Sensitivity Analysis and Distortion Decomposition of Mildly Nonlinear Circuits

by

Guoji Zhu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2007

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Volterra Series (VS) is often used in the analysis of mildly nonlinear circuits. In this approach, nonlinear circuit analysis is converted into the analysis of a series of linear circuits. The main benefit of this approach is that linear circuit analysis is well established and direct frequency domain analysis of a nonlinear circuit becomes possible.

Sensitivity analysis is useful in comparing the quality of two designs and the evaluation of gradient, Jacobian or Hessian matrices, in analog Computer Aided Design. This thesis presents, for the first time, the sensitivity analysis of mildly nonlinear circuits in the frequency domain as an extension of the VS approach. To overcome efficiency limitation due to multiple mixing effects, Nonlinear Transfer Matrix (NTM) is introduced. It is the first explicit representation of the complicated multiple mixing effects. The application of NTM in sensitivity analysis is capable of two orders of magnitude speedup.

Per-element distortion decomposition determines the contribution towards the total distortion from an individual nonlinearity. It is useful in design optimization, symbolic simplification and nonlinear model reduction. In this thesis, a numerical distortion decomposition technique is introduced which combines the insight of traditional symbolic analysis with the numerical advantages of SPICE like simulators. The use of NTM leads to an efficient implementation. The proposed method greatly extends the size of the circuit and the complexity of the transistor model that can be handled. For example, industry standard compact model, such as BSIM3V3 [35] was used for the first time in distortion analysis. The decomposition can be achieved at device, transistor and block level, all with device level accuracy.

The theories have been implemented in a computer program and validated on examples. The proposed methods will leverage the performance of present VS based distortion analysis to the next level.

Acknowledgements

This work would have been impossible without the inputs from many people.

A few words cannot express all my gratitude to my supervisor, Professor Ajoy Opal. I would like to thank him for introducing me to this exciting field, for his excellent guidance and valuable comments throughout this research, and for his financial support for my graduate studies at Waterloo.

I am grateful to Professor James Barby, who not only gave me good advice, but also spent so much time and energy in reading my thesis and helping me to refine it. Thanks to the interesting class ECE644 he offered in winter 2006, I built my background in computer aided design of circuits, which made my research much easier.

I would like to thank all my friends who have provided me with joy and warmth throughout these years, especially Raymond Zhu, who is like my older brother and introduced me to the area of entrepreneurship.

Finally, I would like to thank my parents for their continued love, support and understanding. I love you too.

Contents

List of Tables

List of Figures

List of Abbreviations

AC   All Contributions

CAD   Computer Aided Design

DR   Dynamic Range

EDA   Electronic Design Automation

FD   Finite Difference

HD_3 Third Order Harmonics

HF   High Frequency

IM_3 Third Order Intermodulation

MIC   Most Important Contributor

MNA   Modified Nodal Analysis

NTM   Nonlinear Transfer Matrix

Op-Amp   Operational Amplifier

PA   Power Amplifier

PI   Polynomial Interpolation

RF   Radio Frequency

SBG   Simplifications Before Generation

SDG   Simplifications During Generation

VS   Volterra Series

X_n variable in the n-th order Volterra circuit

h_j j-th order nonlinear coefficient

$F(\cdot)$ Fourier transform

$\otimes$ Convolution operation

Chapter 1

Introduction

1.1 Volterra Series Analysis of Weakly Nonlinear Circuits

Generally, it is more difficult to analyze a nonlinear circuit as compared to a linear one. A Taylor Series expansion can be used to simplify the distortion analysis, but it is only feasible for small circuits at low frequencies, with only a few nonlinearities. Traditional approaches for nonlinear circuit simulation are time domain analysis followed by a Fourier transform (classical SPICE approach) [18], Harmonic Balance [31,40] and Shooting methods [32, 33]. These methods compute the total distortion response iteratively. They do not indicate which nonlinearity in the circuit is mainly responsible for the observed nonlinear behavior. Hence such methods are suitable for verification of already designed circuits. If simulation results fail to meet the pre-determined specifications, they cannot provide insