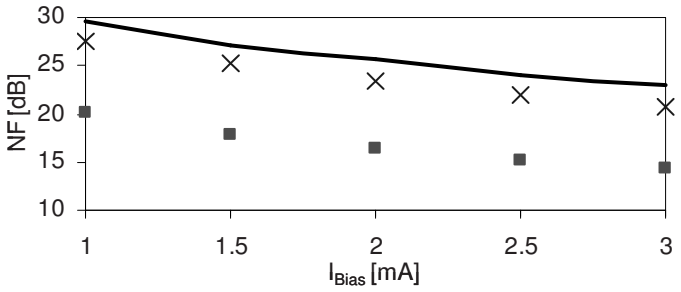


Fig. 5.4. The Gilbert mixer's NF_{SSB} (IF@10kHz) for ideal square-wave LO (a) simulated results (line); (b) calculated with r_{ds} and C_{ds} (cross); (c) calculated only with C_{ds} (square) as a function of the bias current.

for a standard 90nm CMOS process[1]. Driven by an ideal square-wave LO, the flicker-noise leakage is only caused by the finite output impedance of M_3. Including only $C_{ds}^{M_3}$, as done in [7]-[9], underestimates the flicker leakage by over 7dB compared with the model taking into account both $r_{ds}^{M_3}$ and $C_{ds}^{M_3}$. This suggests that, for deep-submicrometer CMOS technologies with low supply voltage and low output resistance, $r_{ds}^{M_3}$ is dominant in the flicker noise leakage rather than $C_{ds}^{M_3}$. Consequently, flicker noise cancelation by tuning out the output capacitance is less effective in modern deep-submicrometer processes.

5.3.3 Optimum transistor length for low flicker noise leakage

As discussed in section 5.3.2, a larger r_{ds} of M_3 reduces the flicker noise leakage. In order to keep the same power consumption and the same proportion to keep the same W/L ratio, which results in a larger $r_{ds}^{M_3}$. At the optimum, $r_{ds}^{M_3}$ is equal to $1/j\omega_{LO}C_{ds}^{M_3}$, and further increasing the W_3 and L_3 can not reduce the flicker noise leakage since the increasing $C_{ds}^{M_3}$ decreases $Z_{ds}^{M_3}$.

The length of the switching pair M_1/M_2 can also be increased to reduce the flicker noise source at the cost of larger gate-source capacitance. Note that, for short transistors, the input capacitance is composed of the intrinsic gate-source capacitance and two relatively large overlap capacitances. Since overlap capacitance is hardly affected by the transistor length, the LO power consumption will increase less than proportional to L.

For demonstration purposes, Fig. 5.5 shows simulation results for three differently dimensioned mixers:

- Mixer A: Using minimum length for M_3 and M_1/M_2 (W_3/L_3=60/0.1,W_1/L_1=W_2/L_2=106/0.1, $V_{GT}^{M_3}$=0.13)
- Mixer B: The same as mixer A except that the width and length of M_3 are doubled with respect to the mixer A implementation. (W_3/L_3 =120/0.2, W_1/L_1 =W_2/L_2 = 106/0.1, $V_{GT}^{M_3}$=0.124).
- Mixer C: The same as mixer B except that the length of transistors M_1 and M_2 is tripled with respect to the mixer B implementation. (W_3/L_3 =120/0.2, W_1/L_1 =W_2/L_2 = 106/0.3, $V_{GT}^{M_3}$=0.125).

[1]This process is used for all simulations in this chapter. The PSP compact MOSFET model [30] is used for all simulations.

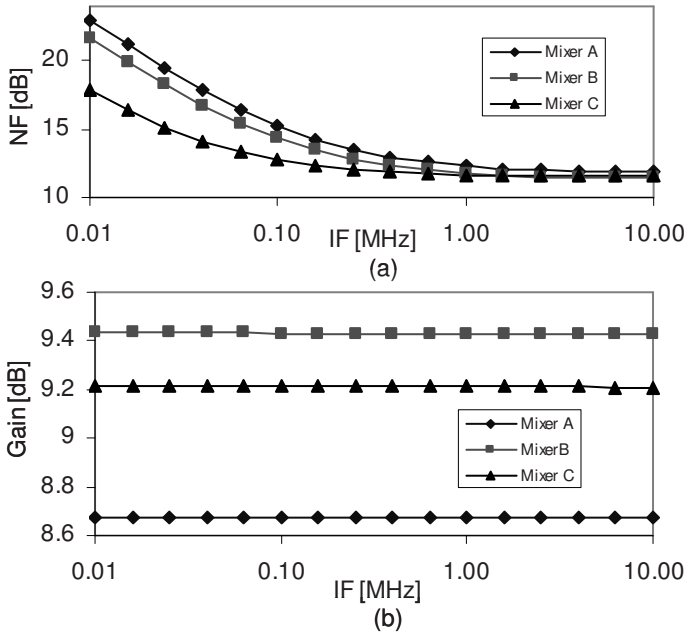


Fig. 5.5. (a) NF and (b) gain of three mixer designs with various channel length.

For the three designs, the same LO driver (f_{LO} = 2.01GHz, V_{LO}=1 V and α =0.06) is used, and the power consumption for all three mixers are set to 1.96mW. This constant power consumption implies that the biasing conditions of M_3 are a slightly different for all three implementations, which results in a small change in the transconductance of M_3. This minor change of M_3 transconductance results in a different gain of the mixer, see Fig.5.5(b). Compared to the mixer using transistors with minimum length (Mixer A), using only longer channel length in M_3 (Mixer B) can decrease NF@10kHz by 1.3dB since the larger $r_{ds}^{M_3}$ reduces the flicker noise leakage; NF@10MHz decreases by 0.4dB due to a gain increase with 0.7dB. Using longer M_3 and M_1/M_2 (Mixer C) decreases NF@10kHz by 5.2dB since the flicker noise of the switching pair decreases for longer transistor; NF@10MHz decreases by 0.3dB. Note that LO driver power consumption increases by 25% although we triple the length of the switch pair M_1/M_2.

Overall, it can be concluded that minimum length is not optimum for transistors in the active mixer to reduce flicker noise. By properly increasing W and L of M_3 and using longer M_1/M_2 the noise performance can be improved without any gain penalty but at the cost of a small increase in LO power. As a side effect, also the transistor mismatches are reduced for larger transistors, and flicker noise cancelation by tuning out $C_{ds}^{M_3}$ [15], [16] becomes effective again.

5.4 Time-varying weakly nonlinear analysis

For the analysis of the active mixer nonlinearity (IIP3 and IIP2), typically a number of simplifications are made in literature [11], [13]:

- Only taking the transconductance nonlinearity into account; this is acceptable for older CMOS technologies but is certainly not acceptable for modern short-channel RF CMOS.
- Neglecting the effect of finite LO slopes
- Calculating switching pair and input stage nonlinearities separately
- Assuming constant, bias independent, linear and nonlinear transconductances.

In this section, we use time-varying weakly nonlinear analyses explicitly including:

- the effect on IIP3 and IIP2 of all transistors' non-linear conductances; this includes the resistive nonlinearities (transconductance, output conductance and cross-modulation conductive terms describing the fact that the drain-source current is controlled by both v_{gs} and v_{ds}) and all capacitive nonlinearities (capacitance, transcapacitance and cross-modulation capacitive terms).
- the effect of the finite LO signal slope on IIP2 and IIP3.
- taking the switching pair and input stage in one circuit model so that that mutual effects are included in the analyses.
- periodic MOS transistor nonlinearities.

5.4.1 IIP3 estimation

For nonlinearity analysis, the mixer is considered as a weakly nonlinear circuit with respect to the input RF signal with the DC bias is periodically changed by the LO signal. Then, with a two-tone input signal at ω_{RF1} and ω_{RF2}, the fundamental signal and intermodulation distortions at the output of the circuit shown in Fig. 5.1 are functions of the periodic LO and thus can be extended to a Fourier series as

$$v_{HD_1,out}[v_{LO}(t)] = \sum_{p=-\infty}^{+\infty} f_p^{HD_1} e^{jp\omega_{LO}t} e^{j\omega_{RF1}t} \tag{5.17}$$

$$v_{IM_3,out}[v_{LO}(t)] = \sum_{p=-\infty}^{+\infty} f_p^{IM_3} e^{jp\omega_{LO}t} e^{j(2\omega_{RF1}-2\omega_{RF2})t} \tag{5.18}$$

where $f_p^{HD_1}$ and $f_p^{IM_3}$ are the Fourier coefficients. For low-side injection the fundamental signal and IM3 are located in the IF band at $\omega_{RF1} - \omega_{LO}$ and $2\omega_{RF1} - \omega_{RF1} - \omega_{LO}$, respectively. The magnitude of these signals is accounted for by the first order Fourier series coefficient $f_{-1}^{HD_1}$ and $f_{-1}^{IM_3}$:

$$v_{HD_1,IF} = f_{-1}^{HD_1} \tag{5.19}$$

$$v_{IM_3,IF} = f_{-1}^{IM_3} \tag{5.20}$$

As a result, the IIP3 can be obtained by using linear extrapolation for small input power P_{in}

$$IIP_3[\text{dBm}] = \frac{1}{2} \times dB\left(\frac{v_{HD_1,IF}}{v_{IM_3,IF}}\right) + P_{in}[\text{dBm}] = \frac{1}{2} \times dB\left(\frac{f_{-1}^{HD_1}}{f_{-1}^{IM_3}}\right) + P_{in}[\text{dBm}] \tag{5.21}$$

The exact waveforms of $v_{HD_1,out}[v_{LO}(t)]$ and $v_{IM_3,out}[v_{LO}(t)]$ can be obtained from simulations. Fig. 5.6 shows the real part of the simulated waveforms for the mixer with the gate bias of M_1 and M_2 changed according to the wave-form of LO signal shown in Fig. 5.3a. The two-tone RF signals at the gate of M_3 are at 1.01 GHz and 1.014 GHz with 1mV amplitude. Note that the waveform of $v_{HD_1,out}[v_{LO}(t)]$ is the same as $F_n^{M_3}[v_{LO}(t)]$ that is approximated in Fig. 5.3c. Therefore (5.8)-(5.10) are reused for the approximation of$v_{HD_1,out}[v_{LO}(t)]$.

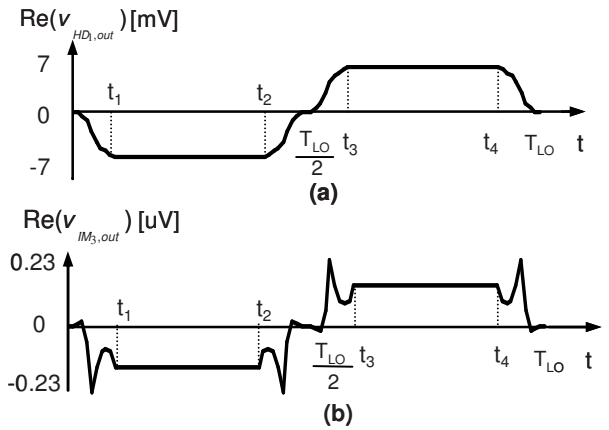


Fig. 5.6. Waveform of (a) $v_{HD_1,out}[v_{LO}(t)]$ and (b) $v_{IM_3,out}[v_{LO}(t)]$.

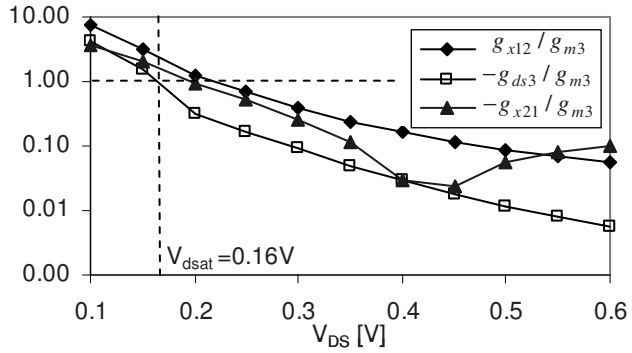


Fig. 5.7. $-g_{ds3}/g_{m3}$, $-g_{x21}/g_{m3}$ and g_{x12}/g_{m3} for an NMOS transistor in logarithm scale, W/L = 60um/0.1um, V_{GT}=0.19V as a function of the drain-source voltage V_{DS}.

At (t_1, t_2), M_1 and M_3 act as a cascode amplifier, while at (t_3, t_4) M_2 and M_3 act as a cascode amplifier and thus $v_{HD_1,out}[v_{LO}(t)]$ reaches the negative and positive maximum respectively, see Fig. 5.6(a). At (0, t_1) and (t_2, 0.5T_{LO}), M_2 is assumed to be off, M_1 stays in saturation region while M_3 may toggle between the triode and saturation region. Fig. 5.7 shows the ratio between the third-order transconductance nonlinearity g_{m3}, output conductance g_{ds3} and

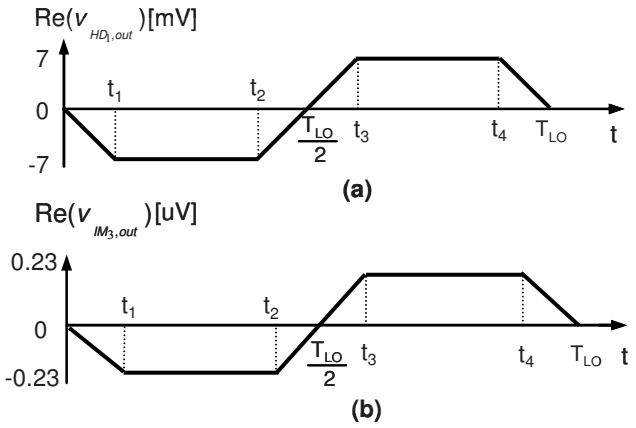


Fig. 5.8. Approximation of (a) $v_{HD_1,out}[v_{LO}(t)]$ (b) $v_{IM_3,out}[v_{LO}(t)]$.

cross-modulation nonlinearity ($g_{x21} = (1/2) \times \left(\partial^3 I_{DS}/\partial {V_{GS}}^2 \partial V_{DS}\right)$ and $g_{x12} = (1/2) \times \left(\partial^3 I_{DS}/\partial {V_{DS}}^2 \partial V_{GS}\right)$). In the triode region, the cross-modulation nonlinearity and output conductance nonlinearity are dominant, while in the saturation region the transconductance nonlinearity is dominant. Therefore, an IM3 peak occurs at (0, t_1) and (t_2, $0.5T_{LO}$) when the transistor is well in the triode region, see Fig. 5.6(b). The same waveform can be seen at ($0.5T_{LO}$, t_3) and (t_4, T_{LO}) when M_1 is assumed to be off, M_3 change from the triode region to saturation gradually.

Equation (5.20) and (5.19) indicate that the first-order Fourier components $f_{-1}^{HD_1}$ and $f_{-1}^{IM_3}$ of $v_{HD_1,out}[v_{LO}(t)]$ and $v_{IM_1,out}[v_{LO}(t)]$ determine the IIP3. Note that sharp details of a signal are mainly caused by its high-order Fourier series coefficients [21]: for the estimation of only the first order Fourier coefficient the trapezoidal waveform in Fig. 5.8[1]. Now using(5.8), (5.19)-(5.21) the voltage conversion gain is:

[1] Simulations show that the difference between the first-order Fourier series component of the original waveform and that of the approximated waveform typically is smaller than 2%.

$$V_{gain} = \frac{f_{-1}^{HD_1}}{V_{IN}} = \frac{2sin(\alpha\pi)}{\pi^2\alpha} \cdot \frac{g_m^{M_1} g_m^{M_3} R_L Z_{\mathrm{ds}}^{M_3}}{1 + g_m^{M_1} Z_{\mathrm{ds}}^{M_3}}\Bigg|_{t_1} \tag{5.22}$$

and the IIP3 can be written as: a sufficiently accurate approximation of the waveforms in Fig. 5.6 is

$$\begin{aligned} IIP_3[\mathrm{dBm}] &= \frac{1}{2} \times dB\left(\frac{f_{-1}^{HD_1}}{f_{-1}^{IM_3}}\right) + P_{in}[\mathrm{dBm}] \\ &\approx \frac{1}{2} \times dB\left(\frac{\frac{2sin(\alpha\pi) \cdot v_{HD_1,out}\big|_{t_1}}{\pi^2\alpha}}{\frac{2sin(\alpha\pi) \cdot v_{IM_3,out}\big|_{t_1}}{\pi^2\alpha}}\right) + P_{in}[\mathrm{dBm}] \\ &= \frac{1}{2} \times dB\left(\frac{v_{HD_1,out}}{v_{IM_3,out}}\Big|_{t_1}\right) + P_{in}[\mathrm{dBm}] \\ &= IIP_3|_{t_1} \end{aligned} \tag{5.23}$$

where (see the appendix for a first order derivation):

$$\begin{aligned} v_{IM_3,out} \approx \frac{-3 \times {V_{IN}}^3 R_L}{4 \times \left(1 + g_{m1}^{M_1} Z_{\mathrm{ds}}^{M_3}\right)} \times \Big[& g_{m1}^{M_1} Z_{\mathrm{ds}}^{M_3} \left(g_{m3}^{M_3} - g_{x21}^{M_3}\right) + g_{m3}^{M_3} - g_{ds3}^{M_1} \left(g_{m1}^{M_3} R_L\right)^3 \\ & - g_{x21}^{M_1} g_{m1}^{M_3} R_L + g_{x12}^{M_1} \left(g_{m1}^{M_3} R_L\right)^2 \Big] \end{aligned} \tag{5.24}$$

Eq. (5.24) indicates that other third-order nonlinearities besides g_{m3} can contribute to the output IM3 significantly. Fig. 5.7 shows that the nonlinearity g_{m3} and g_{x12} have the same sign (positive) while g_{ds3} and g_{x21} have the opposite sign (negative). Thus, all nonlinearity terms with their weighting factor in (5.24) have a positive value, which shows that the contributions of each third-order nonlinearity to the IM3 add up. For low supply voltages, LO signals with large swing can easily drive M_1 out of saturation region at t_1. Then, the output conductance nonlinearity g_{ds3} and the cross-modulation nonlinearity g_{x21} and g_{x12} of M_1 increase dramatically, which increases the $v_{IM_3,out}\big|_{t_1}$, thus decreases IIP3 as indicated by (5.23)-(5.24).

In summary, the IIP3 of the time-varying mixer can be estimated by one time-invariant IIP3 calculation at the maximum of the LO signal. The effect of the slope of the LO signal on IIP3 can be neglected. In low supply voltage processes, for high IIP3 an LO signal with a large swing is not desirable

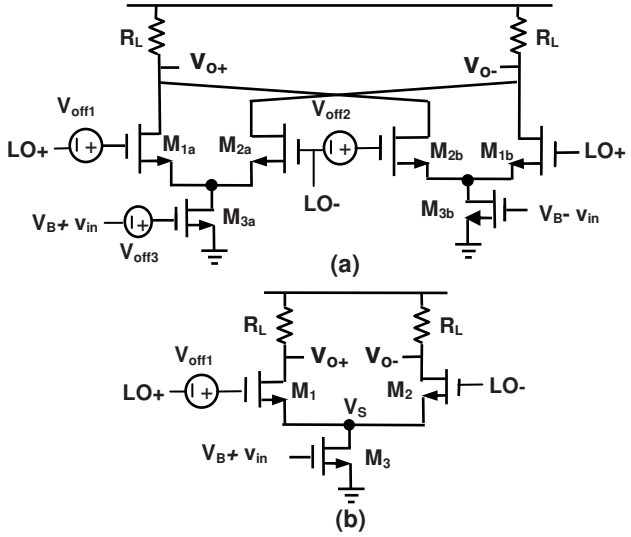


Fig. 5.9. (a) Double balanced mixer with offset voltages for modeling the transistor mismatches. (b) Single balanced mixer with offset voltages for modeling the transistor mismatches.

because the switching transistor enters into triode when LO reaches its maximum.

5.4.2 IIP2 estimation

Mismatches in transistors and load resistors, self-mixing and transistor nonlinearity together cause finite IIP2 for the balanced mixer [12]. The effect of self-mixing and mismatches in load resistors can be made negligible using layout counter-measures [22]. Then, the remaining dominant factors for IIP2 are transistor mismatches and transistor nonlinearities. For the double balanced mixer, transistor mismatch can be modeled by three DC offset voltages shown in Fig. 5.9(a), ($V_{off,1}$ for the mismatch of M_{1a}/M_{2a}, $V_{off,2}$ for the mismatch of M_{1b}/M_{2b} and $V_{off,3}$ for the mismatch of M_{3a}/M_{3b}). Since the effect of the switch pair mismatch is typically much larger than that of the transconductors [12], we will neglect $V_{off,3}$ and use the single balanced mixer shown in Fig. 5.9b in this analysis.

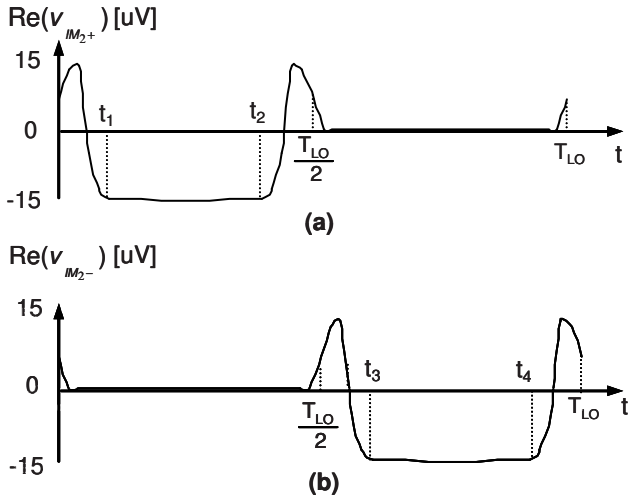


Fig. 5.10. Waveform of (a) $v_{IM_2,+}[v_{LO}(t)]$ and (b) $v_{IM_2,-}[v_{LO}(t)]$.

As explored in the previous section, the mixer is considered as a nonlinear circuit with respect to the input RF signal where the DC bias is changed periodically by the LO signal. With DC offset, the LO is not symmetric and therefore the single-ended IM2 at the positive and the negative outputs are not equal and hence will not cancel. As a function of the asymmetric LO signal, for a two-tone input signal at ω_{RF_1} and ω_{RF_2}, the IM2 at the differential output can be extended to a Fourier series as

$$\begin{aligned} v_{IM_2,out}[v_{LO}(t)] &= v_{IM_2+}[v_{LO}(t)] - v_{IM_2-}[v_{LO}(t)] \\ &= \left(\sum_{p=-\infty}^{+\infty} f_{p,v+}^{IM_2} e^{jp\omega_{LO}t} - \sum_{p=-\infty}^{+\infty} f_{p,v-}^{IM_2} e^{jp\omega_{LO}t} \right) \cdot e^{j(\omega_{RF_1}-\omega_{RF_2})t} \end{aligned} \tag{5.25}$$

For low-side injection the IM2 distortion is located at $\omega_{RF_1} - \omega_{RF_2}$ in the IF band, which is accounted by the 0th order Fourier series component $f_{0,v+}^{IM_2}$ and $f_{0,v-}^{IM_2}$:

$$v_{IM_2,out} = f_{0,v+}^{IM_2} - f_{0,v-}^{IM_2} \tag{5.26}$$

Then the IIP2 is

$$IIP_2[\text{dBm}] = dB\left(\frac{v_{HD_1,out}}{v_{IM_2,out}}\right) + P_{in}[\text{dBm}] = dB\left(\frac{f_{-1}^{HD_1}}{f_{0,v+}^{IM_2} - f_{0,v-}^{IM_2}}\right) + P_{in}[\text{dBm}] \qquad (5.27)$$

Using the same approach as for IM3, the real part of simulated waveforms of the single-ended IM2, $v_{IM_2,+}[v_{LO}(t)]$ and $v_{IM_2,-}[v_{LO}(t)]$, are shown in Fig. 5.10. Note that single-ended IM2 at the positive and negative output ($f_{0,v+}^{IM_2}$ and $f_{0,v-}^{IM_2}$) correspond to the dc term of $v_{IM_2,+}[v_{LO}(t)]$ and $v_{IM_2,-}[v_{LO}(t)]$. These dc-terms, in turn, correspond to the integral of (or area below) the waveform. In a perfectly symmetric mixer, the single-ended IM2 at the positive and negative output ($f_{0,v+}^{IM_2}$ and $f_{0,v-}^{IM_2}$) are equal, and therefore, exactly cancel each other, leading to an infinite IIP2. However, any DC offset introduces an effectively asymmetric LO, an asymmetric bias modulation in (0, $0.5T_{LO}$) and ($0.5T_{LO}$, T_{LO}), hence results in waveform differences between $v_{IM_2,+}[v_{LO}(t)]$ and $v_{IM_2,-}[v_{LO}(t)]$. The single-ended IM2 at the positive and negative output do not exactly cancel, which results in a finite IIP2.

Due to high similarity between $v_{IM_2,+}[v_{LO}(t)]$ and $v_{IM_2,-}[v_{LO}(t)]$, we choose to show details on the estimation of single-ended IM2 at the positive output $f_{0,v+}^{IM_2}$. As discussed in section 5.2, in (0, $0.5T_{LO}$) M_1 and M_3 act as a cascode amplifier. Due to the low supply voltage in deep-submicrometer technologies, M_1 stays in the saturation region and M_3 may toggle between triode region and saturation region. As derived in the appendix, transistor M_3 dominantly contributes to the IM2 of the cascode amplifier:

$$v_{IM_2} \approx \frac{g_{m1}^{M_3} r_{\text{ds}}^{M_3} R_L {V_{IN}}^2}{1 + g_{m1}^{M_3} r_{\text{ds}}^{M_3}} \times \left[-g_{m2}^{M_3} - g_{ds2}^{M_3} \cdot \left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)^2 + g_{x11}^{M_3} \cdot \frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right] \qquad (5.28)$$

where $r_{\text{ds}}^{M_3}$ and $g_{m1}^{M_3}$ are the linear output resistance and transconductance of M_3; $g_{m2}^{M_3} = (1/2) \times \left(\partial^2 I_{DS}/\partial {V_{GS}}^2\right)$ is the derivative of the transconductance; $g_{ds2}^{M_3} = (1/2) \times \left(\partial^2 I_{DS}/\partial {V_{DS}}^2\right)$ is the derivative of the output conductance; $g_{x11}^{M_3} = (\partial^2 I_{DS}/\partial V_{GS} \partial V_{DS})$ is the second-order cross-modulation nonlinearity. Equation (5.28) indicates that the sign of $v_{IM_{2+}}$ for different bias is determined by $g_{m2}^{M_3}$, $g_{ds2}^{M_3}$ and $g_{x11}^{M_3}$. As an example, an NMOS transistor (W/L = 60µm/0.1µm) with V_{GS} fixed is simulated by sweeping V_{DS} from 0.02V to 0.32V. Fig. 5.11

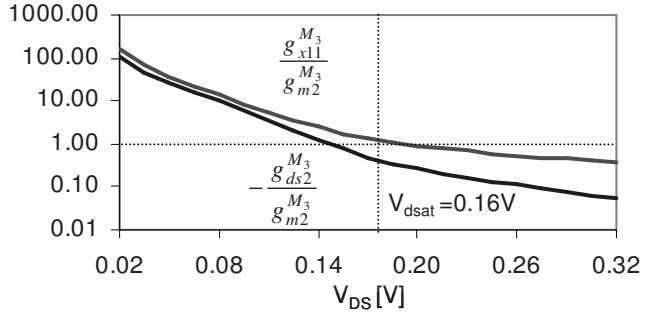


Fig. 5.11. $-g_{ds2}^{M_3}/g_{m2}^{M_3}$ and $g_{x11}^{M_3}/g_{m2}^{M_3}$ in logarithmic scale for NMOS transistor in triode and saturation region.

shows that, in the saturation region, the transconductance nonlinearity ($g_{m2}^{M_3}$) is dominant, while in the triode region, the cross-modulation nonlinearity $g_{x11}^{M_3}$ and the output conductance nonlinearity $g_{ds2}^{M_3}$ become dominant. As LO rises and falls, M_3 may enter in the triode region where the cross-modulation nonlinearity and output conductance nonlinearity ($g_{x11}^{M_3}$ and $g_{ds2}^{M_3}$) are dominant. Then the term $-g_{ds2}^{M_3} \cdot \left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)^2 + g_{x11}^{M_3} \cdot \frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}$ in (5.28) is dominant and v_{IM_2+} is positive as shown in Fig. 5.10a; as LO increases M_3 enters into the saturation where $g_{m2}^{M_3}$ becomes dominant. Since $g_{m1}^{M_1}$ and $g_{m1}^{M_3}$ are in the same order of magnitude, the term $-g_{m2}^{M_3}$ in (5.28) is dominant and v_{IM_2+} turns negative.

Note that the change of v_{IM_2+} between a positive and a negative value during the LO rise and fall times have not been considered in [12] due to the following simplifications:

- Only the transconductance nonlinearity of the transistor is considered. As a result, the fact that the nonlinearity of transistor M_3 is also modulated by its drain-source voltage is neglected.
- The effect of the finite LO slope on IIP2 is neglected. As a result, the fact that due to low supply voltage in the deep-submicron technologies the transistor M_3 typically toggles between the triode and saturation region during the LO rise and fall time is neglected.

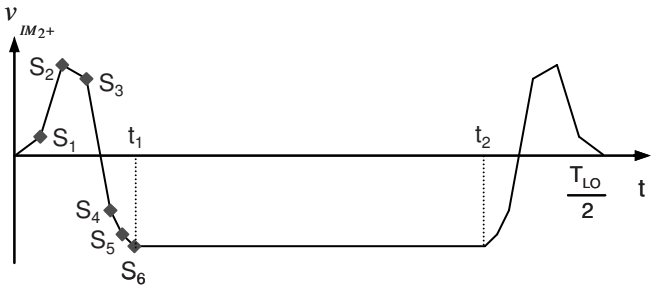


Fig. 5.12. Estimation of $v_{IM_2,+}[v_{LO}(t)]$.

Since the positive single-ended IM2 $f_{0,v+}^{IM_2}$ is equal to the integral of the waveform of v_{IM_2+} shown in Fig. 5.10a, neglecting the positive area in (0, t_1) and (t_2, $0.5T_{LO}$) can overestimate the positive single-ended IM2 $f_{0,v+}^{IM_2}$. The same conclusion applies to the negative single-ended IM2 $f_{0,v-}^{IM_2}$. In summary, the single-ended IM2 can be overestimated by neglecting the LO slope, cross-modulation nonlinearity and output conductance nonlinearity of the transistor. As a result, the differential IM2 ($f_{0,v+}^{IM_2} - f_{0,v-}^{IM_2}$) can be misestimated significantly.

In order to give an accurate estimation of the positive single-ended IM2 $f_{0,v+}^{IM_2}$, a good capture of the waveform, especially in (0, t_1) and (t_2, $0.5T_{LO}$), is essential. Fig. 5.12 demonstrates the estimation of $f_{0,v+}^{IM_2}$ by six equal-distant samples (S_1 to S_6). Assuming that the rise/fall time of the LO are equal to αT_{LO} yields $t_1 = 0.5\alpha T_{LO}$ and $t_2 = 0.5(1-\alpha)T_{LO}$, the area of $v_{IM_2,+}[v_{LO}(t)]$ then is given by

$$\begin{aligned} f_{0,v+}^{IM_2} &= \frac{t_1}{3T_{LO}} \cdot \left[S_1 + S_2 + S_3 + S_4 + S_5 + \frac{S_6}{2}\right] + \frac{t_2 - t_1}{T_{LO}} \cdot S_6 \\ &= \frac{\alpha}{6} \cdot \left[v_{IM_2+}\big|_{v_{LO+}^1} + v_{IM_2+}\big|_{v_{LO+}^2} + v_{IM_2+}\big|_{v_{LO+}^3} + v_{IM_2+}\big|_{v_{LO+}^4} + v_{IM_2+}\big|_{v_{LO+}^5} \right. \\ &\qquad \left. + 0.5v_{IM_2+}\big|_{v_{LO+}^6}\right] + (0.5-\alpha)v_{IM_2+}\big|_{v_{LO+}^6} \end{aligned} \tag{5.29}$$

where $v_{LO+}^k = V_c + (k/6)V_{LO} + V_{off}$

Similarly, the area of $v_{IM_2,-}[v_{LO}(t)]$ can be estimated as

$$f_{0,v-}^{IM_2} = \frac{\alpha}{6} \cdot \left[v_{IM_2-}\Big|_{v_{LO-}^1} + v_{IM_2-}\Big|_{v_{LO-}^2} + v_{IM_2-}\Big|_{v_{LO-}^3} + v_{IM_2-}\Big|_{v_{LO-}^4} + v_{IM_2-}\Big|_{v_{LO-}^5} + 0.5 v_{IM_2-}\Big|_{v_{LO-}^6} \right] + (0.5-\alpha) v_{IM_2-}\Big|_{v_{LO-}^6} \quad (5.30)$$

where $v_{LO-}^{\mathrm{k}} = V_c + (k/6)V_{LO}$.

Now, by using (5.27), (5.29) and (5.30), the IIP2 of the time-varying mixer can be estimated by a few time-invariant IM2 calculations. Note that the estimation of the single-ended IM2 by samples at different instants includes the periodic property of the transistor nonlinearity. For each instant sample, the IM2 is calculated by (5.31) as given in the appendix:

$$v_{IM_2,+} \approx \frac{R_L {V_{IN}}^2}{1 + g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3}} \cdot \{ g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3} \cdot \left[-g_{m2}^{M_3} - g_{ds2}^{M_3} \cdot \left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}} \right)^2 + g_{x11}^{M_3} \cdot \frac{g_{m1}^{M_3}}{g_{m1}^{M_1}} \right] - g_{m2}^{M_1} - g_{ds2}^{M_1} \cdot \left(g_{m1}^{M_3} R_L \right)^2 + g_{x11}^{M_3} \cdot g_{m1}^{M_3} R_L \} \quad (5.31)$$

Firstly, it shows that, for narrowband IM2, the nonlinear capacitance can be neglected. It also shows that the terms with the cross-modulation nonlinearity and output conductance nonlinearity (g_{x11} and g_{ds2}) can cancel the terms with transconductance nonlinearity g_{m2}. Due to the low supply voltage in deep-submicrometer technologies, LO signals with large swing can easily drive M_1 out of saturation region at t_1. Then, at (t_1, t_2), M_1 may stay in between triode and saturation region, where g_{x11} and g_{ds2} becomes larger while M_3 stays in saturation. In that case, (5.31) can be simplified to

$$v_{IM_2,+} \approx \frac{R_L {V_{IN}}^2}{1 + g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3}} \cdot \left[-g_{m2}^{M_3} g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3} - g_{m2}^{M_1} - g_{ds2}^{M_1} \left(g_{m1}^{M_3} R_L \right)^2 + g_{x11}^{M_1} g_{m1}^{M_3} R_L \right] \quad (5.32)$$

as given in the appendix. With the scaling factor $\left(g_{m1}^{M_3} R_L\right)^2$ and $g_{m1}^{M_3} R_L$ for $g_{ds2}^{M_1}$ and $g_{x11}^{M_1}$ the single-ended IM2 can be very small. However, due to the high sensitivity of g_{x11} and g_{ds2} to dc offset voltages (mismatches), low single-ended IM2 does not guarantee a high differential IIP2. Fig. 5.13 illustrates this: the single-balanced mixer shown in Fig. 5.9b is simulated for varying gate bias of M3 with fixed dimensions. The single-ended IIP2 was derived for a situation without mismatch, while the differential IIP2 is the minimum IIP2 from a Monte Carlo mismatch simulation. As the gate-bias of M3 increases, more current flow through R_L. Thus, at (t_1, t_2), M1 may enter the triode region,

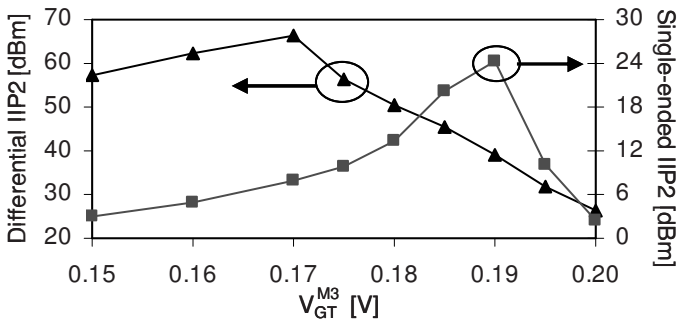


Fig. 5.13. Single-ended IIP2 and minimum differential IIP2 of the single-balanced mixer for various V_{GT} of the transistor M3.

where the distortion from g_{x11} and g_{ds2} of M1 increases and cancels a larger part of the distortion from g_{m2} of M3. This results in smaller negative area of $v_{IM_2,+}$, shown in Fig. 5.12, and yields a high single-ended IIP2. However, high differential IIP2 is achieved for smaller $V_{GT}^{M_3}$. At such bias M1 and M3 are all in saturation region in (t_1, t_2), where g_{m2} of M3 is dominant for single-ended IM2 and less sensitive to the dc offset voltages.

5.4.3 Impact of LO signal on mixer nonlinearity

The LO signal of the mixer practically has finite rise and fall time. However, the influence of LO slope on the mixer nonlinearity has not been investigated yet in previous literature [11]-[14]. The analysis in section 5.4.1 and 5.4.2 shows that, for low supply voltage, as the LO rises or falls, the transistor M_3 may experience deep triode region operation, where the cross-modulation nonlinearity and output conductance nonlinearity become dominant. As discussed in section 5.4.1, the IM3 output at the IF band is equal to the first order Fourier coefficient $f_{-1}^{IM_3}$ of $v_{IM_3,out}[v_{LO}(t)]$, which is under little influence of the LO slope. Therefore, we conclude that the LO slope effect on the mixer IIP3 can be neglected. The IIP3 of the time-varying system can be estimated by one time-invariant nonlinearity calculation.

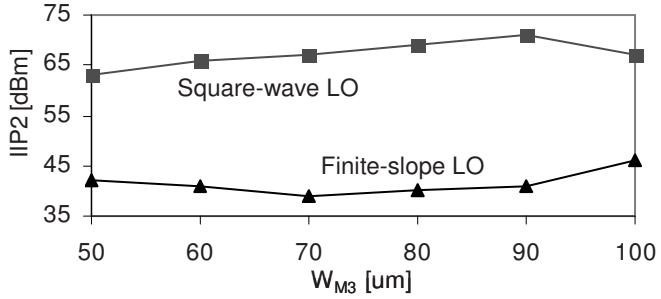


Fig. 5.14. Simulated IIP2 for (a) square-wave LO in square symbol and (b) LO with finite slope (t_{rise}= t_{fall} =0.08T_{LO}) in triangle symbol.

As for the IIP2, during the rise and fall time, M_3 toggles between the triode and saturation region. In the triode region, the cross-modulation nonlinearity and output conductance nonlinearity of the transistor are dominant, and they will result in positive single-ended IM2. In the saturation region, the transconductance nonlinearity is dominant, and the single-ended IM2 changes to negative value. Since the overall single-ended IM2 is the sum of the positive and negative contributions in one period, neglecting the LO slope can overestimate the single-ended IM2, and this may misestimate the differential IM2. Fig. 5.14 shows an illustration: the mixer shown in Fig. 5.9b is simulated by sweeping the width and V_{GT} of M_3 with fixed P_{dc}=2mW at 2 GHz. A fixed 5 mV dc offset is used to model the mismatch of the switch pair. The simulated IIP2 for a square-wave LO and a LO with finite slopes (t_{rise}= t_{fall} =0.08T_{LO}) are compared. The difference of IIP2 between using a square-wave LO and using LO with finite slope can be as large as 30dB, which demonstrates the importance of including LO slope in the IIP2 estimation.

5.4.4 LO slope tuning for IIP2 calibration

Including LO slope in the analysis not only provides more accurate estimation on the IIP2 but also shows one new possibility of introducing intentional mismatch that can be used for IIP2 calibration. In order to achieve high IIP2 typically mismatches are introduced to the mixer for neutralizing

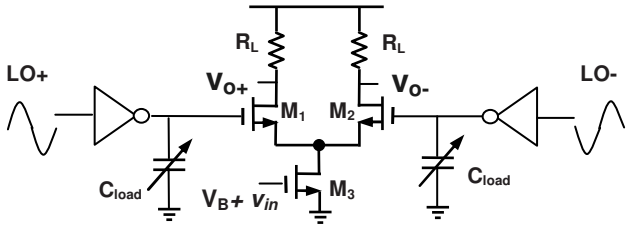


Fig. 5.15. Schematic of the mixer using LO slope tuning for IP2 calibration.

the differential IM2 output caused by the intrinsic mismatches. Currently, the possibilities for introducing intentional mismatches are:

- Controlling the mismatch between the loads by resistors trimming [23], [24] or tuning the pMOS load with NF degradation by 1-2 dB due to the noise introduced by the extra pMOS current sources [25].
- Tuning the current sources within the CMFB section for current mode output mixers [26].
- Tuning the DC level of the LO signal [27].

The discussion in section 5.4.2 suggests that the IIP2 can also be calibrated by tuning the LO slope. For demonstration, the mixer shown in Fig. 5.15 is simulated for f_{LO} at 3.01 GHz and two-tone signals at 3.02 GHz and 3.024 GHz with -25 dBm input power. The mixer is driven by two inverters with load capacitors. By changing these load capacitors we can tune the LO slope for the mixer. Note that very small LO slope change is sufficient due to the high sensitivity of the differential IM2 to the LO slope. In this case the slope change is within -0.75% to +0.75% shown in Fig. 5.16c while the inverter power dissipation is changing between 2.2mW and 2.7mW. Fig. 5.16 shows that high IIP2 can be achieved by tuning the LO slope while the gain, IIP3 and NF are not affected.

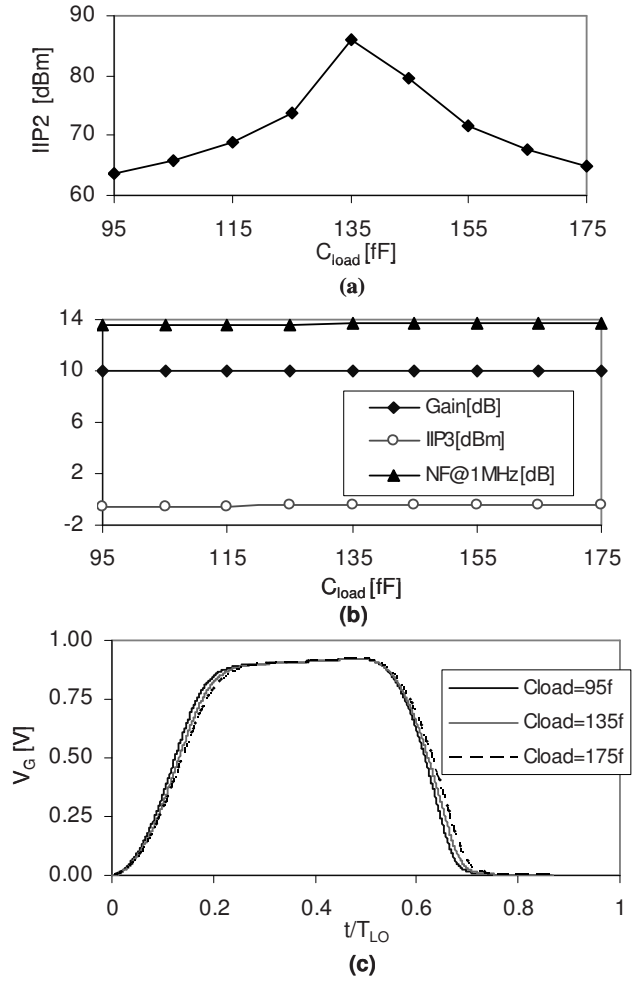


Fig. 5.16. (a) IIP2, (b) Gain, IIP3 and SSB NF at 1MHz, versus inverter load capacitor tuning and (c) waveform of LO- at the gate of M2 for different tuning capacitors.

5.4.5 Summary

It can be concluded that, by using the time-varying weakly nonlinear analysis, the IIP3 and IIP2 of the mixer can be estimated by a few time-invariant weakly nonlinearity calculation, where the effect of LO slope is

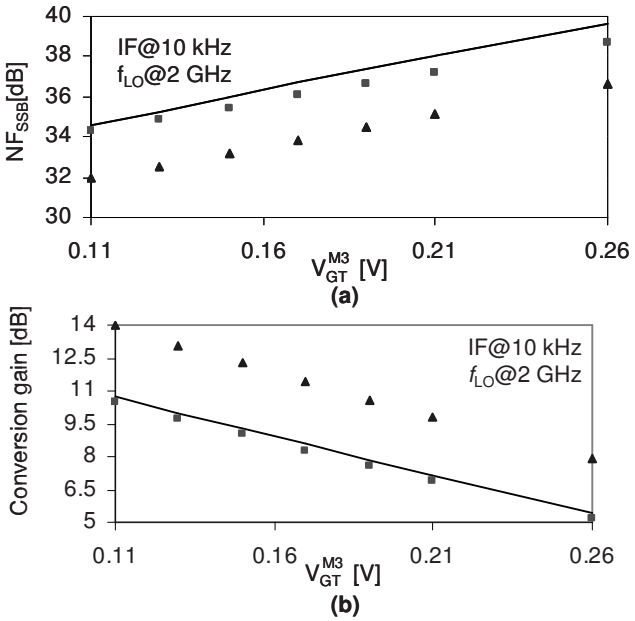


Fig. 5.17 The Gilbert mixer's (a) NFSSB and (b) conversion gain. Simulated (line) and model with rds (squares) and without rds (triangles) as a function of the overdrive voltage of M3.

included. Note that, in the time-invariant nonlinearity calculations, the contribution of the switching pair (M1/M2) and the input stage M3 is evaluated as a whole circuit but not separately as in [12]. As a result, the nonlinearity of the time-varying circuit can be estimated by time-invariant nonlinearity calculations, which is straightforward by using Volterra series approach [28] or the general weak nonlinearity model for amplifiers [29].

5.5 **Benchmarking the accuracy**

To evaluate the accuracy of the model for noise and IIP3 calculation, the single-balanced mixer in Fig. 5.1 is simulated. For the model of IIP2, the double-balanced mixer shown in Fig. 5.9a is simulated. The simulation results in Spectre and the calculation results using our model are presented in this section. We implemented the noise and nonlinearity model within a mixer P-cell similar to what we did for a LNA [6], where all small-signal

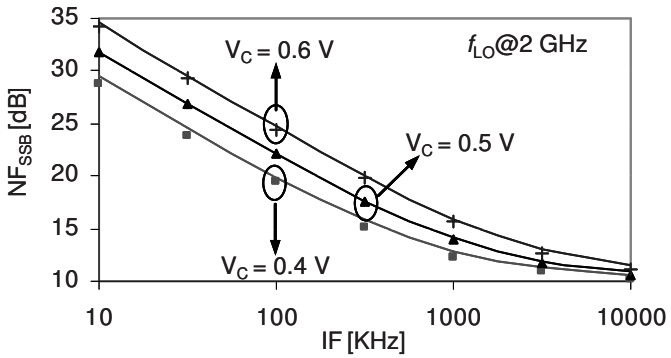


Fig. 5.18. The Gilbert mixer's NF_{SSB} as a function of the IF frequency, for three values of V_C; Simulated (line) and model (symbol) for V_C =0.4V, V_C =0.5V and V_C =0.6V.

parameters and nonlinearities of the transistor are included. This mixer P-cell dimensions a given circuit topology for a given set of specifications. The time-invariant nonlinearity calculation for IIP2 and IIP3 estimation is performed by using the circuit nonlinearity model [29], where all the resistive and capacitive nonlinearities are included.

For Fig. 5.17, the Gilbert mixer was dimensioned at 3mW power consumption with f_{LO} at 2GHz and IF at 10kHz. At this low IF, the flicker noise is dominant. For the LO signal, V_C=0.6V, V_{LO}=0.5V and α =0.1. Fig. 5.17 shows the SSB noise figure and the conversion gain as a function of gate-overdrive voltage of M_3. It is shown in Fig. 5.17a that the noise model with output resistance and capacitance (square symbols) has an estimation error smaller than 0.9dB, while the error is 3dB for the noise model with output capacitance but without output resistance (triangular symbol). Fig. 5.17b shows that the conversion gain resulting from our model with the output resistance has an estimation error smaller than 0.3dB, while the error is more than 2.5dB if the output resistance of M_3 is neglected. The analyses in section 5.3.1 suggest that with the scaling of CMOS technology --- that have lower supply voltage, higher f_T and lower output resistance --- the flicker noise

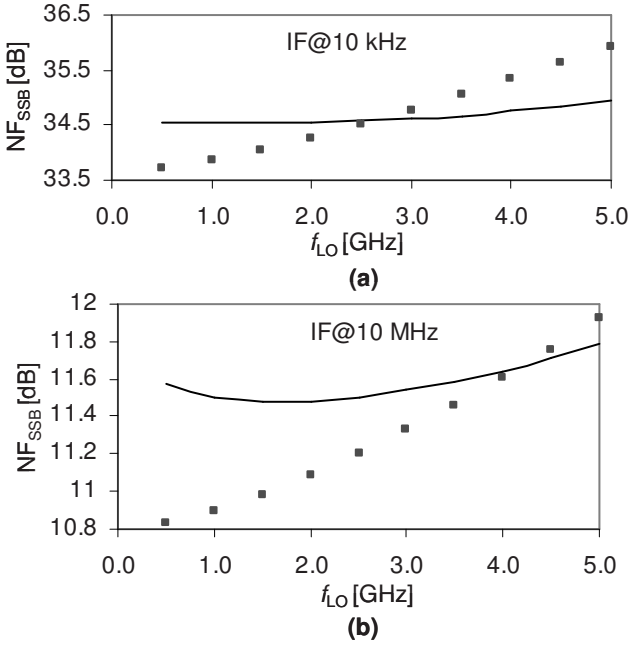


Fig. 5.19. The Gilbert mixer's (a) NF_{SSB} for IF=10kHz and (b) NF_{SSB} for IF=10MHz as a function of the LO frequency. Simulated (line) and model (square symbols) for f_{LO} =2GHz, P_{dc} =3mW, V_C =0.6V, V_{LO} =0.5V and α =0.1.

leakage caused by the finite output resistance of M_3 becomes significant and cannot be neglected.

Fig. 5.18 shows the simulated and calculated SSB noise figure as a function of the IF frequency, for f_{LO} at 2GHz, P_{dc}=3mW, V_{LO}=0.5V and α =0.1, for V_C=0.4V, V_C=0.5V and V_C=0.6V. The estimation error of our noise model is smaller than 0.3dB. As the mixer acts as a balanced differential pair at $0.5T_{LO}$ a lower common mode level of the LO signal, $V_{C,}$ causes lower gain for the differential pair and thus smaller noise spikes shown in Fig. 5.2b. As a result, the flicker noise leakage is smaller for lower V_C.

Fig. 5.19 shows a comparison of simulated and calculated NF_{SSB} both at

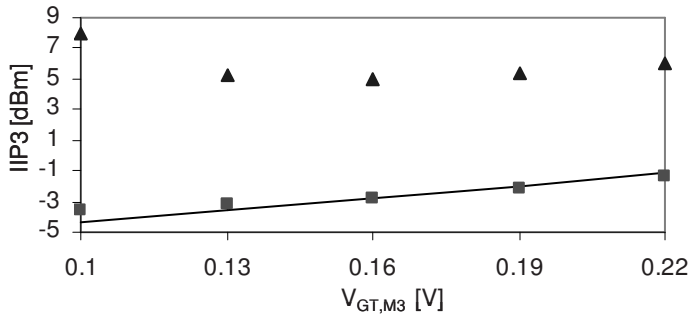


Fig. 5.20. The Gilbert mixer's IIP3 as a function of the overdrive voltage of M_3. Simulated (line), model including all nonlinearity (squares) and model including only g_m nonlinearity (triangular) for P_{dc} =3mW, V_C =0.3V, V_{LO} =0.6V and α =0.1.

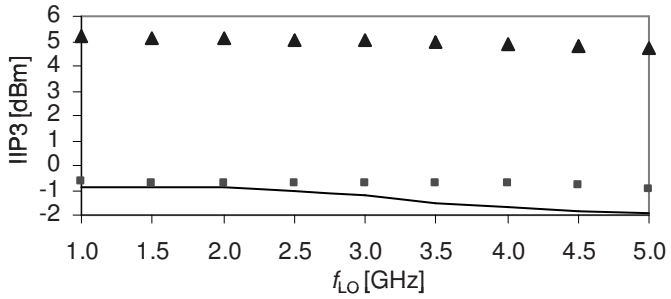


Fig. 5.21. The Gilbert mixer's IIP3 as a function of the LO frequency. Simulated (line), model including all conductance nonlinearity (squares) and model including only g_m nonlinearity (triangular) for P_{dc}=3mW, V_C=0.3V, V_{LO}=0.6V and α =0.1.

IF=10kHz and IF=10MHz as a function of the LO frequency (for 11dB conversion gain, P_{dc}=3mW, V_{LO}=1V and α =0.1). The figure illustrates that the estimation error of our noise model is lower than 1dB for f_{LO} below 5 GHz in a 90nm CMOS technology.

Fig. 5.20 shows the simulated and calculated IIP3 as a function of the overdrive voltage of M_3 with 3 mW power consumption. For the LO signal, f_{LO}=2GHz, V_C=0.3V, V_{LO}=0.6V and α =0.1. The two tone signals are at 2.01GHz and 2.014GHz, and the IIP3 is extrapolated by sweeping the input power from -25dBm to -15dBm. The estimation error is within 0.5dB for our

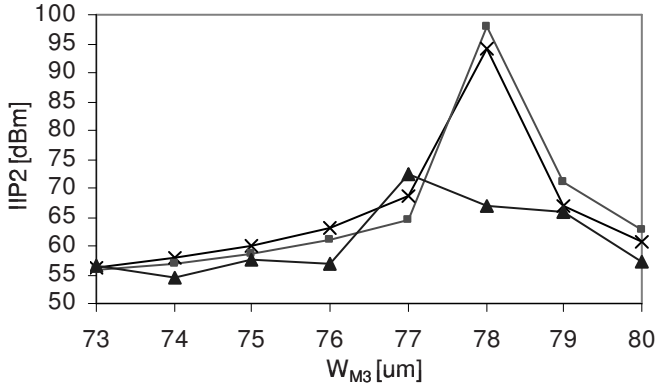


Fig.5.22. The Gilbert mixer's IIP2 as a function of the width of M_3. Simulated (line with cross), model including all conductance nonlinearity and LO effect (line with squares) and model including only gm nonlinearity and no LO slope effect (line with triangular) for, f_{LO} =1GHz, V_C =0.3V, V_{LO} =0.6V and α =0.1.

model, while the error is larger than 6 dB for the model only including gm nonlinearity.

Fig. 5.21shows a comparison of simulated and calculated IIP3 as a function of LO frequency from 1 GHz to 5 GHz. For the LO signal, V_C=0.3 V, V_{LO}=0.6 V and α =0.1. The two tone signals are at f_{LO} +10MHz and f_{LO}+14MHz, and the IIP3 is extrapolated by sweeping the input power from -25dBm to -15dBm. The estimation error of our IIP3 model is lower than 1dB, while the error is larger than 5 dB for the model only including gm nonlinearity. The estimation error in our model increases with frequency, which is because the effect of capacitances on the exact waveform of the drain of M_3 is neglected.

For IIP2 the double-balanced mixer shown in Fig. 5.9a is used. Three DC offset voltage sources model the switch pair mismatch and input stage mismatch, where V_{off1} is 5 mV, V_{off2} is 3 mV and V_{off3} is 3 mV. Fig. 5.22 shows the simulated and calculated IIP2 as a function of the width of M_3 for constant V_{GT}. For the LO signal, f_{LO}=1 GHz, V_C=0.3 V, V_{LO}=0.6 V and α =0.1.

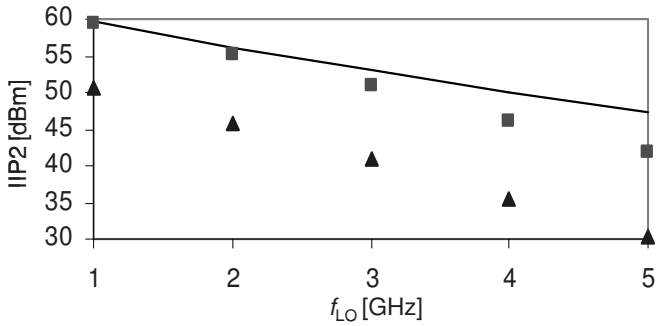


Fig.5.23. The Gilbert mixer's IIP2 as a function of LO frequency. Simulated (line), model including all conductance nonlinearity and LO effect (line with squares) and model including only gm nonlinearity and no LO slope effect (line with triangular) for, V_C =0.3 V, V_{LO} =0.6 V and α =0.1.

The IIP2 is estimated by 12 time-invariant nonlinearity calculations. The two tone signals are at 1.01 GHz and 1.014 GHz, and the IIP2 is extrapolated by sweeping the input power from -30dBm to -25dBm. For IIP2 lower than 65 dBm, the estimation error of our model is below 1dB, while for IIP2 higher than 70 dBm the error is within 4 dB. For the model only including gm nonlinearity and without considering the LO slope effect, it does not predict the IIP2 peak. Fig. 5.23 shows a comparison of simulated and calculated IIP2 as a function of LO frequency from 1 GHz to 5 GHz. For the LO signal, V_C=0.3 V, V_{LO}=0.6 V and α =0.1. The two tone signals are at f_{LO} +10MHz and f_{LO} +14MHz, and the IIP2 is extrapolated by sweeping the input power from -30dBm to -25dBm. The estimation error of our model increases with increasing LO frequency, but remains smaller than 4 dB, while the error is larger than 10dB for the model only including gm nonlinearity and without considering the LO slope effect .

The estimation time of our model is compared with the simulation time using Spectre shown in Table 5.1. For noise and IIP3 estimation, our model speeds up the estimation time by a factor of about 40, since only one or two

Table 5.1 Model estimation time

	Estimation time (second)	Simulation time (second)
Noise	0.06	2.2
IIP3	0.07	3.7
IIP2	1.68	7.4

time-invariant circuit calculations are involved. Although IIP2 estimation takes 12 time-invariant nonlinearity calculations, the estimation time is still 3 times less in comparison with the circuit simulation.

5.6 Conclusion

A simple closed-form model for the fast and accurate estimation of noise, IIP3 and IIP2 of the active mixer is presented. The mixer noise can be estimated by two ac noise calculations with error smaller than 2dB while the calculation time is about 40 times shorter than using a commercial simulator. The model shows that the decreasing transistor output resistance in deep-submicrometer technologies rather than the output capacitance is a dominant reason for the flicker noise leakage. Any flicker noise cancellation technique should include the effect of output resistance. By properly increasing W and L of M_3 and using longer switch pair transistors (M_1/M_2), the noise performance can be improved while no degradation on gain is introduced. The mixer IIP3 can be estimated by one time-invariant nonlinearity calculation with error smaller than 1dBm, while calculation time is reduced with a factor of 50 times. The slope of the LO has little effect on the IIP3. However, the LO slope together with the cross-modulation nonlinearity and output conductance nonlinearity in the triode region contribute significantly to the single-ended IM2 of the mixer. Therefore, the accuracy of the IIP2 estimation is highly dependent on a good capture of the LO waveform. Neglecting LO slope or only considering tranconductance nonlinearity will overestimate the

single-side IM2 and significantly underestimate the differential IIP2. Other than introducing mismatches to the mixer, tuning the LO slope can be a new approach of IIP2 calibration.

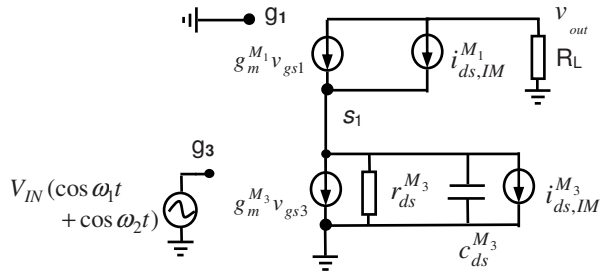


Fig. 5.24. Model for IM2/IM3 calculation of the cascode amplifier.

5.7 **Appendix**

A general distortion model for amplifiers presented in [29] is utilized to derive the IM3 and IM2 of the cascode amplifier shown in Fig. 5.24, where $V_{IN}(\cos\omega_1 t + \cos\omega_2 t)$ is the two-tone input signal.

For IM3 to the first order, we include all third-order resistive nonlinearities between drain-source terminals and the output IM3 is given by

$$v_{IM_3} = H^{M_3}(2\omega_1 - \omega_2) \cdot i_{ds,IM_3}^{M_3} + H^{M_1}(2\omega_1 - \omega_2) \cdot i_{ds,IM_3}^{M_1}$$

$$\approx \frac{-3 \times {V_{IN}}^3 R_L}{4 \times \left(1 + g_{m1}^{M_1} Z_{\mathrm{ds}}^{M_3}\right)} \times \left[g_{m1}^{M_1} Z_{\mathrm{ds}}^{M_3}\left(g_{m3}^{M_3} - g_{x21}^{M_3}\right) + g_{m3}^{M_3} - g_{ds3}^{M_1}\left(g_{m1}^{M_3} R_L\right)^3 \right. \quad (5.33)$$

$$\left. - g_{x21}^{M_1} g_{m1}^{M_3} R_L + g_{x12}^{M_1}\left(g_{m1}^{M_3} R_L\right)^2\right]$$

$i_{ds,IM_3}^{M_1}$ and $i_{ds,IM_3}^{M_3}$ are the IM3 current component of transistor M_1 and M_3 respectively; $H^{M_1}(2\omega_1 - \omega_2)$ and $H^{M_3}(2\omega_1 - \omega_2)$ are the gain from IM3 current component to the voltage output.

For IM2, we include all the nonlinearities between drain-source terminals and the output IM2 is given by

$$
\begin{aligned}
v_{IM_2} &= H^{M_3}(\omega_1-\omega_2)\cdot i_{ds,IM_2}^{M_3} + H^{M_1}(\omega_1-\omega_2)\cdot i_{ds,IM_2}^{M_1} \\
&\approx {V_{IN}}^2\left(\frac{-r_{\mathrm{ds}}^{M_3} g_{m1}^{M_1} R_L}{1+g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3}}\cdot i_{ds,IM_2}^{M_3} + \frac{-R_L}{1+g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3}}\cdot i_{ds,IM_2}^{M_1}\right) \\
&\approx \frac{R_L {V_{IN}}^2}{1+g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3}}\cdot\{g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3}\cdot\left[-g_{m2}^{M_3} - g_{ds2}^{M_3}\cdot\left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)^2 + g_{x11}^{M_3}\cdot\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right] \\
&\qquad\qquad -g_{m2}^{M_1} - g_{ds2}^{M_1}\cdot\left(g_{m1}^{M_3} R_L\right)^2 + g_{x11}^{M_3}\cdot g_{m1}^{M_3} R_L\}
\end{aligned}
\tag{5.34}
$$

$i_{di,IM_3}^{M_1}$ and $i_{ds,IM_3}^{M_3}$ are the IM2 current component of transistor M_1 and M_3 respectively; $H^{M_1}(2\omega_1-\omega_2)$ and $H^{M_3}(2\omega_1-\omega_2)$ are the gain from IM2 current component to the voltage output. Assuming the second-order nonlinearity of M_1 and M_3 are at the same order of magnitude and $(H^{M_3}(2\omega_1-\omega_2)/H^{M_1}(2\omega_1-\omega_2)) \approx g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3} \gg 1$, then the IM2 contribution of M_3 is dominant. Eq. (5.34) can be simplified to

$$
v_{IM_2,+} \approx \frac{R_L {V_{IN}}^2}{1+g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3}}\cdot\left[-g_{m2}^{M_3} g_{m1}^{M_1} r_{\mathrm{ds}}^{M_3} - g_{m2}^{M_1} - g_{ds2}^{M_1}\left(g_{m1}^{M_3} R_L\right)^2 + g_{x11}^{M_1} g_{m1}^{M_3} R_L\right] \tag{5.35}
$$

5.8 **Reference**

[1] S. Chehrazi, A. Mirzaei and A. A. Abidi, "Noise in current-commutating passive FET mixers," *IEEE Trans. Circuits and Systems I,* vol. 57, pp. 332-344, Feb. 2010.

[2] S. Chehrazi, A. Mirzaei and A. A. Abidi, "Second-order intermodulation in current-commutating passive FET mixers," *IEEE Trans. Circuits and Systems I,* vol. 56, pp. 2256-2568, Dec. 2009.

[3] H. Khatri, P. S. Gudem and L. E. Larson, "Distortion in current commutating passive CMOS downconversion mixers," *IEEE Trans. Microwave Theory and Techniques,* vol. 57, pp. 2671-2681, Nov. 2009.

[4] A. V. Do, C. C. Boon, M. A. Do, K. S. Yeo and A. Cabuk, "An energy-aware CMOS receiver front end for low-power 2.4-GHz applications," *IEEE Trans. Circuits and Systems I*, vol 57, 2675-2684, Oct. 2010.

[5] N. Kim, V. Aparin and L. E. Larson, "A resistively degenerated wideband passive mixer with low noise figure and high IIP2," *IEEE Trans. Microwave Theory and Techniques,* vol. 58, pp. 820-829, Apr. 2010.

[6] W.Cheng, A. J. Annema and B. Nauta, "A multi-step P-cell for LNA design automation," *IEEE Int. Symp.Circuits Syst.(ISCAS),* pp.2550-2553*,* May 2008.

[7] H. Darabi and A. A. Abidi, "Noise in RF-CMOS mixers: a simple physical model," *IEEE J. Solid-State Circuits,* vol. 35, pp. 15-25, Jan. 2000.

[8] M. T. Terrovitis and R. G. Meyer, "Noise in current-commutating CMOS mixers," *IEEE J. Solid-State Circuits,* vol. 34, pp. 772-783, June 1999.

[9] T. Melly, A. –S. Porret, C. C. Enz and E.A.Vittoz, "An analysis of flicker noise rejection in low-power and low-voltage CMOS mixers," *IEEE J. Solid-State Circuits,* vol. 36, pp. 102-109, Jan. 2001.

[10] J. Lerdworatawee and W. Namgoong, "Generalized linear periodic time-varying analysis for noise reduction in an active mixer," *IEEE J. Solid-State Circuits,* vol. 42, pp. 15-25, June 2007.

[11] M. T. Terrovitis and R. G. Meyer, "Intermodulation distortion in current-commutating CMOS mixers," *IEEE J. Solid-State Circuits,* vol. 34, pp. 1461-1473, June 2000.

[12] D. Manstretta, M. Brandolini and F. Svelto, "Second-Order intermodulation mechanisms in CMOS downconverters," *IEEE J. Solid-State Circuits,* vol. 38, pp. 394-406, March 2003.

[13] G. Theodoratos, A. Vasilopoulos, G. Vitzilaios and Y. Papananos, "Calculating distortion in active CMOS mixers using Volterra series," *IEEE Int Symp.Circuits Syst.,* pp.2249-2252, May 2006.

[14] P. Dobrovolny, G. Vandersteen, P. Wambacq and S. Donnay, "Analysis and white-box modeling of weakly nonlinear time-varying circuits," in *Proc. Design, Automation and Test in Europe Conference and Exhibition (DATE)*, pp.624-629, 2003.

[15] J. Yoon, H. Kim, C. Park, J. Yang, H. Song, S. Lee and B. Kim, "A new RF CMOS gilbert mixer with improved noise figure and linearity," *IEEE Trans. Microwave Theory and Techniques,* vol. 56, pp. 626-631, Mar. 2008.

[16] J. Park, C.-H Lee, B. Kim and J. Laskar, "Design and analysis of low flicker-noise CMOS mixers for direct-conversion receivers", *IEEE Trans. Microwave Theory and Techniques,* vol. 54, pp. 4372-4380, Dec. 2006.

[17] C. D. Hull and R. Meyer, "A Systematic approach to the analysis of noise in mixers", *IEEE Trans. Circuits and Systems I,* vol. 40, pp. 909-919, Dec. 1993.

[18] M. T. Terrovitis, K. S. Kundert and R.G.Meyer., "Cyclostationary noise in radio-frequency communication systems", *IEEE Trans. Circuits and Systems I,* vol. 49, pp. 1666-1671, Jan. 2002.
[19] H. Darabi and J. Chiu, "A noise cancellation technique in active RF-CMOS mixers," *IEEE J. Solid-State Circuits,* vol. 40, pp. 2628-2632, Dec. 2005.
[20] B. Razavi, "Design considerations for future RF circuits", *IEEE Int Symp.Circuits Syst.,* pp.741-744, May 2007.
[21] A. Oppenheim, A.S.Willsky and S.H. Nawab, "Signals and Systems," 2nd Edition, Prentice-Hall, 2002.
[22] M. Brandolini, P. Rossi, D. Sanzogni and F.Svelto, "A +78 dBm IIP2 CMOS direct downconversion mixer for fully integrated UMTS receivers," *IEEE J. Solid-State Circuits,* vol. 41, pp. 552-559, March 2006.
[23] K. Kivekas, A. Parssinen, J. Ryynanen, J. Jussila and K. Halonen, "Calibration techniques of active BiCMOS mixers," *IEEE J. Solid-State Circuits,* vol. 37, pp. 766-769, June. 2002.
[24] M. W. Hwang, G .H. Cho, S. Y. Yoo, J. C. Lee, S. M. Ock, S. K. Min, S. H. Beck, K. Lim, S. Han and J. Lee, "A high IIP2 direct-conversion receiver using even-harmonic reduction technique for cellular CDMA/PCS/GPS applications," *IEEE Trans. Circuits and Systems I,* vol. 55, pp. 2934-2943, Oct. 2008.
[25] K. Dufrene, Z. Boos and R.Weigel, "Digital adaptive IIP2 Calibration Scheme for CMOS downconversion mixers," *IEEE J. Solid-State Circuits,* vol. 43, pp. 2434-2445, Nov. 2008.
[26] W. Kim, S. G. Yang, Y. K. Moon, J. Yu, H. Shin, W. Choo and B. H. Park, "IP2 calibrator using common mode feedback circuitry," *IEEE European Solid-State Circuits Conference (ESSCIRC),* pp.231-234, 2005.
[27] K. Dufrene and R. Weigel, "A novel IP2 calibration method for low-voltage downconversion mixers," *IEEE Radio Frequency Integrated Circuits (*RFIC*)* Symposium, pp. 289– 292, 2006.
[28] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*, Norwell, MA: Kluwer, 1998.
[29] W. Cheng, A. J. Annema, J. A. Croon, D. B. M. Klaasen and B. Nauta, "A general weak nonlinearity model for LNAs," *IEEE Custom Integrated Circuits Conference*, pp.221 – 224, Sept. 2008.
[30] Available http://www.nxp.com/models/mos_models/psp/
[31] W. Cheng, A. J. Annema, J. A. Croon and B. Nauta, "Noise and nonlinearity modeling of active mixers for fast and accurate estimation," *IEEE Trans. Circuits and Systems I,* vol. 58, pp. 276-289, Feb. 2011.

Chapter 6

A Flicker-Noise/IM3 Cancellation Technique for Active Mixers

Based on the noise and distortion model of active mixers presented in chapter 5, this chapter presents an approach to simultaneously cancel flicker noise and IM3 in Gilbert-type mixers, utilizing negative impedances. For proof of concept, two prototype double-balanced mixers in 0.16μm CMOS are fabricated. The first demonstration mixer chip was optimized for full IM3 cancellation and partial flicker noise cancellation; this chip achieves 9dB flicker noise suppression, improvements of 10dB for IIP3, 5dB for conversion gain, and 1dB for input P_{1dB} while the thermal noise increased by 0.1dB. The negative impedance increases the power consumption for the mixer by 16%, and increases the die area by 8% (46x28μm^2). A second demonstration mixer chip aims at full flicker noise cancellation and partial IM3 cancellation, while operating on a low supply voltage ($0.67\times V_{DD}$); in this chip, the negative impedance increases the power consumption by 7.3%, and increases the die

This chapter has been submitted to IEEE Journal of Solid-State Circuits in 2012.

area by 7% (50x20μm^2). For one chip sample, measurements show >10dB flicker noise suppression within $\pm$200% variation of the negative impedance bias current; for ten randomly selected chip samples, >11dB flicker noise suppression is measured.

6.1 Introduction

Due to its higher conversion gain, the active mixer provides better noise suppression of the subsequent stages than passive mixers. However, CMOS active mixers suffer from high flicker noise as well as from low linearity. While high flicker noise causes serious sensitivity degradation especially in narrowband direct-conversion receivers, mixers with poor linearity limit the dynamic range of the receiver.

Three major techniques have been presented for flicker noise reduction in CMOS active mixers:

1. Dynamic current injection [1-2]: a pMOS cross-coupled pair is used to inject current into the NMOS mixer transconductor stage only at the switching on/off instants (at $0.5T_{LO}$) so that no DC current flowing through the switches. This suppresses the flicker noise leakage from the switching pair.
2. Double LO switch pairs [3]: extra switches in series driven at 2LO frequency are used so that during the switching period little DC current flows through the major switches that are driven by LO signal, thereby reducing flicker noise leakage.
3. RF leakageless static current bleeding with two resonating inductors [4]. Two inductors are connected between the mixer transconductor stage and the current bleeding circuit. The inductors resonate out the tail capacitance and reduce the RF signal leakage to the current bleeding circuit.

In Technique 1, a large LO swing and large headroom is required, increasing the LO power and decreasing the conversion gain due to the use of small R_{load} [1]. Technique 2 needs a stack of three transistors plus the load, which is not suitable for deep-submicrometer technologies with low supply voltages. Also extra clock power is needed to generate 2*LO signal. Technique 3 needs two inductors which consume die area. Common to all the techniques in [1-4] is

that the effect of the transistor output resistance on the flicker noise leakage is neglected.

In technologies with long-channel transistors where the output capacitance of transistors is dominant, the effect of output resistance r_{ds3} on flicker noise leakage can be neglected [1-4]. However, nowadays the technology scaling offers higher f_T well above 100 GHz, while it also brings lower transistor output resistance and lower supply voltage [5]. Neglecting the effect of r_{ds3} in deep-submicrometer technologies can yield a significant underestimation of the output flicker noise [7]. Taking into account the effect of both output resistance and output capacitance on flicker noise leakage, in this chapter we propose a combined flicker noise/IM3 cancellation technique that uses a negative impedance to minimize the flicker noise leakage from the switching pair and to simultaneously improve the linearity. Section 6.2 presents the circuit theory behind this flicker noise/IM3 cancellation technique. Section 6.3 and 6.4 show a circuit implementation and the measurement and simulation results; the results are summarized in section 6.5.

6.2 Flicker noise/IM3 cancellation using negative impedance

6.2.1 Flicker noise leakage in Gilbert mixers

The double-balanced Gilbert mixer shown in Fig. 6.1a is widely used as the active downconverter in CMOS receivers. Transistor M_{3a}/M_{3b} convert v_{in} into current that is commuting via switch pair M_{1a}/M_{2a} and M_{1b}/M_{2b} respectively. The flicker noise output of the Gilbert mixer, $v_{fl,out}(t)$, is dominated by the switch pair M_1/M_2, while transistor M_3 is causing thermal noise folding [6-7]. Assuming perfect symmetry in the mixer, the flicker noise leakage mechanism from each switch is the same: it is hence sufficient to focus on flicker noise leakage from one of the switch pair transistors. In [7] the time-varying small signal model shown in Fig. 6.1b is used to analyze the flicker noise contributed by M_{1a} in one LO period at the mixer output ($v_{fl,out}^{M_{1a}}(t)$). The flicker noise of M_{1a} is modeled by the equivalent gate-referred

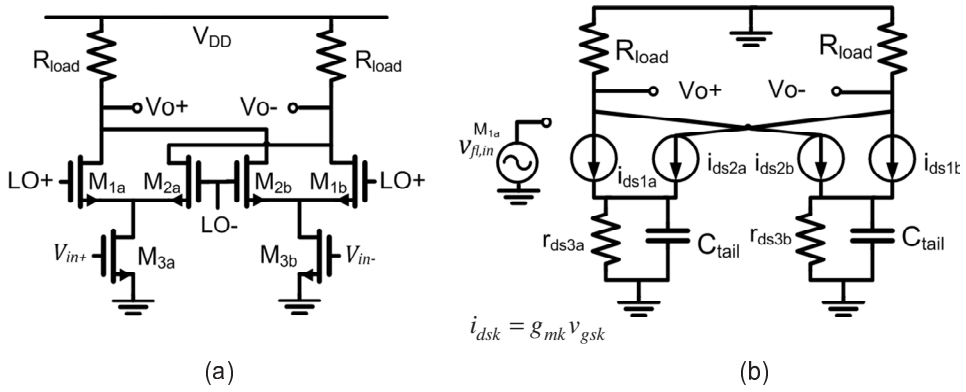


Fig. 6.1. (a) Schematic and (b) time-varying noise model of the double-balanced Gilbert mixer.

.

root mean square (rms) noise voltage $v_{fl,in}^{M_{1a}}$. For a first order approximation of $v_{fl,out}^{M_{1a}}$, a few assumptions are made:

- The LO signal is modeled by a trapezoid shown in Fig. 6.2(a) with a rise/fall time equal to $\alpha \cdot T_{LO}$.
- At (t_1, t_2), M_{2a} and M_{2b} are off while in (t_3 , t_4) M_{1a} and M_{1b} are off.

Taking into account the transconductance of M_1, M_2 and M_3, and the output admittance of M_3 ($Y_{ds}^{M_{3a}} = g_{ds}^{M_{3a}} + j\omega_{LO} C_{ds}^{M_{3a}}$), the flicker noise contribution of M_{1a} at t_1, $\frac{T_{LO}}{2}$ and t_3 in Fig. 6.2(a) are given by:

$$v_{fl,out}^{M_{1a}}\Big|_{t_1} = \frac{-g_m^{M_{1a}} R_L Y_{ds}^{M_{3a}} v_{fl,in}^{M_{1a}}}{g_m^{M_{1a}} + Y_{ds}^{M_{3a}}}\Bigg|_{t_1} \tag{6.1}$$

$$v_{fl,out}^{M_{1a}}\Big|_{\frac{T_{LO}}{2}} = -g_m^{M_{1a}} R_L v_{fl,in}^{M_{1a}}\Big|_{\frac{T_{LO}}{2}} \tag{6.2}$$

$$v_{fl,out}^{M_{1a}}\Big|_{t_3} = 0 \tag{6.3}$$

In (t_1, t_2), M_{1a} and M_{3a} form a cascode amplifier. Due to the finite output impedance of M_{3a} in deep-submicrometer CMOS, the noise contribution from the cascode transistor M_1, given by (6.1), cannot be neglected. In (t_2, t_3) both M_{1a} and M_{2a} are on, while at $\frac{T_{LO}}{2}$ M_{1a} and M_{2a} act as a balanced differential pair. In this period of time, the output impedance of M_3 has a negligibly small

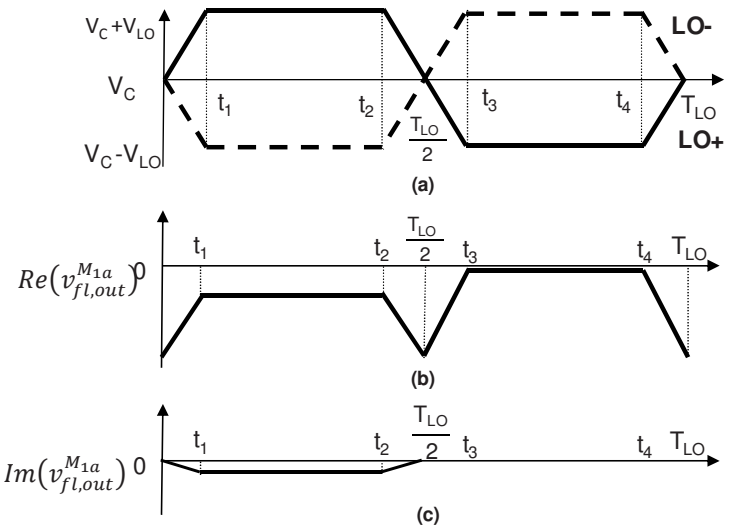


Fig. 6.2. (a) Waveform of the LO signal. (b) and (c): approximation of the real and imaginary part of $v_{fl,out}^{M_{1a}}(t)$ respectively.

effect on $v_{fl,out}^{M_{1a}}$ as shown by (6.2). In (t_3, t_4) M_{1a} is assumed to be off, thus $v_{fl,out}^{M_{1a}}$ is zero. Note that the integral (or area) of $v_{fl,out}^{M_{1a}}(t)$ shown in Fig. 6.2(b) and (c) corresponds to the flicker noise leakage that involves no frequency translation [7]. As a result, the flicker noise at the mixer output contributed by M_{1a}/M_{2a} and M_{1b}/M_{2b} is

$$\begin{aligned} S_{fl,out} &= 4 \times \left[\int_0^{T_{LO}} v_{fl,out}^{M_{1a}}(t)\right]^2 \\ &= \frac{4}{{T_{LO}}^2} \times \left[(t_2 - t_1) \cdot v_{fl,out}^{M_{1a}}\Big|_{t_1} \right. \\ &\quad \left. + t_1 \cdot \left(v_{fl,out}^{M_{1a}}\Big|_{\frac{T_{LO}}{2}} - v_{fl,out}^{M_{1a}}\Big|_{t_1}\right) + \left(t_3 - \frac{T_{LO}}{2}\right) \cdot v_{fl,out}^{M_{1a}}\Big|_{\frac{T_{LO}}{2}}\right]^2 \end{aligned} \tag{6.4}$$

For the symmetrical LO signal shown in Fig. 6.2(a), with a rise/fall time equal to $\alpha \cdot T_{LO}$, the time instants t_1, t_2 and t_3 can be rewritten as $t_1 = 0.5\alpha T_{LO}$, $t_2 = 0.5 \times (1 - \alpha)T_{LO}$, $t_3 = 0.5 \times (1 + \alpha)T_{LO}$. This enables rewriting (6.4) into

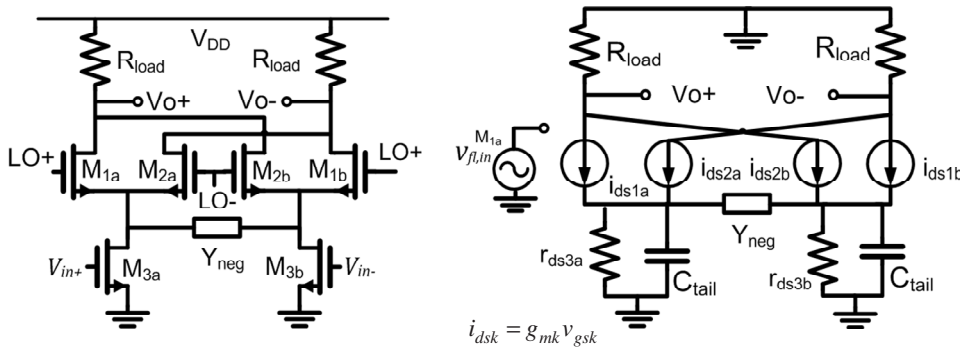


Fig. 6.3. (a) Schematic and (b) time-varying noise model for the mixer with a negative impedance for flicker noise cancellation.

$$S_{fl,out} = \left[(1-3\alpha)\cdot v_{fl,out}^{M_{1a}}\Big|_{t_1} + 2\alpha v_{fl,out}^{M_{1a}}\Big|_{\frac{T_{LO}}{2}}\right]^2 \tag{6.5}$$

6.2.2 Negative impedance for flicker noise cancellation

To minimize the integral of $v_{fl,out}^{M_{1a}}(t)$ in (6.4) --- hence to minimize 1/f noise leakage --- we apply a negative impedance $Y_{neg} = G_{neg} + j\omega C_{neg}$ between the drain of M_{3a} and M_{3b} as shown in Fig. 6.3. Using the model shown in Fig. 6.3(b), this yields a different $v_{fl,out}^{M_{1a}}$ for the time interval (t_1-t_2):

$$v_{fl,out}^{M_{1a}}\Big|_{t_1} = \left.\frac{-g_m^{M_{1a}}R_L\left(Y_{ds}^{M_{3a}}+2Y_{neg}\right)v_{fl,in}^{M_{1a}}}{g_m^{M_{1a}}+Y_{ds}^{M_{3a}}+2Y_{neg}}\right|_{t_1} \tag{6.6}$$

$$v_{fl,out}^{M_{1a}}\Big|_{\frac{T_{LO}}{2}} = -g_m^{M_{1a}}R_L v_{fl,in}^{M_{1a}}\Big|_{\frac{T_{LO}}{2}} \tag{6.7}$$

$$v_{fl,out}^{M_{1a}}\Big|_{t_3} = 0 \tag{6.8}$$

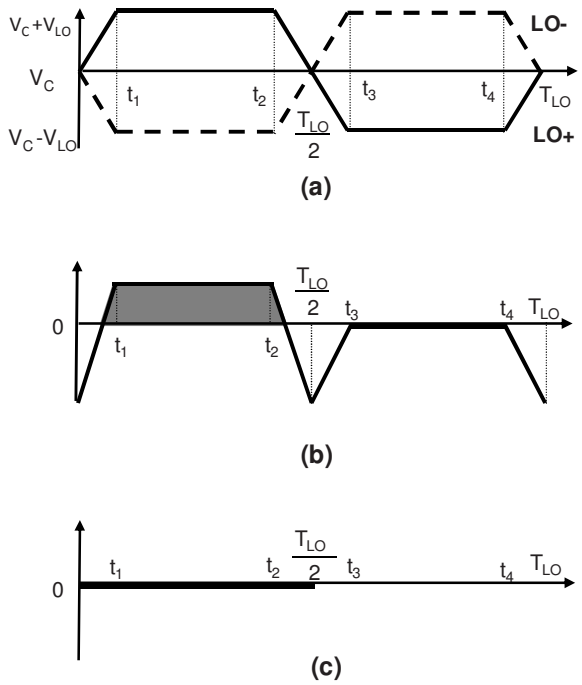


Fig. 6.4. (a) Waveform of the LO signal. (b) and (c): approximation of the real respectively imaginary part of $v_{fl,out}^{M_{1a}}(t)$ using a negative impedance for flicker noise cancellation.

Eq. (6.6) shows that for $C_{neg} \approx -0.5C_{tail}$ and $G_{neg} \in \left[\frac{-g_{ds}^{M_{3a}} - g_m^{M_{1a}}}{2}, \frac{-g_{ds}^{M_{3a}}}{2}\right]$ the sign of the real part of $v_{fl,out}^{M_{1a}}\big|_{t_1}$ changes from negative to positive (a detailed derivation is presented in the appendix). At $\frac{T_{LO}}{2}$, the negative impedance has no effect on $v_{fl,out}^{M_{1a}}$ as shown in (6.7) since M_{1a} and M_{2a} act as a balanced differential pair. In (t_3, t_4) M_{1a} is off, thus $v_{fl,out}^{M_{1a}}$ is zero. Now the new approximated waveform of $v_{fl,out}^{M_{1a}}(t)$ is shown in Fig. 6.4b. The negative impedance Y_{neg} changes the real part of $v_{fl,out}^{M_{1a}}$ in (t_1, t_2) from negative to positive value, which enables minimization of the area of $v_{fl,out}^{M_{1a}}\big|_{t_1}$. This leads to the cancellation of the flicker noise leakage from the switching pair (M_{1a}/M_{2a} and M_{1b}/M_{2b}). For

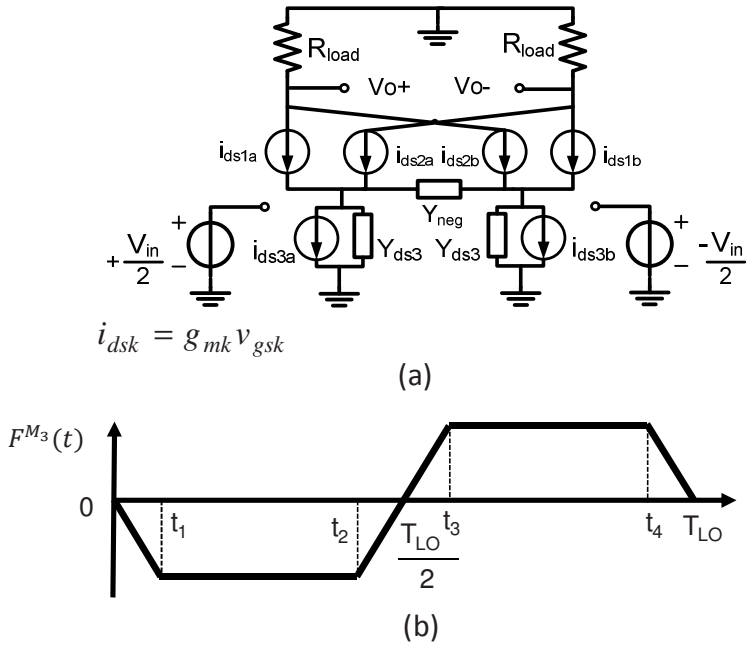


Fig. 6.5. (a) Time-varying linear model for calculating the voltage gain of the mixer with a negative impedance for flicker noise cancellation and (b) the approximation for the instantaneous voltage gain $F^{M_3}(t)$.

.

a complete flicker noise leakage cancellation, (5) equates to zero. Together with (6.6-6.7), this equation gives the condition for a complete flicker noise leakage cancellation:

$$\begin{cases} G_{neg} = -\frac{g_{ds}^{M_{3a}}}{2} - \frac{\alpha g_m^{M_{1a}}}{(1-\alpha)} \\ C_{neg} = -\frac{C_{tail}}{2} \end{cases} \tag{6.9}$$

6.2.3 Negative impedance impact on gain and thermal noise

It was derived in [7] that the first-order Fourier coefficient of the instantaneous voltage gain $F^{M_3}(t) = v_{out}/v_{in}$ in one LO period corresponds to the conversion gain of a mixer. Using the model shown in Fig. 6.5(a), a sufficiently accurate approximation of $F^{M_3}(t)$ is given in Fig. 6.5(b) with

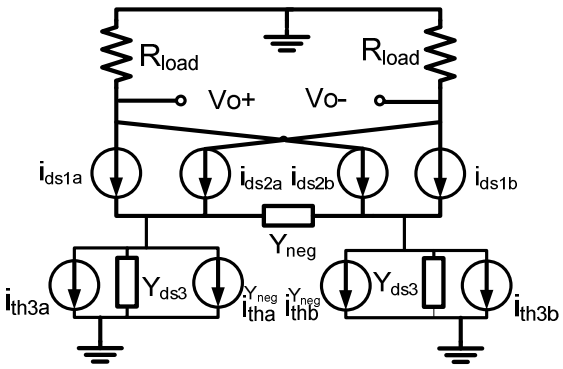


Fig. 6.6. Time-varying noise model for calculating the thermal noise of the mixer with a negative impedance.

$$F^{M_3}|_{t_1} = \left.\frac{-g_m^{M_{1a}} g_m^{M_{3a}} R_L}{g_m^{M_{1a}} + Y_{ds}^{M_{3a}} + 2Y_{neg}}\right|_{t_1} \tag{6.10}$$

$$F^{M_3}|_{\frac{T_{LO}}{2}} = 0 \tag{6.11}$$

$$F^{M_3}|_{t_3} = -F^{M_3}|_{t_1} \tag{6.12}$$

At (t_1, t_2), M_{2a}/M_{2b} are off and M_{1a}/M_{3a} and M_{1b}/M_{3b} forms a differential cascode common-source amplifier. At (t_3, t_4), M_{1a}/M_{1b} are off and M_{2a}/M_{3a} and M_{2b}/M_{3b} form a differential cascode common-source amplifier. At $0.5T_{LO}$, M_{1a}/M_{2a} and M_{1b}/M_{2b} are on, and there is no output due to the differential symmetry. This yields the voltage conversion gain

$$V_{gain} = \frac{2\sin(\alpha\pi)}{\pi^2\alpha} F^{M_3}|_{t_1} = \frac{2\sin(\alpha\pi)}{\pi^2\alpha} \left.\frac{-g_m^{M_{1a}} g_m^{M_{3a}} R_L}{g_m^{M_{1a}} + Y_{ds}^{M_{3a}} + 2Y_{neg}}\right|_{t_1} \tag{6.13}$$

Equation (6.13) shows that the conversion gain is increased under *partial* flicker noise cancelling condition ($G_{neg} \in \left[\frac{-g_{ds}^{M_{3a}} - g_m^{M_{1a}}}{2}, \frac{-g_{ds}^{M_{3a}}}{2}\right]$ and $C_{neg} \approx -0.5C_{tail}$).

Fig. 6.6 shows the noise model for the mixer, where two non-correlated noise current $i_{tha}^{Y_{neg}}$ and $i_{thb}^{Y_{neg}}$ model the noise of Y_{neg}. Note that the noise of the negative impedance contributes to the mixer output by the same transfer

function as the noise of M_{3a}/M_{3b}. Therefore, Y_{neg} only contributes to thermal noise, while no flicker noise leaks to the output for the symmetric mixer. As a result, the mixer thermal noise is dominantly contributed by the thermal-noise folding of M_{3a}/M_{3b}, the load R_{Load}, the input source impedance R_s and the negative impedance Y_{neg}. Assuming perfect input matching, the single-side band noise figure (NF) for high IF (thermal noise dominated) is then given by

$$\begin{aligned} NF_{SSB} &\approx \frac{2S_{M3a,out} + S_{Rs,out} + S_{Yneg} + S_{Rload,out}}{0.5S_{Rs,out}} \\ &= 2\frac{\left(\frac{i_{th3a}}{g_{m3}}\right)^2 + kTR_s + \left(\frac{i_{th}^{Yneg}}{g_{m3}}\right)^2 + \frac{4kTR_{load}}{\left(1-\frac{4\alpha}{3}\right)\left|F^{M_3}|_{t_1}\right|^2}}{kTR_s} \end{aligned} \tag{6.14}$$

where the four terms respectively account for the thermal noise from the transconductor stage M_{3a}/M_{3b}, the thermal noise from the input source impedance R_s, the thermal noise from the negative impedance and the thermal noise from the load. Although the extra noise by Y_{neg} increases the thermal noise NF, due to the increased conversion gain (larger $F^{M_3}|_{t_1}$), the input referred noise due to the load R_{Load} is decreased. As a result, the thermal noise increase due to Y_{neg} can be small.

6.2.4 Negative impedance for IM3 distortion cancellation

It is shown in [7] that the IM3 of the time-varying mixer can be estimated by one time-invariant IM3 calculation at the maximum of the LO signal. Therefore, the circuit model shown in Fig. 6.7 is used to estimate the IM3. When the LO signal reaches its positive maximum at t1, M1a/M1b are fully on and M2a/M2b are fully off. The IM3 distortion current $i_{IM3}^{M_{1a}}$, $i_{IM3}^{M_{1b}}$, $i_{IM3}^{M_{3a}}$ and $i_{IM3}^{M_{3b}}$ are generated by the voltage swing across the transistor terminals. Given the differential circuit topology we can assume $i_{IM3}^{M_{1a}} = -i_{IM3}^{M_{1b}} = i_{IM3}^{M_1}$ and $i_{IM3}^{M_{3a}} = -i_{IM3}^{M_{3b}} = i_{IM3}^{M_3}$. For a first-order approximation, only the transconductance of M1, M2 and M3, and the output admittance of M3 are taken into account, which yields

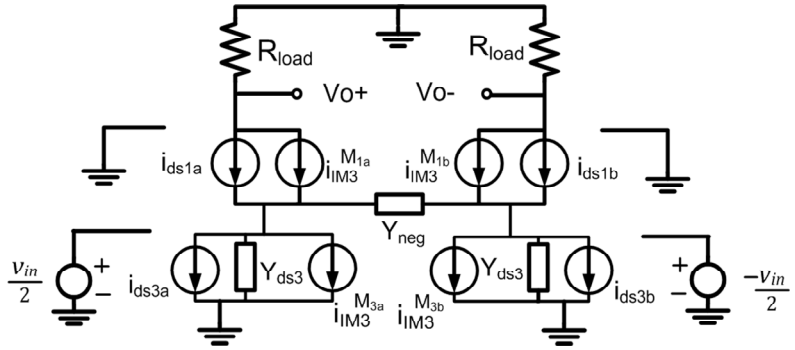


Fig. 6.7. Circuit model for the mixer distortion analysis.

$$v_{out}^{IM3} = \frac{-2g_m^{M_1}R_L}{g_m^{M_1}+Y_{ds}^{M_3}+2Y_{neg}} i_{IM3}^{M_3} + \frac{-2\left(Y_{ds}^{M_3}+2Y_{neg}\right)R_L}{g_m^{M_1}+Y_{ds}^{M_3}+2Y_{neg}} i_{IM3}^{M_1} \quad (6.15)$$

$$i_{IM3}^{M_3} = f_{IM3}^{M_3}\left(v_{gs}^{M_3a}, v_{ds}^{M_3a}\right) = f_{IM3}^{M_3}\left[\frac{v_{in}}{2}, \frac{-g_m^{M_3}v_{in}}{2\left(g_m^{M_1}+Y_{ds}^{M_3}+2Y_{neg}\right)}\right] \quad (6.16)$$

$$i_{IM3}^{M_1} = f_{IM3}^{M_1}\left(v_{gs}^{M_1a}, v_{ds}^{M_1a}\right) = f_{IM3}^{M_1}\left[\frac{g_m^{M_3}v_{in}}{2\left(g_m^{M_1}+Y_{ds}^{M_3}+2Y_{neg}\right)}, \frac{-g_m^{M_3}(-1+g_m^{M_1}R_L)v_{in}}{2\left(g_m^{M_1}+Y_{ds}^{M_3}+2Y_{neg}\right)}\right] \quad (6.17)$$

Equation (6.16) and (6.17) describe the fact that via transistor nonlinearity (denoted by the function f_{IM3}) the distortion current of a transistor is due to both the voltage swing across its gate-source and to its drain-source voltage swing (assuming the distortion related to the bulk-source voltage swing is insignificant).

For $G_{neg} \in \left[\frac{-g_{ds}^{M_3a}-g_m^{M_1a}}{2}, \frac{-g_{ds}^{M_3a}}{2}\right]$ and $C_{neg} \approx -0.5C_{tail}$,

- Eq. (6.16-6.17) show that the gate-source and drain-source voltage swing respectively for M_{1a} and M_{3a} have the same polarity. Thus, $i_{IM3}^{M_3}$ and $i_{IM3}^{M_1}$ have the same polarity, given that both M_{1a} and M_{3a} are biased in the saturation region.
- Eq. (6.15) shows that the gain factor for distortion currents $i_{IM3}^{M_3}$ and $i_{IM3}^{M_1}$ have opposite signs. This enables cancellation of the distortion contributions caused by $i_{IM3}^{M_3}$ and $i_{IM3}^{M_1}$.

Equating (6.15) to zero gives that for a *complete* IM3 cancellation

$$\begin{cases} G_{neg} = -\frac{g_{ds}^{M_3}}{2} - \frac{g_m^{M_1} i_d^{M_3}}{2 i_d^{M_1}} \\ C_{neg} = -\frac{C_{tail}}{2} \end{cases} \tag{6.18}$$

Under *partial* IM3 cancelling condition ($G_{neg} \in \left[\frac{-g_{ds}^{M_{3a}} - g_m^{M_{1a}}}{2}, \frac{-g_{ds}^{M_{3a}}}{2}\right]$ and $C_{neg} \approx -0.5C_{tail}$), the distortion current polarity of each transistor within the mixer remains unchanged. For the switching stage (M_{1a}/M_{1b} and M_{2a}/M_{2b}), the negative impedance changes the amplifying factor of their distortion current from negative to positive. This enables the IM3 cancellation between the transconductor stage (M_{3a}/M_{3b}) and the switching stage (M_{1a}/M_{1b} and M_{2a}/M_{2b}).

6.2.5 Summary

It can be concluded that a negative impedance ($G_{neg} \in \left[\frac{-g_{ds}^{M_{3a}} - g_m^{M_{1a}}}{2}, \frac{-g_{ds}^{M_{3a}}}{2}\right]$ and $C_{neg} \approx -0.5C_{tail}$) reduces the flicker noise leakage from the switching stage (M_{1a}/M_{1b} and M_{2a}/M_{2b}) by averaging out the flicker noise leakage transfer function. Note that this is very similar to flicker noise suppression in chopper amplifiers [8]. Using a negative impedance, also the conversion gain is increased while the thermal noise may be slightly degraded. For a perfect symmetric mixer no flicker noise will be introduced by the negative impedance.

Using negative impedance in the specified range also enables partial IM3 cancellation between the transconductor stage (M_{3a}/M_{3b}) and the switching stage (M_{1a}/M_{1b} and M_{2a}/M_{2b}). The exact optimum of Y_{neg} is in general a little different for complete flicker noise cancellation and complete IM3 cancellation. Hence Y_{neg} can be designed for either full flicker noise/partial IM3 cancellation or full IM3/partial flicker noise cancellation.

6.3 Circuit Implementation

To prove this flicker noise/IM3 cancellation concept, the circuit shown in Fig. 6.8 is implemented in a standard 0.16µm CMOS process. The negative impedance is implemented by the cross-coupled pair M_{4a}/M_{4b} with source

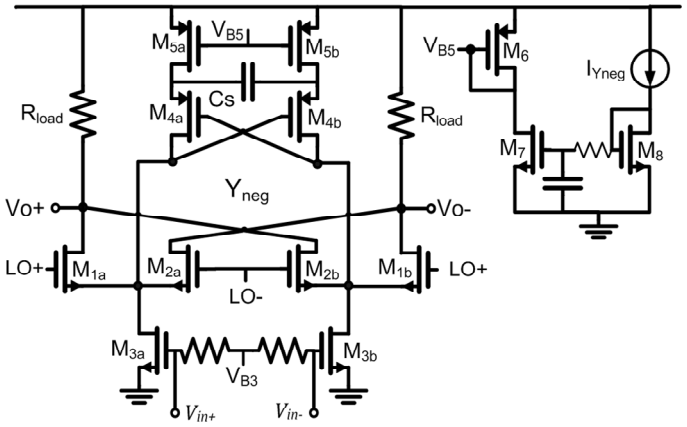


Fig. 6.8. The schematic of the mixer with a negative impedance Y_{neg}.

degeneration provided by capacitor (C_s) and current source M_{5a}/M_{5b} [9-10]. The pMOS-based negative impedance enables dc current reuse of the negative impedance by the mixer's transconductor stage. As a first-order approximation of the value of the negative impedance, only the transconductance of M_4 and M_5 and the output impedance of M_5 are taken into account, yielding

$$\begin{cases} G_{neg} = -\dfrac{g_m^{M_4}\left[g_{ds}^{M_5}\left(g_{ds}^{M_5}+g_m^{M_4}\right)+\left(C_{ds}^{M_5}+2C_s\right)^2\omega^2\right]}{2\left[\left(g_{ds}^{M_5}+g_m^{M_4}\right)^2+\left(C_{ds}^{M_5}+2C_s\right)^2\omega^2\right]} \\ C_{neg} = -\dfrac{{g_m^{M_4}}^2\left(C_{ds}^{M_5}+2C_s\right)^2\omega}{2\left[\left(g_{ds}^{M_5}+g_m^{M_4}\right)^2+\left(C_{ds}^{M_5}+2C_s\right)^2\omega^2\right]} \end{cases} \tag{6.19}$$

Assuming $g_m^{M_4} \gg g_{ds}^{M_5}$ and $\left(g_m^{M_4}/\left(C_{ds}^{M_5}+2C_s\right) \approx 2\pi f_t = \omega_t\right)$, (6.19) reduces to

$$\begin{cases} G_{neg} \approx -\dfrac{g_m^{M_4}\left(\frac{\omega}{\omega_t}\right)^2}{2\left[1+\left(\frac{\omega}{\omega_t}\right)^2\right]} = -\dfrac{g_m^{M_4}}{2} \\ C_{neg} \approx -\dfrac{C_{ds}^{M_5}+2C_s}{2\left[1+\left(\frac{\omega}{\omega_t}\right)^2\right]} = -\dfrac{C_{ds}^{M_5}+2C_s}{2} \end{cases} \tag{6.20}$$

Equation (6.20) shows that the Y_{neg} for either full flicker noise cancellation or full IM3 cancellation can be obtained by setting a suitable value for the transconductance of M_4 and for the degeneration capacitance C_s. For minimal

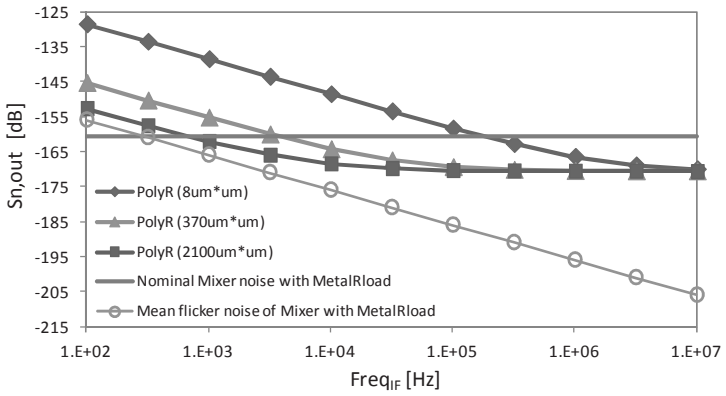


Fig. 6.9. The simulated output noise of the mixer using poly-silicon resistor and metal resistor as R_{load}.

chip area, a poly-diffusion capacitor is used for C_s instead of a fringe capacitor.

Regarding the load resistor Rload, it is frequently assumed that a poly-silicon resistor has negligible flicker noise [1, 6] which may be valid for many conventional circuits. However, in this chapter the aim is at very low flicker noise mixers. The work in [11] shows that the traps at the silicon grain boundaries in the poly-silicon cause some flicker noise, resulting in a flicker noise current given by $S_i = \frac{\beta I^2}{WL} \cdot \frac{1}{f}$, where β is a constant including technology-dependent data and temperature, I is the DC current through the resistor, and W and L are the resistor width and length respectively. Simulation results in Fig. 6.9 show that without mismatch and with metal resistors, the mixer noise output only contains thermal noise (denoted by Nominal mixer noise with MetalRload). Under mismatch and with metal resistors, the mixer flicker noise frequency corner is below 1kHz. Using poly resistors, even in the nominal case without any mismatch, the flicker noise of the poly-silicon resistor is dominant compared to the mixer thermal noise unless large silicon area is used. Therefore, in our design a serpentine metal resistor is used to be able to prove our concept properly

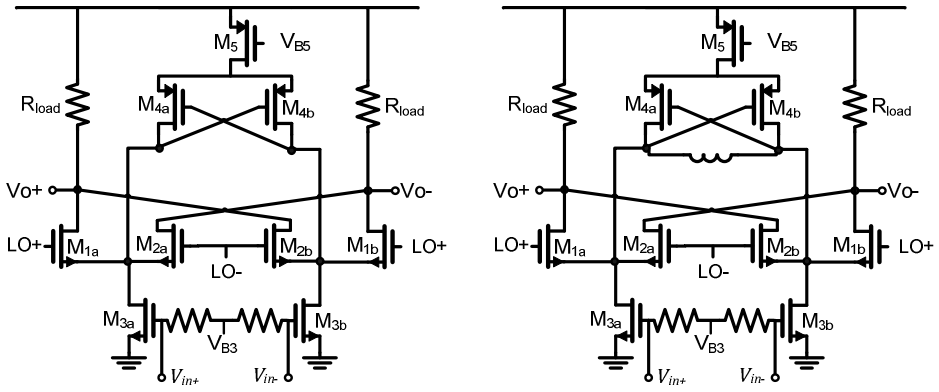


Fig. 6.10 Two mixers for flicker noise reduction. a) Dynamic bleeding, b) Dynamic bleeding with a inductor.

Using the topology shown in Fig. 6.8 we designed two chips. One (MixerD) is optimized for full IM3 cancellation and partial flicker noise cancellation using full supply voltage (V_{DD}=1.8V). To test robustness of the proposed flicker noise cancellation technique under the constraint of low supply voltage, the second chip (MixerNF) is optimized for full flicker noise cancellation and partial IM3 cancellation using $0.67 \times V_{DD}$ (1.2V). Since flicker noise is mainly a problem for narrowband system, the two mixer chips are designed for 0.9GHz. Two off-chip baluns are used to generate the differential RF input and differential clock, respectively. The external differential clock signal and an on-chip LO buffer provides the LO for mixer.

Due to the similar topology appearance, our mixer in Fig. 6.8 is compared with previous techniques of flicker noise reduction. In [1] the cross-coupled pair M_{4a}/M_{4b} shown in Fig. 6.10a is used to provide a dynamic current into the transconductor stage at the LO zero-crossings (at $\frac{T_{LO}}{2}$). As a result, at $\frac{T_{LO}}{2}$ the current through the switching pair and the transconductance of the switching pair is reduced. This enables a smaller $v_{fl,out}^{M_{1a}}|_{\frac{T_{LO}}{2}}$ in (6.5) and consequently yields some flicker noise reduction. The cross-coupled pair M_{4a}/M_{4b} turns on only around $\frac{T_{LO}}{2}$ and remains off during the remainder of the LO period, which requires a high LO voltage swing and low R_{load} (50 ohm in [1]). To tune out

the parasitic capacitance of the cross-coupled pair M_{4a}/M_{4b}, an inductor is added to the cross-coupled pair M_{4a}/M_{4b} as shown in Fig. 6.10b in [2]. Although these dynamic bleeding techniques [1-2] and our mixer all use the cross-coupled pair M_{4a}/M_{4b}, there are a number of fundamental differences:

1. The cross-coupled pair M_{4a}/M_{4b} in the dynamic bleeding technique only operates around $\frac{T_{LO}}{2}$, while in our mixer the negative impedance is operational during the total LO period. As a result, our mixer only needs normal LO voltage swing, while high LO voltage swing are required in the dynamic bleeding technique, which may impose linearity degradation due to the switching pair (see [7]).
2. The cross-coupled pair M_{4a}/M_{4b} in the dynamic bleeding technique is designed as a DC current injector rather than a negative resistor. Flicker noise leakage due to the finite transconductor output resistance is not addressed. The source degenerated capacitance together with the cross-coupled pair M_{4a}/M_{4b} in this chapter are designed as a negative impedance, which fully addresses the flicker noise leakage.

Based on the analysis in section 6.2, in fact the mixer in Fig.6.10b can be made to act in the same way as our mixer if the cross-coupled pair M_{4a}/M_{4b} is designed to operate during the whole LO period: the cross-coupled pair together the inductor in [2] is equivalent to the negative impedance proposed in our work. However, full flicker noise cancelling was not done in [2].

6.4 Simulation and measurement

The microphotograhps of two mixer chips (MixerD and MixerNF) are shown in Fig.6.11. The active area of the LO buffer and mixer with decap is $0.0156mm^2$ for the chip MixerD, of which 8.2% is occupied by the Y_{neg} circuit. In the chip MixerNF the Y_{neg} circuit consumes 7.1% of the total active area ($0.014mm^2$). The packaged chips were measured on PCB boards for 0.9GHz LO and 0.92GHz RF. The noise is measured by an Agilent E5500 noise set-up. For noise at IF<1MHz a SRS preamplifer is used connecting the mixer output with the noise set-up; for noise at IF>1MHz, a LeCroy AP033 active probe is used connecting the mixer output with the noise measurement set-up.

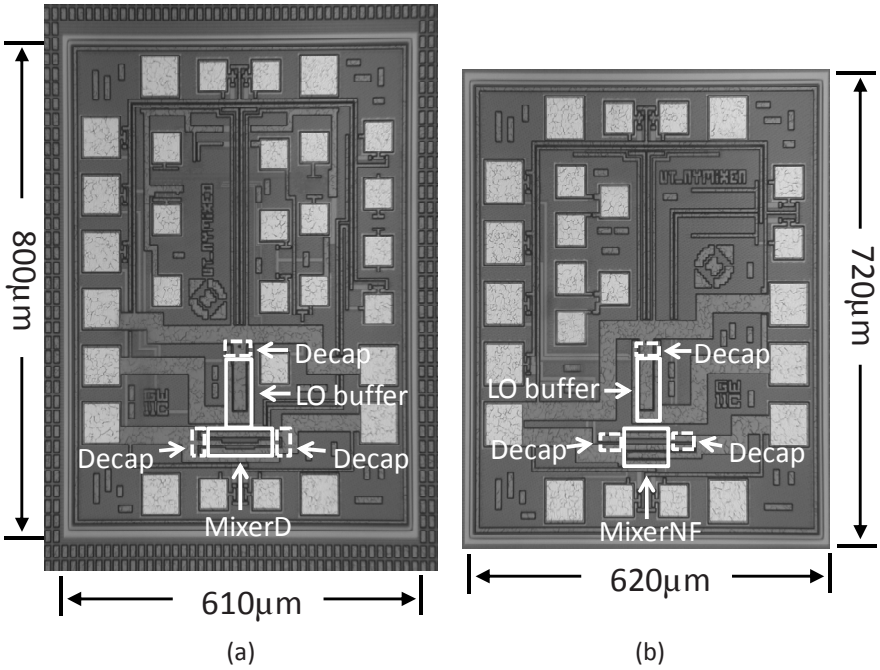


Fig. 6.11. Chip photo of (a) MixerD and (b) MixerNF.

6.4.1 Mixer with full-IM3/partial-flicker-noise cancellation

For the mixer optimized for full IM3 cancellation and partial flicker noise cancellation (MixerD), the bias current of Y_{neg} (I_{Yneg}) is swept within ±45% variation of the optimal value to verify the robustness against process spread. The measured and simulated results are shown in Fig. 6.12 as a function of the bias current normalized to the optimal value (NI_{Yneg}). At the optimal bias value (NI_{yneg}=1), a measured improvement of 10dB for IIP3, 5dB for conversion gain, 9dB for DSB NF@1kHz, and 1dB for input P_{1dB} are achieved compared to the same mixer without Y_{neg}. The DSB NF@10MHz degrades by 0.1dB. The mixer DC current increases from 9.2mA to 10.7mA due to the biasing of the Y_{neg} circuitry, while the LO buffer current (16mA) stays

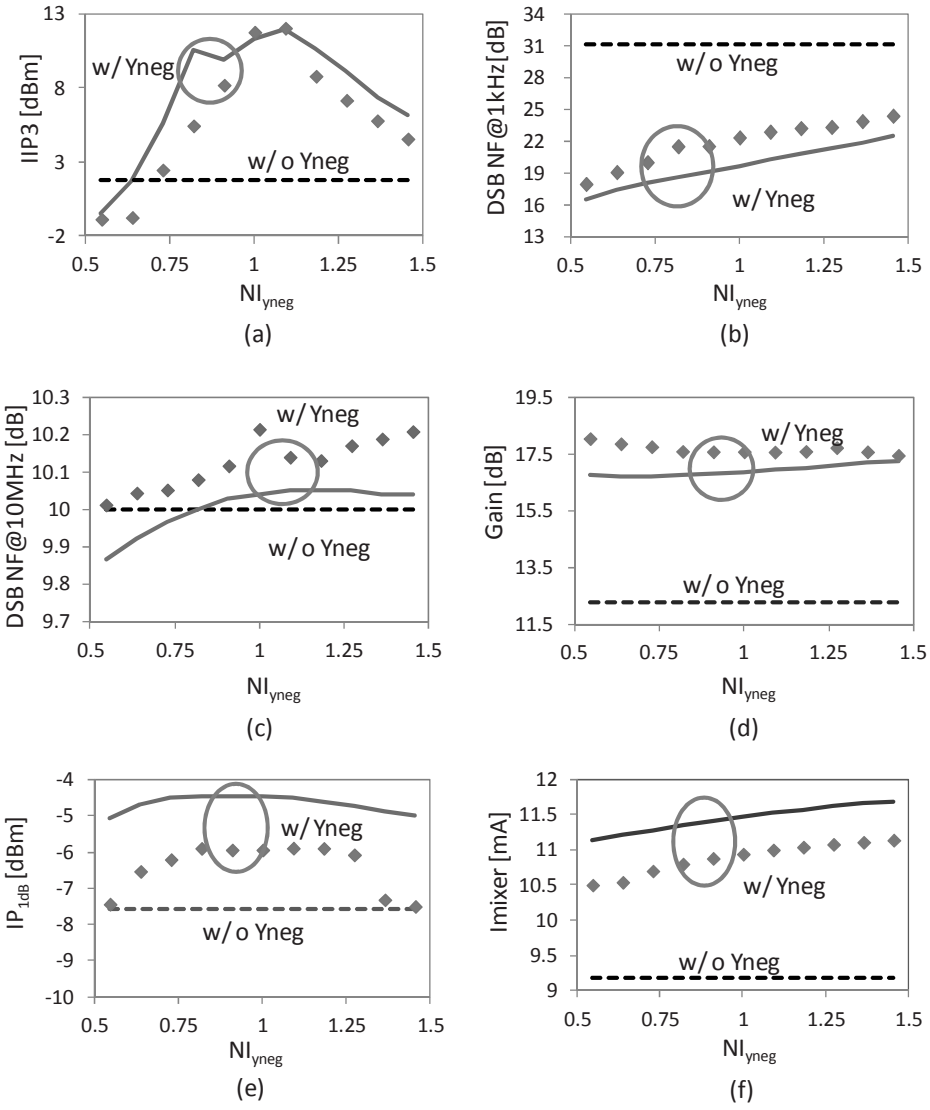


Fig. 6.12. IIP3, NF, conversion gain, input P_{1dB} and dc current taken by the mixer as a function of the nominated bias current of Y_{neg} for MixerD. Solid line for simulated results and symbol for measured results of the mixer with Y_{neg}. Dashed line for measured results of the mixer without Y_{neg}.

unchanged. Within $\pm$45% variation of I_{Yneg}, +5dB gain improvement, +6dB NF@1kHz reduction, <0.1dB thermal NF degradation, no input P_{1dB} degradation are achieved. As shown in Fig. 6.12b, we show the flicker NF reduction at a very low frequency (1kHz), where the flicker noise is dominant and the thermal noise can be neglected. Fig. 6.13 shows the measured HD1 and IM3

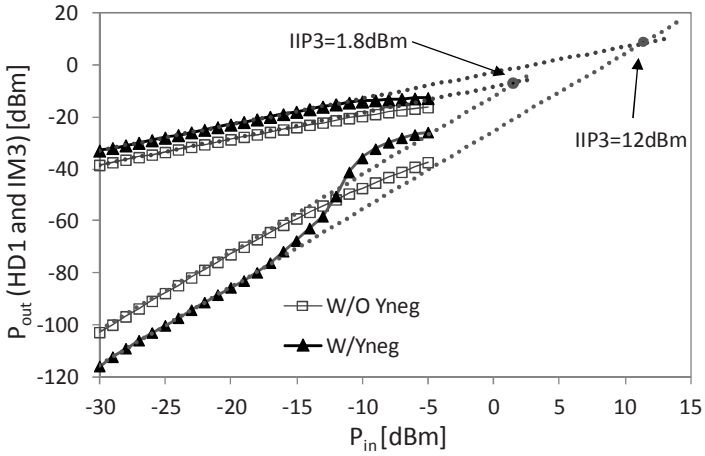


Fig. 6.13. Measured output HD1 and IM3 vs input power P_{in} for MixerD.

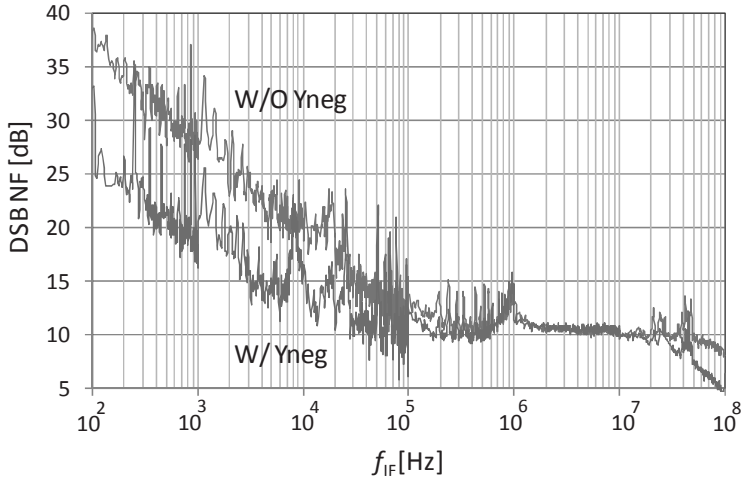


Fig. 6.14. Measured DSB NF for mixer (MixerD) with and without using Y_{neg}.

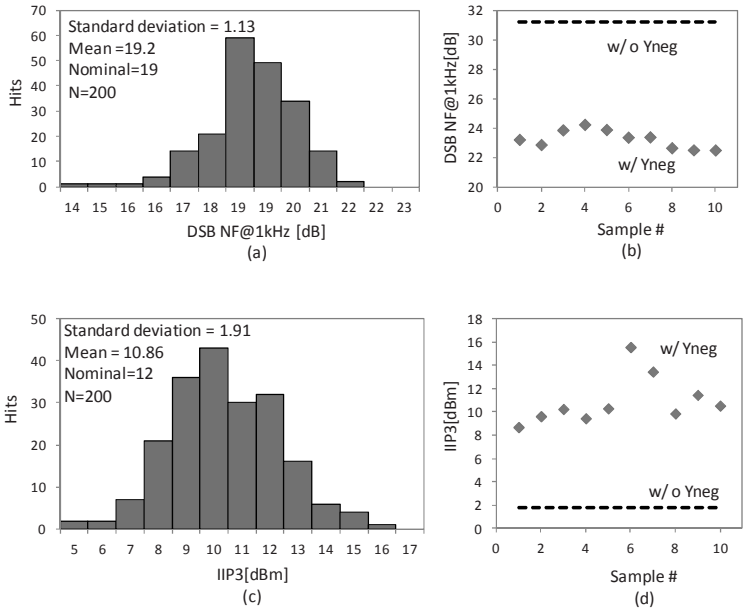


Fig. 6.15. Effect of mismatches and process spread on NF and IIP3 of MixerD with Y_{neg} (NI_{Yneg}=1). (a) 200-time Monte Carlo simulation results of DSB NF@1kHz (b) measured DSB NF@1kHz of ten dies, (c) 200-time Monte Carlo simulation results of IIP3, (d) measured IIP3 of ten dies. Symbol for measured results of the mixer with Y_{neg}. Dashed line for measured results of one mixer sample without Y_{neg}.

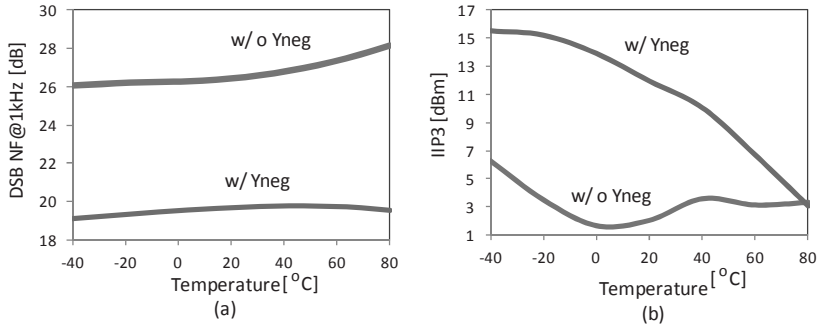


Fig. 6.16. Simulated NF and IIP3 for mixer (MixerD) with and without using Y_{neg} as a function of temperature.

output at the optimal bias value (NI_{Yneg}=1). Due to higher-order nonlinearity distortion introduced by Y_{neg}, the IM3 curve start to show 5th order behavior for Pin >-18dBm. The measured mixer DSB NF is shown in Fig. 6.14. The spikes in the output noise PSD are from the equipment power supply and the measurement setup. Although Y_{neg} introduces 5dB thermal noise to the mixer, 5dB more gain also provided by Y_{neg} lowers the input-referred noise of the R_{load} by 5dB, and results in <0.1dB degradation in the thermal noise figure. The 1/f corner frequency decreases from 100kHz to 20kHz.

The effect of mismatches and process spread on NF and IIP3 are shown in simulation and measurement results in Fig. 6.15. A 200-time Monte Carlo simulation in Fig. 6.15(a) shows a mean DSB NF@1kHz of 19.2dB (nominal is 19dB) which is about 7dB lower than the mixer without Y_{neg}, while the measurement of ten dies from one wafer shows +7dB NF reduction at 1kHz in Fig. 6.15(b). Note that such high NF is not the result a badly designed mixer but is due to the very low frequency (1kHz), where the flicker noise is significant. In comparison, the measured NF@1kHz of the low-flicker-noise mixer in [2] is 29dB. In Fig. 6.15(c) a 200-time Monte Carlo simulation shows mean IIP3 of 10.8dBm (nominal is 12dBm) which is 9dB higher than the mixer without Y_{neg}, whereas +6dB IIP3 improvement are measured in ten dies from one wafer as shown in Fig. 6.15(d). For the temperature range (-40^{o}C to 80^{o}C) in the nominal corner, simulations show +6.7dB 1/f NF reduction in Fig. 6.16a, whereas the IM3 cancellation becomes less effective as the temperature increases.

6.4.2 Mixer with full-flicker-noise / partial-IM3 cancellation

For the mixer optimized for full flicker noise cancellation (MixerNF), Fig. 6.17 shows the measured and simulated results as a function of the bias current for Y_{neg} normalized to the optimal value (NI_{Yneg}). At the optimal bias value (NI_{yneg}=1), a measured improvement of 8dB for DSB NF@1kHz, 1.4dB for conversion gain, 0.1dB for the DSB NF@5MHz and 2.5dB for input P_{1dB} are achieved compared to the same mixer without Y_{neg}. The mixer dc current increases by 4% due to the biasing of the Y_{neg} circuitry, while the LO buffer current (4.8mA) stays unchanged. The difference between the measured and

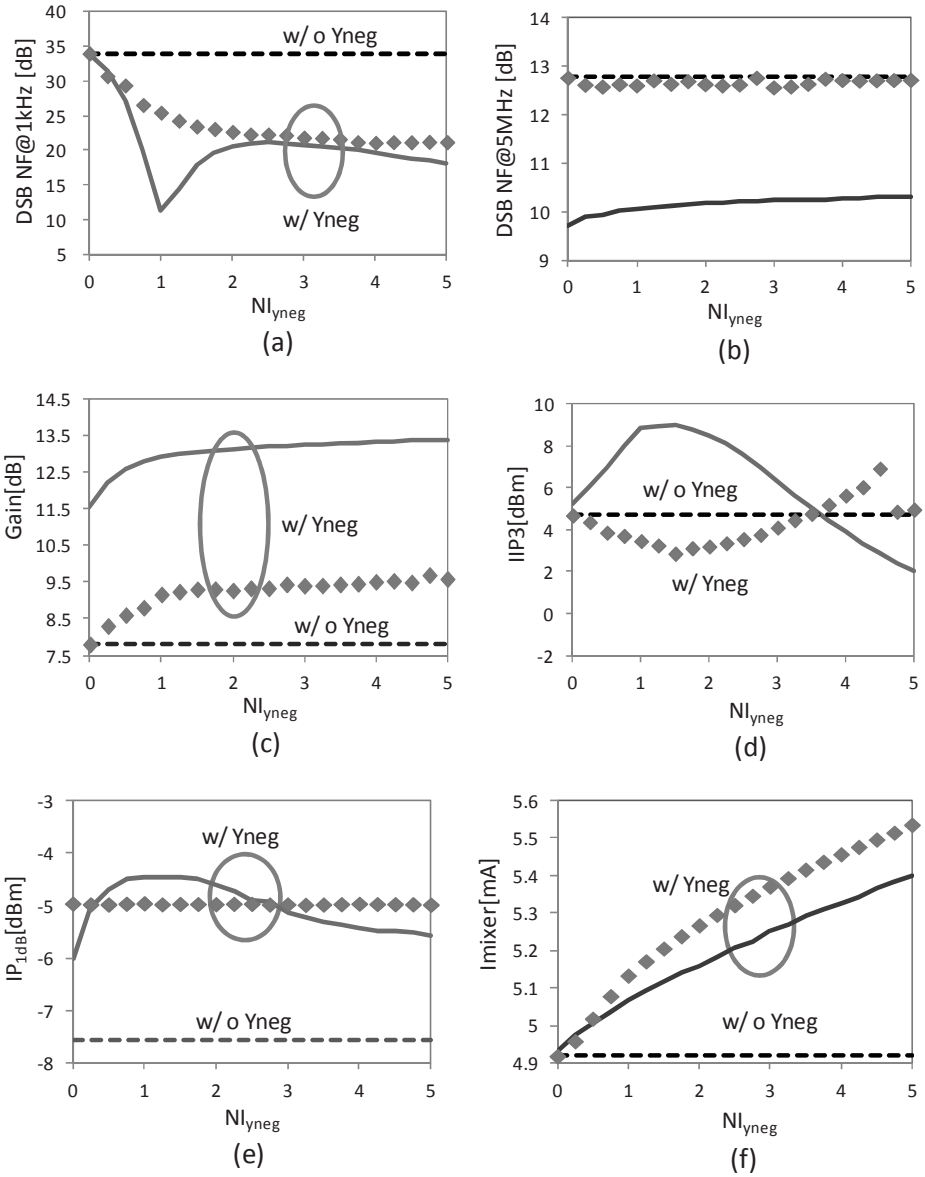


Fig. 6.17. NF, IIP3,conversion gain, input P_{1dB} and dc current taken by the mixer as a function of the nominated bias current of Y_{neg} for MixerNF. Solid line for simulated results and symbol for measured results of the mixer with Y_{neg}. Dashed line for measured results of the mixer without Y_{neg}.

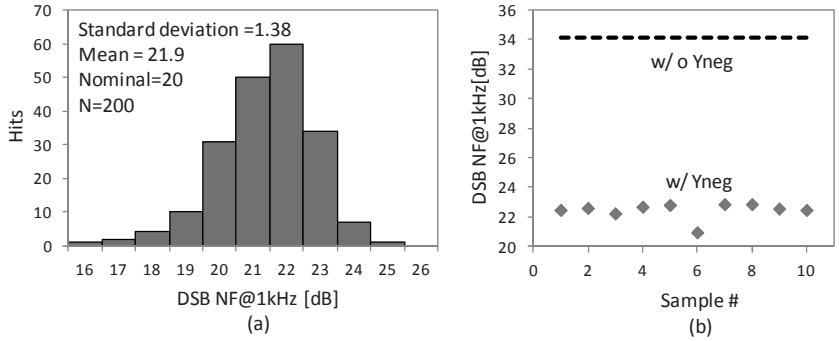


Fig. 6.18. (a) 200-time Monte Carlo simulation results of DSB NF@0.1kHz for MixerNF (NI_{yneg}=2). (b) Measured DSB NF@0.1kHz for MixerNF (NI_{yneg}=2) of ten dies.

simulated IIP3 shown in Fig. 6.17(d) may be due to the fact that this mixer is operated with low supply voltage ($0.67\times V_{DD}$), where the headroom for Y_{neg} circuit is not sufficient to provide robust IM3 cancellation. Fig. 6.17(a) shows that, at the optimal bias value (NI_{yneg}=1), a full flicker noise cancellation suggested by simulation is not found in measurement, due to mainly external low-frequency noise contributed by the measurement set-up and due to flicker noise leakage resulting from mismatches in the mixer. However, Fig. 6.17(a) also shows >10dB improvement for flicker NF can be achieved for very broad bias range (NI_{Yneg}>1.75). The robustness of this flicker cancellation under low supply voltage is then tested by setting NI_{Yneg}=2 in MixerNF. In Fig. 6.18(a), a 200-time Monte Carlo with mismatches and process spread shows a mean DSB@1kHz of 21.9dB (nominal value is 20dB) that is 15dB lower than in the mixer without Y_{neg}. The measurement of ten dies shows +10dB flicker NF reduction in Fig. 6.18(b). For the temperature range (-40°C to 80°C) in nominal corner, the simulation shows +14dB 1/f NF reduction in Fig. 6.19. The measured mixer noise output is shown in Fig. 6.20 (for Y_{neg} biased at NI_{yneg}=2). The 1/f corner frequency decreases from 200kHz to 20kHz. Note that the rolling-off behavior for IF>5MHz is due to the IF filter.

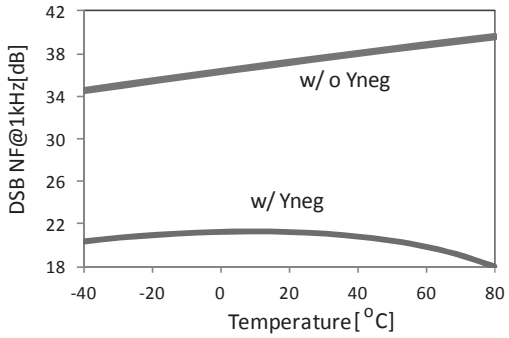


Fig. 6.19. Simulated NF for mixer (MixerNF) with and without using Y_{neg} as a function of temperature.

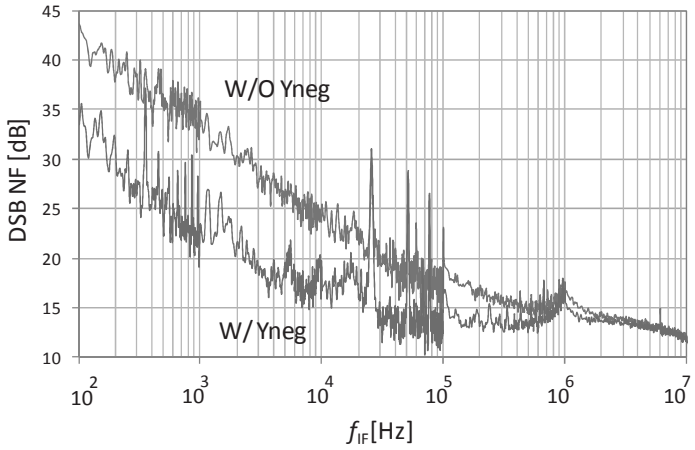


Fig. 6.20. Measured DSB NF for mixer with and with using Y_{neg} for MixerNF (NI_{yneg}=2).

6.4.3 Benchmarking

The mixer with full-IM3/partial-flicker noise cancellation (MixerD) presented in this chapter is compared with previous works on flicker noise reduction [1-4] in Table 6.1. Since the flicker NF value depends on a few factors such as circuit bias and technology-related flicker noise corner, our technique is compared with previous works in term of the value of flicker

Table 6.1. Comparison of techniques for flicker noise reduction in CMOS active mixers.

		Darabi[1]	Yoon[2]	Park[4]	Pullela[3]	This work
CMOS		0.13µm	0.13µm	0.18µm	0.13µm	0.16µm
V_{DD} [V]		1.2	1.5	1.8	2.7	1.8
Freq [GHz]		2	2.4	5.2	1.96	0.9
Inductor Number		0	1	2	0	0
Flicker NF reduction [dB]		7@10kHz	7.5@10kHz	7@10kHz	9.5@10kHz	9@1kHz
Flicker NF @10kHz [dB]	w/o cancelling	21.8	37	27	20	18
	w/ cancelling	14	29	20	13	8.5
Gain Improvement [dB]		0.5	1.3	6	2	5
Thermal NF changes∗ [dB]		0	0	0	-2	+0.1
Thermal NF @10MHz [dB]	w/o cancelling	12	7.5#	10	10	8.5
	w/ cancelling	12	7.5#	10	10.1	6.5
IIP3 improvement [dB]		0	1.6	0	1	10
Bias current increases		0%	0%	0%	N/A	16%

∗ "+" for NF increases and "–" for NF decreases.
This NF is measured at 100MHz

noise reduction. It shows that the presented technique provides very good flicker NF reduction, while at the same time it achieves the largest improvement in IIP3 and gain without using on-chip inductors or high supply voltages or increasing the LO power. In conclusion, this flicker noise/IM3 cancellation provides solutions for reducing flicker noise and improving linearity of CMOS active mixers.

6.5 Conclusion

A new technique providing simultaneous cancellation of flicker noise and IM3 distortion for active mixers is presented without using on-chip inductors or high supply voltages or increasing the LO power. By using a negative impedance (Y_{neg}), the flicker noise leakage from the switching pairs is minimized. Meanwhile the negative impedance enables IM3 distortion cancellation between the switching pairs and the transconductor stage, which yields overall IM3 improvement. The techniques also improve the conversion gain while has little effect on the thermal noise. For the demonstration mixer chip optimized for full-IM3/partial-flicker-noise cancellation, 9dB flicker noise suppression, 10dB improvement for IIP3, 5dB improvement for conversion gain and 1dB improvement for input P_{1dB} are achieved. The Y_{neg} circuit increases the thermal NF by 0.1dB, power consumption by 16% and active area by 8%. Under the effect of mismatch and process spread, a 200-time Monte Carlo simulation shows 7dB reduction in mean NF@1kHz and 9dB increase in mean IIP3. A ten-sample measurement shows >7dB reduction in NF@1kHz and >6dB increase in IIP3. The simulation demonstrates that the flicker noise cancellation is not sensitive to temperature variation [-40^{o}C to 80^{o}C], while the IM3 cancellation degrades as the temperature increases. For the demonstration mixer chip optimized for full-flicker-noise/partial-IM3 cancellation under low supply voltage ($0.67 \times V_{DD}$), >10dB flicker noise suppression is measured within $\pm$200% variation of the negative impedance bias current. The ten-sample measurement shows >11dB flicker NF reduction, and the simulation shows >14dB flicker NF reduction for the temperature range [-40^{o}C to 80^{o}C].

6.6 **Appendix**

The real part of (6.6) is given by

$$Re\left(v_{fl,out}^{M_{1a}}\Big|_{t_1}\right)=\frac{-g_m^{M_{1a}}R_L\left[G^{M_3}\left(g_m^{M_{1a}}+G^{M_3}\right)+\left(\omega_{LO}C^{M_3}\right)^2\right]v_{fl,in}^{M_{1a}}}{\left(g_m^{M_{1a}}+G^{M_3}\right)^2+\left(\omega_{LO}C^{M_3}\right)^2} \tag{A6.1}$$

where $G^{M_3}=g_{ds}^{M_{3a}}+2G_{neg}$ and $C^{M_3}=C_{tail}+2C_{neg}$.

$$\text{When } \begin{cases} G_{neg}>\frac{1}{4}\left(-2g_{ds}^{M_{3a}}-g_m^{M_{1a}}-\sqrt{{g_m^{M_{1a}}}^2-(2\omega_{LO}C^{M_3})^2}\right) \\ G_{neg}<\frac{1}{4}\left(-2g_{ds}^{M_{3a}}-g_m^{M_{1a}}+\sqrt{{g_m^{M_{1a}}}^2-(2\omega_{LO}C^{M_3})^2}\right) \end{cases} \tag{A6.2}$$

the real part of (6.6) is positive. For $C_{neg}\approx -0.5C_{tail}$, $C^{M_3}\approx 0$ and (A6.2) can be simplified to

$$\frac{-g_{ds}^{M_{3a}}-g_m^{M_{1a}}}{2}<G_{neg}<\frac{-g_{ds}^{M_{3a}}}{2} \tag{A6.3}$$

Therefore, for $G_{neg}\in\left[\frac{-g_{ds}^{M_{3a}}-g_m^{M_{1a}}}{2},\frac{-g_{ds}^{M_{3a}}}{2}\right]$ the real part of (6) is positive.

6.7 References

[1] H. Darabi and J.Chiu, "A noise cancellation technique in active RF-CMOS mixers," *IEEE J. Solid-State Circuits,* vol. 40, pp. 2628-2632, Dec. 2005.

[2] J. Yoon, H. Kim, C. Park, J. Yang, H. Song, S. Lee and B. Kim, "A new RF CMOS Gilbert mixer with improved noise figure and linearity," *IEEE Trans. Microwave Theory and Techniques,* vol. 56, pp. 626-631, Mar. 2008.

[3] R. S. Pullela, T. Sowlati and D. Rozenblit, "Low flicker-Noise quadrature mixer topology," *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers,* San Francisco, CA, 2006, pp. 1870 – 1879.

[4] J. Park, C.H Lee, BS Kim and J. Laskar, "Design and analysis of low flicker-noise CMOS mixers for direct-conversion receivers", *IEEE Trans. Microwave Theory and Techniques,* vol. 54, pp. 4372-4380, Dec. 2006.

[5] B. Razavi, "Design considerations for future RF circuits", *IEEE Internationl Symposium on Circuits and Systems*, pp.747–744, 2007.

[6] H. Darabi and A. A. Abidi, "Noise in RF-CMOS mixers: a simple physical model," *IEEE J. Solid-State Circuits,* vol. 35, pp. 15-25, Jan. 2000.

[7] W. Cheng, A. J. Annema, J. A. Croon and B. Nauta, "Noise and nonlinearity modeling of active mixers for fast and accurate estimation", *IEEE Trans. Circuits and Systems I,* vol. 58, pp. 276-289, Feb. 2011.

[8] C. C. Enz and G. C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: Autozeroing, correlated double sampling, and chopper stabilization," *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584–1615, Nov. 1996.

[9] C. Tilhac, S. Razafimandimby, A. Cathelin, S. Bila, V. Madrangeas and D. Belot, "A Tunable bandpass BAE-filter architecture using negative capacitance circuitry," *IEEE Radio Frequency integrated Circuits (RFIC)Symposium,* pp.605 – 608, 2008.

[10] J. C. Zhan, K. Maurice, J. Duster and K. T. Kornegay, "Analysis and design of negative impedance LC oscillators using bipolar transistors," *IEEE Trans. Circuits and Systems I,* vol. 50, pp. 1461-1464, Nov. 2003.

[11] R. Brederlow, W. Weber, C. Dahl, D. Schmitt-Landsiedel and R.Thewes, "Low-frequency noise of integrated poly-silicon resistors," *IEEE Trans. Electron Devices,* vol. 48, nl.6, pp. 1180-1187, Nov. 2001.

Chapter 7

Conclusions

7.1 **Conclusions**

In this thesis, we have developed circuit models that predict the noise and distortion behavior of two major RF building blocks, namely, LNAs and active mixers. Without involving complex calculations, our models provide both good understanding of the circuits and sufficient estimation accuracy. For time-invariant weakly nonlinear systems such as LNAs, our model only uses AC transfer functions together with technology dependent transistor nonlinearity parameters to estimate the circuit distortion. For time-varying active mixers, our noise model uses two AC calculations to predict the noise behavior; our distortion model uses a few time-invariant distortion calculations to estimate IIP3 and IIP2.

For the circuit synthesis using numerical computing power, e.g., circuit design automation (CAD), the simplicity of our models enables a fast first-step coarse optimization. The result can be used as a very good starting point for a numerical optimizer that performs a second-step fine optimization.

For the circuit synthesis using circuit designers' brain power, our simple models provide a good understanding of circuits' behavior that potentially can be useful for circuit optimizations and innovations. Thanks to the insights provided by our model, new circuit concepts for reducing distortion and noise for LNAs and Gilbert mixer are inspired and presented in this thesis.

Our model suggests that, to improve the linearity of LNAs with cascode topologies, one way is to bias the transistors in the moderate inversion region. This enables a distortion cancellation between the transconductance nonlinearity and other nonlinearities that is related to output conductance and cross terms. Compared to strong inversion, in the moderate inversion region, less dc current is required for the same g_m, resulting less power consumption and higher P_{1dB} due to more headroom for the output swing. The price that has to be paid for this is a larger transistor. Simulations show that, for both common source amplifier and common gate LNA, this IM3 cancellation scheme provides robustly more than 10dB IIP3 improvement for signal frequencies up to 5GHz in a 90nm CMOS process. Alternatively, a negative impedance can be used to enable distortion current cancellation between the transconductor and the cascode transistor, as discussed in chapter 3. For a resistive feedback LNA implemented in a 0.16µm standard CMOS process, measurements show that for 0.1GHz to 1GHz, improvements of 6.3dB to 10dB for IIP3 and 0.2dB to 1dB for gain are achieved without noise degradation. The negative impedance increases the power consumption for the LNA by 2%, and increases the die area by about 700μm^2.

As suggested by our model, the negative impedance is also used to simultaneously cancel flicker noise and IM3 in Gilbert mixer. As presented in chapter 6, the demonstration chip achieves 9dB flicker noise suppression, improvements of 11dB for IIP3, 5dB for conversion gain, and 1dB for input P_{1dB} while the thermal noise increased by 0.1dB. The negative impedance increases the power consumption for the mixer by 16%, and increases the die area by 8% (46x28μm^2).

To obtain a compact highly linear wideband attenuator, we applied a wideband IM3 cancellation technique to CMOS attenuators which enable the distortion currents of the series and shunt devices within the attenuator cancel each other. The full IM3 cancellation depends on transistors' W/L-ratios for

given attenuation value, which provides robustness to PVT variations and enables high linearity with small active area. In a 0.16μm standard bulk CMOS process, a 4-step Π attenuator system using this IM3 cancellation technique achieves >3dBm input P_{1dB} and >26dBm IIP3 for 50MHz-5GHz. The active area is only 0.0054mm^2. A 4-step T attenuator system using this IM3 cancellation technique achieves >11dBm input P_{1dB} and >27dBm IIP3 for 50MHz-5.6GHz. The active area is 0.0067mm^2. Compared to literature published before 2012, these two attenuators achieve very high linearity with the smallest silicon area.

7.2 Original contributions

- The general weak nonlinearity model for LNAs that uses the time-invariant linear analysis to estimate the time-invariant weakly nonlinear circuit (Chapter 2).
- The analysis of the mechanism that causes the increasing distortion generation by the cascode transistor in deep-submicrometer technologies (Chapter 2).
- The approach of biasing transistors in moderate inversion region to enable distortion cancellation between the transconductance nonlinearity and other nonlinearities e.g. output conductance nonlinearity and cross-term nonlinearities for LNAs with cascode topology (Chapter 2).
- The concept of using a negative impedance to enable distortion current cancellation between the transconductor and the cascode transistor for LNAs with cascode topology (Chapter 3).
- The concept of using distortion currents of switches within the attenuator to cancel each other and achieve high IIP3 (Chapter 4).
- The noise and linearity model for active mixer. The concept of using two time-invariant linear analyses to estimate the time-varying linear analysis for noise. The concept of using a few time-invariant weakly nonlinear analyses to estimate the time-varying weakly nonlinear analyses for IIP2 and IIP3 (Chapter 5).

- The analysis of significant effect of the transconductor's finite output resistance on the flicker noise leakage and distortion contribution of the switching pair (Chapter 5).
- The analysis of significant significant effect of the LO slope, the cross-modulation nonlinearity and output conductance nonlinearity in the triode region on IIP2 (Chapter 5).
- The concept of using a negative impedance to simultaneously cancel flicker noise and IM3 distortion for Gilbert mixer (Chapter 6).

7.3 Recommendation for future work

Started from investigating transistor nonlinearities, the research work of this thesis ended in obtaining insights on distortion behavior of individual circuit blocks, which was defined naturally by the thesis title. Consequently, all proposed techniques for the linearity optimization or distortion cancellation tend to provide "local" solutions: considering one circuit block only and remain effective for weakly nonlinear system. As a result, the IM3 cancellation for LNAs either by biasing in the moderate inversion region or using a negative impedance becomes less effective for larger input power (P_{in}>-18dBm in our demonstration circuits). This is due to high-order nonlinearity e.g. 5^{th} order. The IM3 cancellation for Gilbert mixer faces the similar limitation. In order to provide linearity improvement for large input power, a more systematic way may be required, which is worth of further investigations.

The noise and linearity model for active mixers is based on the assumption that the transient effects of the capacitances do not change the dc bias of transistors. This is valid simplification only for low GHz region in modern CMOS process. Accurate models for active mixers that operate at very high frequencies (e.g. 60GHz) require the inclusion of parasitics in determining the instantaneous dc point for transistors.

The negative impedance circuit used in chapter 3 and chapter 6 works only up to low GHz region. Further research is desirable on new topologies of negative impedance that can work for very high frequency (e.g. 60GHz).

Supplementary material 1

for Chapter 1, published in *IEEE Int. Symp. Circuits Syst.(ISCAS)*, pp.2550-2553, May 2008.

A multi-step P-cell for LNA design automation

Wei Cheng, Anne Johan Annema and Bram Nauta
IC Design Group, CTIT Research Institute, University of Twente
PO Box 217, 7500 AE Enschede, The Netherlands
w.cheng@utwente.nl

***Abstract*—This paper presents a novel way to efficiently implement parametric cells (P-cells). A narrow-band LNA is used as demonstration vehicle. In the P-cell, circuit parameters such as transistor size, bias condition and passive values are determined automatically for any given reachable target performance. To achieve both high accuracy and relatively high speed, a new iterative stepped approach is used with respect to speed and accuracy, starting with moderate-accuracy and fast optimization that yields the starting point for the next higher-accuracy and slower optimization step. The presented approach can be extended to other types of circuits.**

I. INTRODUCTION

The low noise amplifier (LNA) is a critical building block in any RF front-end; it has an important effect on the noise performance of the overall system. As the market for wireless communication expands, the need for LNA with demanding performance specifications is increasing. However, like any other RF block, the design of LNAs is a time-consuming task that typically relies heavily on the experience of RF designers. LNA design automation can significantly simplify the design task and shorten the design-to-market time.

Available LNA design automation is mainly based on circuit synthesis [1-3]. Starting at some initial circuit component values such as transistor size, bias condition and passive values, the circuit performance metrics are calculated and then a cost function is evaluated. Typically, the calculation of circuit performance is done using conventional analog circuit simulators. Adaptation of circuit component values is done using e.g. a gradient descent algorithm that optimizes the pre-specified cost function. Already for small sized circuits the number of parameters is large and numerical optimization costs lots of time, while there is usually no guarantee that the algorithm finds a solution. In mathematical terminology the main problems are due to the large search space and sticking in local minima.

To speed up the LNA design automation process a novel multi-step approach is proposed that can solve the traditional drawbacks of optimization:

- The user selects a circuit topology and specifies target performance metrics, which are inputted to the P-cell. A straight-forward extension is that also the best performing (optimized) circuit out of a set of circuit topologies is selected automatically by the P-cell based on defined cost function evaluation.
- The P-cell first makes a coarse optimization of circuit component based on the input specifications. This first step uses circuit analyses for noise and harmonics modeling. A built-in interface with state-of-the-art MOS models such as MOS model 11 (MM11) [4] or PSP [4] extracts the information of intrinsic noise and nonlinearities of the transistor. As a result this first step has moderate accuracy but is very fast.
- The result of the coarse optimization is used as starting point for a numerical optimizer, wrapped around conventional simulators such as Spectre [5]. Because of the relatively good starting point for the numerical optimizer this second optimization is very fast.
- The results of the second optimization can be used as settings for the layout generation. Extraction and optimization on the extracted circuit, taking into account layout parasitics (including those of passive components) and variability, can increase accuracy at a significant calculation time penalty.

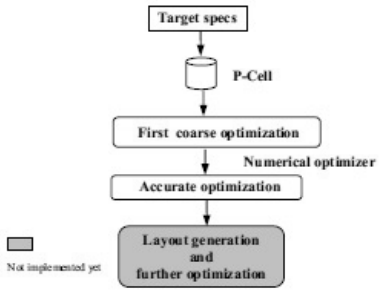


Figure 1. Block diagram of the multi-step approach for LNA design automation.

This paper is organized as follows. Section II discusses the approach for the building of P-cell. Section III discusses the implementation of the P-Cell, which consists of the LNA modeling and numerical root-finding. Section IV demonstrates

This work is funded by NXP Semiconductors Research.

the design automation of a narrow band LNA using the proposed multi-step approach.

II. P-CELL BUILDING APPROACH

The goal of the P-cell is illustrated in Fig. 2: for any specified target circuit performance, the P-Cell first chooses the optimal LNA topology and then calculates the optimal values for each circuit component. In other words, the P-Cell determines the circuit parameters such as transistor size, bias condition and passive value automatically for any given circuit performance specifications, which is a reverse-direction approach compared with the commonly-used approach in the synthesis [1-3] (calculating the circuit performance metrics from chosen circuit parameters).

As the first step in the optimization process carried out in the P-cell, a coarse circuit optimization is done. This coarse optimization is fast and with moderate accuracy in reaching the target performance metrics. The realization of the reverse-direction approach is based on two blocks: LNA modeling and numerical root-finding, as illustrated in Fig. 3.

- In the **LNA modeling** block, small signal analyses including noise and harmonics are performed. State-of-the-art MOS models such as MOS model 11 or PSP [4] are embedded in the P-cell providing direct access to MOS transistor parameters and properties. As a result, the mathematical link between the LNA performance metrics with the circuit parameters such as bias, component values is built.
- **Numerical root-finding**. Based on the mathematical link built by the LNA model, the numerical root-finding algorithm performs the reverse-direction calculation for any given circuit performance metrics, which determines the value of the circuit parameters.

III. IMPLEMENTATION OF THE P-CELL

To demonstrate the multi-step design automation approach, a P-cell for the narrow-band inductively degeneration common source (IDCS) LNA [6] shown in Fig. 4a is implemented.

A. LNA Modeling

The linear small signal model in Fig. 4b is used to calculate the noise performance, input and output impedance and gain. All the parasitics of the transistor such as C_{gd}, C_{gs}, C_{db}, r_{db} are accounted for. The noise sources included are

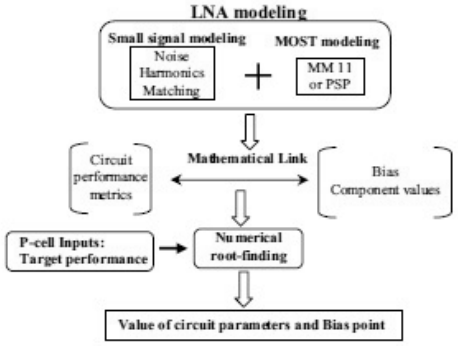


Figure 3. Block diagram of the reverse-direction approach in the P-Cell.

namely, gate induced noise $\overline{i_{g,n}^2}$, channel noise $\overline{i_{d,n}^2}$, flicker noise $\overline{e_{flicker,n}^2}$, parasitic resistance noise of gate and source inductors $\overline{e_{Lg,n}^2}$ and $\overline{e_{Ls,n}^2}$, resistive load noise $\overline{e_{Rload,n}^2}$.

In most of the previous work on LNA nonlinearity modeling, only the transconductance nonlinearities are taken into account [1], [6-8]. However, we have observed that for short channel MOST with moderate voltage gain also the nonlinear output conductance and cross modulation terms play a big role. For operating frequencies higher than 1 GHz the effect of the transistor's nonlinear capacitances also becomes important. Aiming for fairly accurate nonlinearity modeling a complete weakly nonlinear model is used shown in Fig.4 (c). This model includes four nonlinear current sources, namely, $i_{Rds}[v_{gs}(t), v_{ds}(t)]$, $i_{Cds}[v_{gs}(t), v_{ds}(t)]$, $i_{Cgs}[v_{gs}(t), v_{ds}(t)]$ and $i_{Cgd}[v_{gs}(t), v_{ds}(t)]$. Source i_{Rds} accounts for the nonlinear resistive drain-source current; sources i_{Cds}, i_{Cgs}, i_{Cgd} account respectively for the nonlinear capacitive drain-source current, for the nonlinear capacitive gate-source current and for the nonlinear capacitive gate-drain current. Assuming a zero source-bulk voltage for the transistor, these four sources are functions of MOST terminal voltage v_{gs} and v_{ds}, which are described by two-dimensional Taylor series expansions. Equations (1) and (2) show the series expansions used in this work for the weakly nonlinear resistive and capacitive current sources. The Taylor series' coefficients such as g_{x11} and C_{dsX21} are extracted directly from the embedded MOS models.

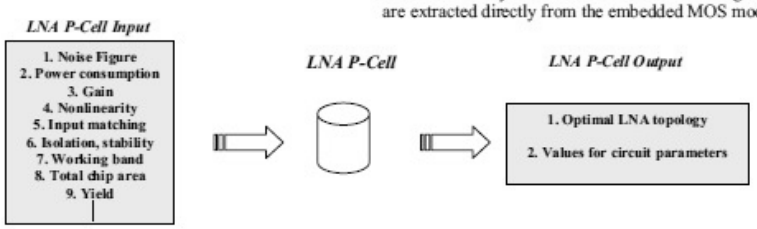


Figure.2 The function of the P-cell

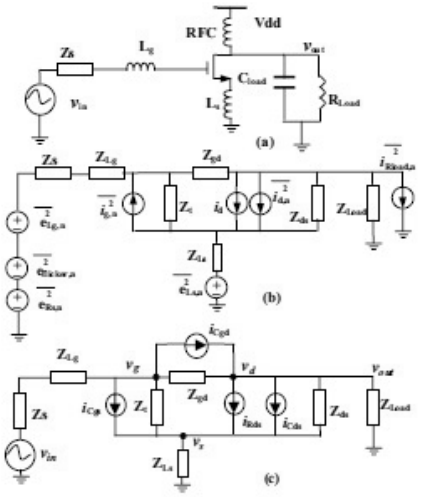


Figure 4. (a) Schematic of the IDCS LNA, (b) and its noise small signal model, (c) its nonlinearity model.

$$i_{Rds} = g_{m1}v_{gs} + g_{m2}v_{gs}^2 + g_{m3}v_{gs}^3 + g_{dd1}v_{ds} + g_{dd2}v_{ds}^2 + g_{dd3}v_{ds}^3 + g_{X11}v_{gs}v_{ds} + g_{X12}v_{gs}v_{ds}^2 + g_{X21}v_{gs}^2v_{ds} \quad (1)$$

$$Q_{Cds} = C_{ds1}v_{gs} + C_{ds2}v_{gs}^2 + C_{ds3}v_{gs}^3 + C_{dd1}v_{ds} + C_{dd2}v_{ds}^2 + C_{dd3}v_{ds}^3 + C_{dd,X11} \times v_{gs}v_{ds} + C_{dd,X12} \times v_{gs}v_{ds}^2 + C_{dd,X21} \times v_{gs}^2v_{ds} \quad (2)$$

$$i_{Cds}(t) = \partial_t Q_{Cds}$$

A distortion analysis method similar to the one in [9] is used to derive the closed-form formulas for IIP2 and IIP3. To the best of our knowledge it is the first time that a complete weakly nonlinear model of transistor valid for all operating frequency has been used for LNA nonlinearity analysis. The combination of direct access to the embedded MOS model and the complete nonlinear modeling of LNA – including mixing terms and nonlinear capacitances – leads to a fairly accurate optimization result at relatively high speed for the first optimization step in the P-cell. This optimization result is then used for the numerical optimizer (second-step optimization).

B. *Numerical root-finding*

The mathematical link obtained in the LNA modeling provides the core relations between circuit performance metrics (noise figure NF, IIP3, IIP2, power consumption, S_{11} and S_{21}) and circuit parameters (V_{gs}, the transistor aspect ratio W/L, Ls, Lg). Under the constraint of input matching, numerical root-finding algorithms such as a Newton-Raphson (N-R) method are used to obtain the numerical solutions for the core relations, as illustrated in Fig. 5.

C. *Capability of P-cell*

Our P-cell implementation includes a number of functions, including:

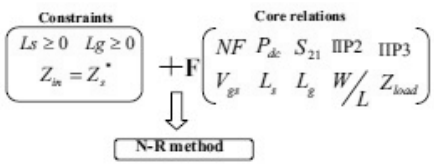


Figure 5. Block diagram of the numerical root-finding processing.

- Providing information about the reachable level of circuit performance, given some boundary conditions.
- Determination of component sizes and values for a set of target performance metrics.
- Calculation of the realizable minimum NF and maximum S21, IIP2 and IIP3 for given operating conditions (operating frequency, output impedance of the input voltage source, power consumption range, load impedance and supply voltage).

IV. MULTI-STEP LNA DESIGN AUTOMATION

To demonstrate the multi-step design automation approach the IDCS LNA shown in figure 4a is designed for the operating condition listed in Table I, with various target values for NF, IIP2 and IIP3; a commercial standard 90nm CMOS process is assumed for the design

TABLE I. DESIGN VEHICLE BOUNDARY CONDITIONS

f (GHz)	V_{dd} (V)	Q_{Lg}	Q_{Ls}	P_{DC} (mW)	C_{Load} (fF)	R_{Load} (ohm)	Zs (ohm)
5	0.6	10	10	<8	100	2000	50

An industrial numerical circuit optimizer is used as the second-step accurate optimizer. This numerical optimizer is wrapped around Spectre to optimize component values, bias conditions and more to satisfy a user-defined cost function. The optimization is done by iteratively evaluation of the (Spectre) simulation result and calculating a new set of component values and bias settings that reduce the cost function.

Firstly Fig. 6 shows the accuracy and speed of the first coarse optimization step in our P-cell for the specifications in Table 1 as a function of target NF (Fig. 6a), target IIP2 (Fig. 6b) and IIP3 (Fig. 6c). Each graph shows on the left y-axis the error between the target specification – e.g. target NF – and the actual value that follows from Spectre simulations using the exact component values and biasing settings as provided by the P-cell. The right y-axis shows the calculation time required for the P-cell on a Linux server with a 3 GHz Intel Xeon CPU and 3 GB memory. Note that within 2 seconds quite accurate results are provided by our P-cell, which is due to the fully modeling of noise and nonlinearity including e.g. conductance cross modulation terms and nonlinear capacitances of the MOS transistor.

Secondly a design example is used to show the overall performance of the P-cell with two-step optimization on hard

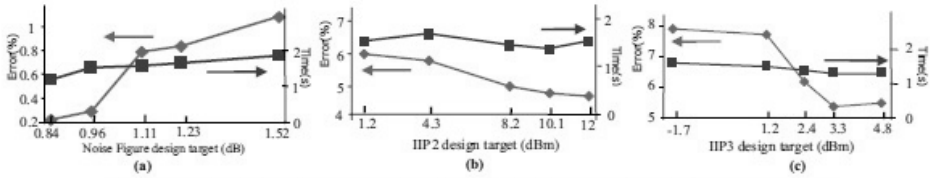


Figure.6 Benchmark of P-cell on accuracy and speed in meeting the design targets of (a) Noise figure, (b) IIP2 , (c) IIP3.

target specifications, for high performance applications. Table II lists the design targets and the Spectre-simulated performance using the exact component values and bias settings as provided by the P-cell after first coarse optimization and two-step optimization respectively. Because the first-step coarse optimization provides good initial settings for the second optimization step with high speed, the total optimization time is relatively short (around 15 seconds). Fig 7 shows Spectre simulation result of the circuit as optimized by the P-cell with two-step optimization.

TABLE II. DESIGN TARGET AND P-CELL RESULTS

	IIP3 (dBm)	IIP2 (dBm)	NF (dB)	V_{gain} (dB)	S11 (dB)	Pdc (mW)
Target	>5	>20	<1	>12	<-15	<8
P-cell (first coarse optimization)	6.55	41.39	0.862	13	-28	6.2
P-cell (two-step optimization)	6.6	42	0.86	13.8	-29	6

V. CONCLUSION

A novel way to build a P-cell for RF blocks is presented, implementing a multi-step approach in which the accuracy in the first step is moderate which yields high speed operation. Next steps have increasing levels of accuracy and use the result of the previous step as initial guess, which yields overall high speed and accurate operation. With the good noise and nonlinearity modeling – due to the inclusion of all nonlinear conductance and capacitances controlled by v_{gs} and v_{ds} – the high speed and good accuracy of this multi-step design automation approach was demonstrated on a narrow-band LNA design.

REFERENCES

[1] G. Tulunay, and S. Balkir, "Automatic synthesis of CMOS RF front-ends," IEEE Proc. ISCAS., pp. 4, May 2006.

[2] N. Roy, M. Najmabadi, R. Raut, and V. Devabhaktuni, "A systematic approach towards the implementation of a low-noise amplifier in sub-micron CMOS technology," Canadian conference on Electrical and computer engineering, pp. 1909-1913, May 2006

[3] A. Nieuwoudt, T. Ragheb, and Y. Massoud, "SOC-LNA: synthesis and optimization for fully integrated narrow-band CMOS low noise amplifiers," ACM/IEEE Proc. DAC, pp.879-884, July 2006

[4] http://www.nxp.com/Philips_Models/ mos_models/index.html.

[5] *Spectre Circuit Simulator User Guide*, Cadence Product Documentation.

[6] T.Lee, *The design of CMOS radio-frequency integrated circuits*, 2nd ed., Cambridge: Cambridge university Press, 2004.

[7] H. Nejati, T.Ragheb, A.Nieuwoudt, and Y.Massoud, "Modeling and Design of Ultrawideband Low Noise Amplifiers with Generalized Impedance Matching Networks," IEEE Proc. ISCAS., 4, May 2007.

[8] R.Baki, T.K.K.Tsang, and M.N.El-Gamal, "Distortion in rf CMOS short-channel low-noise amplifiers," IEEE Trans. Microwave Theory and Techniques, vol. 54, Issue 1, pp. 46-56, jan. 2006

[9] G.Palumbo, and S.Pennisi, "High-frequency harmonic distortion in feedback amplifiers: analysis and applications, " IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications, Volume 50, Issue 3, Mar 2003 pp.:328 – 340.

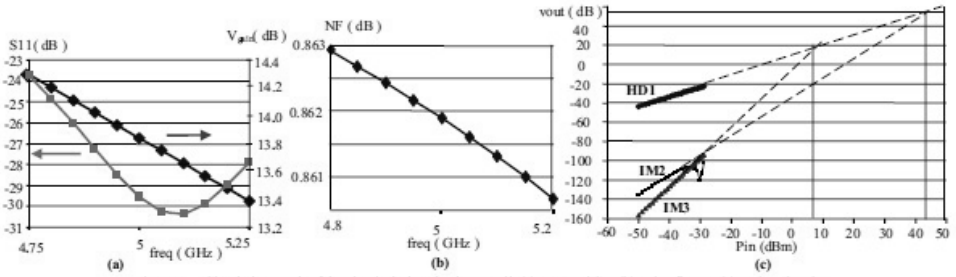


Figure. 7 Simulation result of the circuit designed using P-cell (a) V_{gain} and S_{11}, (b) noise figure , (c) IIP3 and IIP2.

Supplementary material 2

for Chapter 2, published in *IEEE Custom Integrated Circuits Conference*, pp.221 – 224, Sept. 2008.

IEEE 2008 Custom Intergrated Circuits Conference (CICC)

A general weak nonlinearity model for LNAs

Wei Cheng, Anne Johan Annema, Jeroen A. Croon*, Dirk B.M. Klaassen* and Bram Nauta
IC Design Group, CTIT Research Institute, University of Twente
PO Box 217, 7500 AE Enschede, The Netherlands
e-mail: w.cheng@utwente.nl
*with NXP-TSMC Research Center, High Tech Campus 37, Eindhoven, The Netherlands

Abstract- **This paper presents a general weak nonlinearity model that can be used to model, analyze and describe the distortion behavior of various low noise amplifier topologies in both narrowband and wideband applications. Represented by compact closed-form expressions our model can be easily utilized by both circuit designers and LNA design automation algorithms. Simulations for three LNA topologies at different operating conditions show that the model describes IM components with an error lower than 0.1% and a one order of magnitude faster response time. The model also indicates that for narrowband IM2@w_1-w_2 all the nonlinear capacitances can be neglected while for narrowband IM3 the nonlinear capacitances at the drain terminal can be neglected.**

I. Introduction

The low noise amplifier (LNA) is a critical building block in the RF front-end. One of the important design specifications of the LNA is its distortion performance, typically specified in terms of IIP2 and IIP3. A number of papers present nonlinearity analyses for LNAs to provide design guidelines [1-2] or for LNA design automation purposes [3-4]. Volterra series theory is widely used as the major nonlinearity analysis approach [5]. Trying to avoid the complex calculation involved in Volterra series, other approaches include:

A. Per-nonlinearity distortion analysis [6]

B. Combined multisine and Volterra analysis [7]

C. Harmonic distortion analysis in feedback amplifiers [8]

Approach *A* treats a MOS transistor as a linear component with a nonlinear drain current source, which allows identifying the transistors that contribute most to the output nonlinearity. Although all the drain-source nonlinearities can be included, no information is given about which nonlinearity of the drain current has more effects [7]. Approach *B* splits the nonlinear behavior in similar contributions as in approach *A* while better insights on the nonlinearity contribution are achieved by using the selective Volterra analysis. Both approach *A* and *B* demand distortion simulations of the circuit to provide the data for post-processing, which is not meant for hand-calculation analysis. Alternatively approach *C* only uses conventional algebra to analyze harmonic distortion of feedback amplifiers but can't provide solutions for intercept 1-dB compression point and intermodulation distortions. Despite the complexity difference in all aforementioned methods the nonlinearity analysis must be redone for each new topology.

In this paper we introduce a general weak nonlinearity model that is independent of the LNA topologies. Theoretically, for any LNA topology both in narrowband or wideband applications, this model calculates output IM2 and IM3 of the circuit using simple closed-form expressions. The compact closed-form expression is just a linear combination of nonlinear coefficients of each MOS transistor and terminal AC gains. Therefore, the result of our model can be easily utilized by both circuit designers and LNA design automation algorithms without involving any complex nonlinearity analysis. In section II the weak nonlinearity model for MOS transistor valid from DC to RF frequencies is introduced. Using this MOS nonlinearity model a generalized distortion analysis for weakly nonlinear circuits is discussed in Section III, which presents the approach to obtain the general weak nonlinearity model. Section IV shows the benchmark on accuracy for our model using three different LNAs operating in different load condition and at different frequencies. Conclusions are drawn in section V.

II. MOS Transistor Nonlinearity Modeling

In this paper it is assumed that MOS transistors' nonlinearities are the (main) cause for distortions in RF-circuits. Therefore, for analyzing the nonlinear behavior of RF circuits, modeling and describing transistor nonlinearity is essential. Taylor series are successfully and dominantly used to describe the weakly nonlinear behavior of the MOS transistor [1-5]. However, most of these descriptions simplify the MOS nonlinearity model to the following extent:

- Only consider transconductance nonlinearity [1-3].
- Consider all resistive drain current nonlinearity but neglect the charge-storage nonlinearity [5], [7- 8].

Aiming for validity from DC to the RF frequency range, we introduce a complete weakly nonlinear model taking into account both resistive and charge-storage nonlinearity.

A MOS transistor is a four-terminal device, in which all currents into and charges at nodes are nonlinear functions of the voltages across any two terminals. Mathematically the transistor can be modeled as a three-port network with the gate-source, drain-source and bulk-source voltage as the inputs and gate current, drain current and bulk current as outputs for any given DC bias. Therefore the transistor's weakly nonlinear behavior in the close vicinity of any DC bias point can be described by the multi-dimensional Taylor series up to (here) the third-order, which is given by

$$\begin{aligned} i_k(t) = {} & G^k_{100}v_{gs} + G^k_{200}v^2_{gs} + G^k_{300}v^3_{gs} + G^k_{010}v_{ds} + G^k_{020}v^2_{ds} + G^k_{030}v^3_{ds} + G^k_{001}v_{bs} \\ & + G^k_{002}v^2_{bs} + G^k_{003}v^3_{bs} + G^k_{110}v_{gs}v_{ds} + G^k_{120}v_{gs}v^2_{ds} + G^k_{210}v^2_{gs}v_{ds} + G^k_{101}v_{gs}v_{bs} \\ & + G^k_{102}v_{gs}v^2_{bs} + G^k_{201}v^2_{gs}v_{bs} + G^k_{011}v_{ds}v_{bs} + G^k_{012}v_{ds}v^2_{bs} + G^k_{021}v^2_{ds}v_{bs} + G^k_{111}v_{gs}v_{ds}v_{bs} \\ & + C^k_{100}\frac{dv_{gs}}{dt} + C^k_{200}\frac{dv^2_{gs}}{dt} + C^k_{300}\frac{dv^3_{gs}}{dt} + C^k_{010}\frac{dv_{ds}}{dt} + C^k_{020}\frac{dv^2_{ds}}{dt} + C^k_{030}\frac{dv^3_{ds}}{dt} + C^k_{001}\frac{dv_{bs}}{dt} \end{aligned} \quad (1)$$

$$+C^k_{002}\frac{dv^2_{bs}}{dt}+C^k_{003}\frac{dv^3_{bs}}{dt}+C^k_{110}\frac{dv_{gs}v_{ds}}{dt}+C^k_{120}\frac{dv_{gs}v^2_{ds}}{dt}+C^k_{210}\frac{dv^2_{gs}v_{ds}}{dt}+C^k_{101}\frac{dv_{gs}v_{bs}}{dt}$$
$$+C^k_{102}\frac{dv_{gs}v^2_{bs}}{dt}+C^k_{201}\frac{dv^2_{gs}v_{bs}}{dt}+C^k_{011}\frac{dv_{ds}v_{bs}}{dt}+C^k_{012}\frac{dv_{ds}v^2_{bs}}{dt}+C^k_{021}\frac{dv^2_{ds}v_{bs}}{dt}+C^k_{111}\frac{dv_{gs}v_{ds}v_{bs}}{dt}$$

where

$$C^k_{mnl}=\frac{1}{m!}\frac{1}{n!}\frac{1}{l!}\frac{\partial^{(m+n+l)}Q_k}{\partial V^m_{gs}\partial V^n_{ds}\partial V^l_{bs}}\quad ;G^k_{mnl}=\frac{1}{m!}\frac{1}{n!}\frac{1}{l!}\frac{\partial^{(m+n+l)}I_k}{\partial V^m_{gs}\partial V^n_{ds}\partial V^l_{bs}}\quad k\in\{g,d,b\}$$

are respectively the nonlinear capacitive and resistive coefficients with Q_k for charge at and I_k for current into terminal k. For simplicity reasons, in this paper the source and bulk are assumed to be connected, effectively reducing the MOS-transistor to a three-terminal device. As a result only the currents into and charges stored at the gate, drain and source are of concern and hence coefficients C^k_{mnl} and G^k_{mnl} with $l\neq 0$ are zero. Furthermore, the DC gate current is neglected and only the capacitive gate current is taken into account, which is a valid simplification for RF operation [9].

The weak nonlinearity model is shown in Fig. 1a, where the linear current sources ($i_{g,lin}$ and $i_{d,lin}$) and nonlinear current sources ($i_{g,nonlin}$ and $i_{d,nonlin}$) are separated for the circuit distortion analysis discussed in next section. Other capacitances that may be present in the MOS transistor structure can be included in this representation. For example the interconnect capacitance between the gate and drain may be added explicitly across the terminals, but can also be embedded in (1) by using e.g. the Blakesley transform.

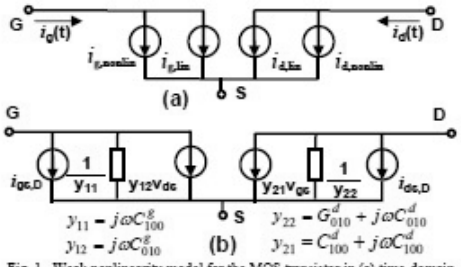


Fig. 1. Weak nonlinearity model for the MOS transistor in (a) time-domain and (b) frequency domain.

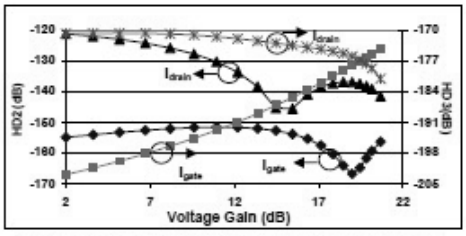


Fig. 2. Comparison on HD2@2GHz and HD3@3GHz in gate and drain current between ADS [13] simulation and the MOS nonlinearity model for different voltage gains (v_{ds}/v_{gs}). Symbols represent model prediction, lines transistor-level circuit simulation using a commercial 90nm CMOS process (f_T=110GHz)

The nonlinear coefficients are extracted directly from state-of-art MOS model, namely the PSP model [14], which ensures excellent accuracy. Very good agreement with the simulated HD2 and HD3 in gate and drain current is observed for transistors with different dimensions and bias, one of which is shown in Fig. 2 with the transistor under different voltage gain (v_{ds}/v_{gs}) conditions.

III. Generalized Distortion Analysis

In this section a general circuit with N transistors is analyzed. A two-tone input signal $V_{IN}e^{j\omega_1 t}+V_{IN}e^{j\omega_2 t}$ is applied assuming that amplitude V_{IN} is small to ensure the circuit operates in the weakly nonlinear region. The MOS transistors are assumed to be the only nonlinearity sources in the circuit, although this assumption is not necessary. Since no topology information is involved the analysis result is valid for all topologies.

A. Dependent relation

The frequency-domain MOS nonlinearity model shown in Fig.1b is used. It consists of first-order y-parameters (directly converted from the linear coefficients in (1)) and two distortion current sources (gate distortion current source $i_{gs,D}$ and drain distortion current source $i_{ds,D}$) that contain both harmonic and intermodulation distortion elements. In the frequency domain $i_{gs,D}$ and $i_{ds,D}$ can be regarded as (dependent) small-signal multi-tone stimuli, therefore in any circuit with N transistors the distortion in the circuit output is a linear combination of $i_{gs,D}$ and $i_{ds,D}$ from each transistor. Moreover, the terminal voltages of each transistor (v_{gs} and v_{ds}) are linear combination of $i_{gs,D}$, $i_{ds,D}$ from each transistor and the two-tone input signal, which in total yield

$$v_{out,D}=\sum_{k=1}^{N}(H_{igsk}\circ i_{gsk,D}+H_{idsk}\circ i_{dsk,D}) \tag{2}$$

$$v_{dsj}=\sum_{k=1}^{N}(F^{dsj}_{igsk}\circ i_{gsk,D}+F^{dsj}_{idsk}\circ i_{dsk,D}+F^{dsj}_{vds}(\omega_1)\cdot V_{IN}e^{j\omega_1 t}+F^{dsj}_{vds}(\omega_2)\cdot V_{IN}e^{j\omega_2 t}) \tag{3}$$

$$v_{gsj}=\sum_{k=1}^{N}(F^{gsj}_{igsk}\circ i_{gsk,D}+F^{gsj}_{idsk}\circ i_{dsk,D}+F^{gsj}_{vds}(\omega_1)\cdot V_{IN}e^{j\omega_1 t}+F^{gsj}_{vds}(\omega_2)\cdot V_{IN}e^{j\omega_2 t}) \tag{4}$$

where H_{ixsk} $x\in\{g,d\}$ is the AC gain to the output voltage when an AC current source applied between the terminal x and source in transistor k; F^{yxj}_{ixsk} $x\in\{g,d\}$, $y\in\{g,d\}$, $x\neq y$ is the voltage gain from the current source attached between the terminal x and source in transistor k to the terminal voltage v_{ys} in transistor j; F^{yxj}_{vds} $y\in\{g,d\}$ is the AC gain from voltage input to the terminal voltage v_{ys} in transistor j.

In summary (2) suggests that the output voltage distortion can be easily calculated once all the distortion current sources are known. (1), (3) and (4) indicate a dependent relation between the controlling voltage v_{gs} and v_{ds} and the distortion current sources of each transistor as illustrated in Fig. 3.

B. Solving for the general model

The dependent relation shown in Fig 3 generates a linear

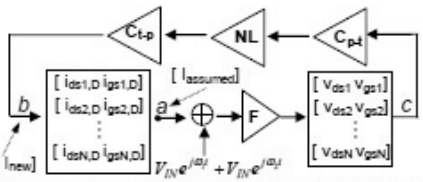


Fig. 3. Illustration of the dependent relation between the distortion current sources and the controlling voltages in any circuit with *N* transistors. Block "F" and "NL" represent the calculation conducted in (3)-(4) and (1) respectively. Block "C_{p-t}" represents the conversion from phasors to the according time-domain expressions and "C_{t-p}" for vice-versa.

equation matrix for solving the closed-form expressions of each distortion current sources ($i_{gsk,D}$, $i_{dsk,D}$, k=1,⋯ N). At first the assumed expression matrix [$I_{assumed}$] for $i_{gsk,D}$ and $i_{dsk,D}$ are input to node *a*. Given the weak nonlinear assumption, [$I_{assumed}$] contains distortion elements up to the third-order, where the magnitude of second-order distortions should be proportional to V_{IN}^2 and the magnitude of the third-order distortions to V_{IN}^3, any terms with higher-order of magnitude is small enough to be neglected. Following the loop counterclockwise until node *b* we can obtain a new expression matrix [I_{new}]. Up to the third-order [$I_{assumed}$] should be equal to [I_{new}], which provides the solution for the expressions of $i_{gsk,D}$, $i_{dsk,D}$. Then the circuit output distortion can be easily obtained from (2).

Meanwhile the matrix solving is significantly simplified due to the observation in (3) and (4) that only the terms with fundamental tones contribute to the second-order distortion; only the terms with fundamental tones and second-order tones contribute to the third-order distortion. Therefore other terms in (3) and (4) can be neglected for the calculation of second and third order distortions.

As a result the closed-form expressions for the output harmonic and intermodulation distortions are obtained. Due to the space limitation only their compact forms are given

$$v_{out,\omega_{IM}} = \sum_{k=1}^{N} [P_{nm0,k}^{d} \cdot (G_{nm0,k}^{d} + j\omega_{IM} C_{nm0,k}^{d}) + P_{nm0,k}^{g} \cdot j\omega_{IM} C_{nm0,k}^{g}] \quad (5)$$

where $P_{nm0,k}^{j}$ is a function of according AC terminal gains; ω_{IM} is the intermodulation frequency; $G_{nm0,k}$ and $C_{nm0,k}$ are the nonlinear conductances and capacitances in transistor *k* at terminal gate/drain. We've observed that in a 90nm CMOS process up to 5 GHz $G_{nm0,k}^{d}$ is more than one order of magnitude larger than both $\omega C_{nm0,k}^{d}$ and $\omega C_{nm0,k}^{g}$ and thus for lower RF frequency (5) can be simplified to

$$v_{out,\omega_{IM}} \approx \sum_{k=1}^{N} (P_{nm0,k}^{d} \cdot G_{nm0,k}^{d} + P_{nm0,k}^{g} \cdot j\omega_{IM} C_{nm0,k}^{g}) \quad (6)$$

$$\text{or} \quad v_{out,\omega_{IM}} \approx \sum_{k=1}^{N} (P_{nm0,k}^{d} \cdot G_{nm0,k}^{d}) \quad \text{, if } P_{nm0,k}^{g} \approx P_{nm0,k}^{d} \quad (7)$$

which provides the following circuit design insights for any LNA topology in a CMOS technology:

a. For narrowband IM2 (ω_{IM}@low frequency) all the nonlinear capacitances can be neglected and only three second-order conductances of each transistor are important, namely, nonlinear transconductance (G_{200}), output conductance (G_{020}) and cross-modulation conductance (G_{110}).
b. For IM3 (ω_{IM}@ RF tone) the nonlinear capacitances at the drain terminal can be neglected and if $P_{nm0,k}^{g}$ is not much larger than $P_{nm0,k}^{d}$ the nonlinear capacitances at the gate terminal can also be neglected.

In summary this generalized distortion analysis shares an approach that has similarities with harmonic balance analysis in approximating all node voltages and branch currents by truncated Fourier series. By utilizing the diagram shown in Fig. 3 and cutting redundant frequency elements our analysis method simplifies complex weakly nonlinear analyses into relatively simple calculations. The resulting general weakly nonlinear model can be applied for different LNA topologies as will be shown next. The model uses simple closed-form expressions to describe the total output distortion for both narrowband and wideband application, which is just linear combination of the nonlinear coefficients of each transistor and terminal AC gains. In fact this general model can also be applied to [10] where a state-space approach is used to get low complexity analytic expressions of distortions for different fully balanced band Gm-C filters

IV. Benchmark On Accuracy

In order to evaluate the accuracy of the proposed general nonlinearity model, a common source (CS) amplifier, a narrowband cascode LNA [11] and a broadband noise-canceling LNA [12] are simulated in ADS using a commercial 90nm CMOS process. The topologies are shown in Fig. 4. The general LNA nonlinearity model is integrated in parametric cells (P-cells) [4] to provide distortion prediction.

The CS amplifier is an example verifying the two design insights given in the previous section. Fig. 5 shows that for narrowband IM2 three nonlinear conductances are the key factors, which verifies the design insight *a*. For the CS amplifier $P_{nm0,k}^{g}/P_{nm0,k}^{d} \approx g_m(Z_{Cgs}//R_s)$ is typically around 1~2 and thus insight *b* is applied ,which is verified in Fig. 6 that all nonlinear capacitances at gate and drain terminals can be neglected for narrowband IM3. For the other two LNAs the 0simulated results by ADS and the results of our model are compared on output IM2 and IM3. The accuracy is defined by $\left|1 - \left|\frac{V_{ADS} - V_{model}}{V_{ADS}}\right|\right|$, where V_{ADS} and V_{model} are the ADS-simulated and model results for voltage magnitude of IM respectively. For the narrowband LNA the two-tone signal is at 5 GHz with 1 MHz spacing. By sweeping the load ($Z_{L,real}$) the model is benchmarked for different gains as shown in Fig. 7. For the wideband LNA one tone is fixed at 1 GHz and the other tone is swept from 2.15 GHz to 10 GHz resulting in a perfect description of IM2 and IM3 at different frequency shown in Fig. 8.

In summary the benchmark shows the proposed general nonlinearity model predicts the output distortion for different topologies with very good accuracy. The simple closed-expressions with the model enable a very fast response time for calculating IM distortions (within 40 ms per one sweep)

providing over one order of magnitude advantage versus the transistor-level simulation in ADS, thus making it very suitable for implementation in a design automation loop for optimizing the circuit within short time.

V. Conclusion

A general weak nonlinearity model for different low-noise amplifier topologies was presented, which is achieved by our generalized weak nonlinearity analysis. Implemented by simple closed-form expressions this model provides a potential solution for LNA design automation with different topology candidates to improve the response time and for the circuit designers to gain insightful guidelines on LNA nonlinearity. Very good accuracy and short response time of this model is shown on intermodulation distortion calculation for different LNAs in both narrowband and wideband applications.

References

[1] R.A.Baki, T.K.K. Tsang and M.N.El-Gamal, "Distortion in RF CMOS Short-Channel Low-Noise Amplifiers, *IEEE Trans. Microw. Theory Tech.*, VOL. 54, NO.1, pp. 46-56, Jan. 2006.

[2] W.Chen, G.liu, B.Zdravko and A.M.Niknejad, "A Highly Linear Broadband CMOS LNA Employing Noise and Distortion Cancellation," in Proceedings of *IEEE RFIC Symposium*, Hawaii, June 3-5, 2007.

[3] G. Tulunay, and S. Balkir, "Automatic synthesis of CMOS RF front-ends," in Proceedings of *IEEE ISCAS*, Greece, May 21-24, 2006.

[4] W.Cheng, A. J. Annema and B.Nauta , "A multi-step P-cell for LNA design automation," in Proceedings of *IEEE ISCAS*, May 2008, in press.

[5] P.Wambacq and W.Sansen, *Distortion Analysis of Analog Integrated Circuits*, Norwell, MA: Kluwer, 1998.

[6] P.Li and L.T.Pileggi, "Efficient per-nonlinearity distortion analysis for analog and RF circuits,"*IEEE Trans. CAD Des. Integr. Circuits Syst.*, vol. 22, pp. 1297-1309, 2003.

[7] J.Borremans, L.D.Locht, P.Wambacq and Y.Rolain, "Nonlinearity Analysis of Analog/RF Circuits using Combined Multisine and Volterra Analysis," in Proceedings of *IEEE DATE*, France, April 16-20, 2007.

[8] G.Palumbo and S.Pennisi, "High-frequency harmonic distortion in feedback amplifiers: Analysis and applications," *IEEE Trans. Circuit Syst. I, Reg. Papers*, vol. 50, no. 3, pp. 328-340, mar. 2003.

[9] A.J. Annema, et al., "Analog Circuits in Ultra-Deep Sub-Micron CMOS," *IEEE J. Solid-State circuits*, vol. 402, No. 2, Jan. 2005

[10] A.Celik, Z.Zhang and P.P.Sotiriadis, "A state-space approach to intermodulation distortion estimation in fully balanced bandpass Gm-C Filters with weak nonlinearities" *IEEE Trans. Circuit Syst. I, Reg. Papers*, vol. 54, Issue 4, pp. 829-844, April 2007.

[11] T.Lee, *The design of CMOS radio-frequency integrated circuits*, 2nd ed., Cambridge: Cambridge University Press, 2004.

[12] C.-F. Liao and S.-I. Liu, "A Broadband Noise-Canceling CMOS LNA for 3.1-10.6-GHz UWB Receivers", *IEEE J. Solid-State circuits*, vol. 42, Issue. 2, Feb. 2007

[13] http://eesof.tm.agilent.com/products/ads_main.html

[14] http://www.nxp.com/Philips_Models/ mos_models/index.html.

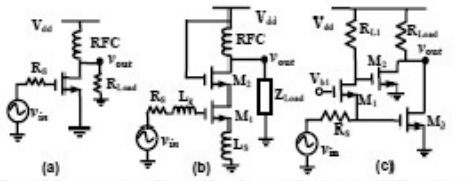


Fig. 4. Schematics of (a) CS amplifier, (b) narrowband cascode LNA and (c) wideband noise-cancelling LNA

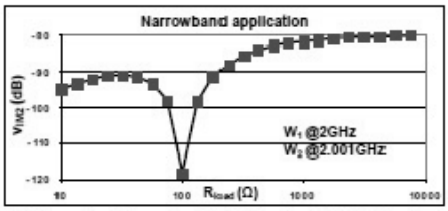


Fig. 5 Comparison between simulated (line) and model (symbol) results on output voltage IM2@1MHz for a CS amplifier as a function of load resistance. Squares represent the model (only contains three nonlinear conductances and neglects nonlinear capacitances).

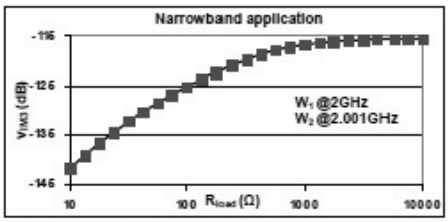


Fig. 6 Comparison between simulated (line) and model (symbol) result on output voltage IM3@1MHz for a CS amplifier as a function of load resistance. Squares represent the model (neglecting nonlinear capacitances at gate and drain terminals)

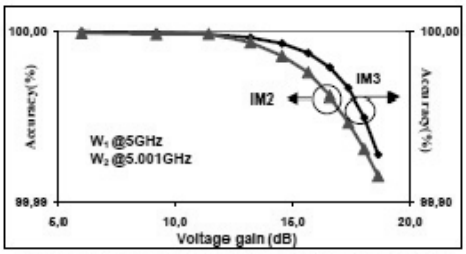


Fig. 7 Accuracy of our model on IM2 and IM3 for narrowband cascode LNAs over different voltage gain (v_{out}/v_{in})

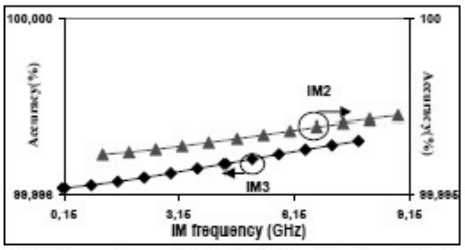


Fig. 8 Accuracy of our model on IM2 and IM3 prediction for a wideband LNA (IM2 @w_1-w_2 ranging from 1.15 GHz to 8.85GH; IM3@w_2-2w_1 ranging from 0.15GHz to7.85GHz)

List of Publications

Journal Papers

W. Cheng, M. S. Oude Alink, A. J. Annema, G. J. M. Wienk and B. Nauta, "A wideband IM3 cancellation technique for CMOS Π and T attenuators," submitted to *IEEE J. Solid-State Circuits*.

W. Cheng, A. J. Annema, G. J. M. Wienk and B. Nauta, "A flicker noise/IM3 cancellation technique for active mixers," submitted to *IEEE J. Solid-State Circuits*.

W. Cheng, M. S. Oude Alink, A. J. Annema, J. A. Croon and B. Nauta, "RF circuit linearity optimization using a general weak nonlinearity model," accepted for publication in *IEEE Trans. Circuits and Systems I*, 2012.

W. Cheng, A. J. Annema, J. A. Croon and B. Nauta, "Noise and nonlinearity modeling of active mixers for fast and accurate estimation," *IEEE Trans. Circuits and Systems I,* vol. 58, pp. 276-289, Feb. 2011.

Conference Papers

W. Cheng, A. J. Annema and B. Nauta, "A multi-step P-cell for LNA design automation," *IEEE Int. Symp. Circuits Syst.(ISCAS),* pp.2550-2553, May 2008.

W. Cheng, A. J. Annema, J. A. Croon, D. B. M. Klaasen and B. Nauta, "A general weak nonlinearity model for LNAs," *IEEE Custom Integrated Circuits Conference (CICC)*, pp.221 – 224, Sept. 2008.

W. Cheng, M. S. Oude Alink, A. J. Annema, G. J. M. Wienk and B. Nauta, "A wideband IM3 cancellation technique for CMOS attenuators," *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, San Francisco, CA,* 2012, pp. 78 –79.

W. Cheng, A. J. Annema, G. J. M. Wienk and B. Nauta, "A wideband IM3 cancellation technique using negative impedance for LNA with cascode topology," accepted for publication in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium* 2012.

M. S. Oude Alink, E. A. M. Klumperink, A. B. J. Kokkeler, W. Cheng, Z. Ru, A. Ghaffari, G. J. M. Wienk and B. Nauta, "A CMOS spectrum analyzer frontend for cognitive radio achieving +25dBm IIP3 and -169 DBm/Hz DANL," accepted for publication in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium* 2012.

Acknowledgment

This thesis is made possible with the help of many people. My gratitude in heart shall surface in text now.

Looking back, in fact I wasn't so sure in which direction I should start my master project. Both IC design and microfabrication were of my interest. I realize it's my promotor, Bram Nauta, who made the decision for me indirectly then. Instead of leaving me alone to think about the project proposal, he immediately gave me a tour in our Hoegkamp vloer3. His passionate talk reinforced my preference to the IC design and resulted in this thesis. Not only Bram is good at igniting students' interest in IC design, he is also a great person in leading ICD group to its prosperous path. I am really lucky to witness the fast and successful growth of ICD. Hopefully I have made some contribution. I am grateful for the opportunity of doing PhD in ICD.

I also would like to thank my daily supervisor Anne Johan Annema for his devoted guidance both in circuits design and tennis courts. The high standard he and Bram set for my research work and academic writing is invaluable to my future career. I thank him for having me as his PhD. He always "bugs" me when I present some results to him, which is essential in improving the quality of my work.

My thanks go to Eric Klumperink. Although I am not his student, he cares my social and research well-being. The flower parade in Lichtenvoorde is great, I thank him for his warm invitation. From time to time, he asks about my research process and always helps as much as he can. Without talking with him, I probably would never work on the attenuator and thus graduate without an ISSCC paper.

This ISSCC paper would be impossible without Mark, Gerard and the help from Michiel and Henk. Working with Mark is very pleasant since he is a person of giving with an open mind. I appreciate his willingness to share knowledge with me and the whole group. This is the spirit that should last forever in ICD. I thank Gerard for proposing making the attenuator in C14, you are the very key factor in this matter. I thank him also for making the

layouts for me. Otherwise, I might graduate even later. It's also great fun to look at his photos every time after he comes back from vacations. The measurement wouldn't finish in time without the help from Michiel and Henk. I am grateful for that.

I would like to express my gratitude to all the ICDer of "last generation". I deeply appreciate the help and understanding from Mustafa, your tips on the master project presentation is extremely important. I know not all students can get such good help from their supervisors. Paul Veldhorst is of not much word but much kindness, I thank him for the help. Paul Geraedts did great job in the initiation of ICD wiki, we can't thank him enough for that. I am glad to be the colleague of Fabian, who shows that one man can be talented in two extremely different fields. Stephan, Eisse, Simon, Daniel, Rameswor, Vincent, talking with you guys are always enjoyable. Thanks to Amir, I am now not sensitive to the "D"-word. My special thanks to two friends, Fenno de Vries and Paulo Lookman. During my master student time, many nights at the student lab working and talking with you guys is great fun. I will always remember.

It's my pleasure to stay in such a quiet room with all the Japanese friends from AKM (Asahi Kasei microdevices). I thank tennis master Osamu for teaching me how to play better; I thank Kazuma for his dinner; I enjoyed talking with Kenji, Shuichi, Yuichiro and Eiji about Japan, China, travelling and many other topics.

My thanks go to NXP semiconductor for funding this project and donating the silicon. Dirk Klaassen, Bert Huizing, Raf Roovers and Gerard van de Weide always show us a warm welcome during our visit there and provide great help to me. Jeroen Croon has done enormous work in providing helpful feedback, which boosts the productivity of my research work a great deal.

Gerdien is the first person I have found to share the same birthday with me. She is very considerate and kind. Talking with her is very pleasant. Annemiek takes special care of the foreign students here, we all love her. Two lady secretaries in our floor always help us keep a good balance of work and life. Thank you.

Tennis is the second thing I've learned in the Netherlands besides IC design. A one-hour break at lunch time in the courts really refreshes my mind.

Karin, Rene, Eric, Peter, Rik, Hans, Kodo and Coach Jan Herman, thank you all for playing with me. I will miss playing tennis with you.

Although the PhD life involves pressure and stress, the social life with all my friends in Enschede just calms me down. Balaji and the F4 of SC group (Victor, Joost, Bijoy and Jiwu, a.k.a Uncle Lu), I wish I could've known you well earlier. Having dinner/drink, swimming and running Singelloop Enschede with you guys always are fun. Playing tennis double late in summer nights is more joyful. Well, of course, watching Uncle Lu's weight-losing campaign once in a while is the most cheerful event. Balaji, it would be perfect if you could play tennis. Staying in Europe exposes me to a great variety of culture and nature. Chang Zheng, Xiaotong, Lala, Yang Jing, Tian Jia, Xu weiwei, Shuhan, Xiaoying, Gao Xiang, Xiaoyan, it's so great to have your accompany during our happy trips. The memory always remains in my mind. Gang Tian, Wan Xin, Song Jing and Kang Shuo, we really have made a lucky decision to go to Egypt one year before the revolution. Playing Majiang (Mahjong) in the train was lots of fun. Liu Hanli&Sun Fei, Ma Yujie&Chen Qi, Tao Qian&Zhu Yi, I am very glad to see your family population increases. It's also good for me to gain some babysitting skills from your kids. Many thanks go my flatmates, Cai Lei, Zhao Xiaoguang, Wang Xiaocheng, He Yu, Yin Haihan and Chen Yanting. Because of you, our flat feels like home. My friends, Lin Yi, Zhang Peng, Zhao Yiping, Zhang Yang, Yang Di, Xie Yanbo, Ma Haifeng, Wang Wenxin, Qin Yunnan, Wang Jiahui, thank you all for your support and friendship. We may get busier when we grow older, but we shall not care less for each other. My friend and flatmate, Zhiyu, always brings positive energy when I got frustrated at work. Discussing work with you while cooking in the kitchen is very "tasteful".

I owe my parents the deepest gratitude for their caring love and unconditional support. All I have achieved today is because of them. My wife Julie, thank you for taking care of my parents, you can make them happier than me. Your love and encouragement emerge in this thesis.

Enschede, April 2012.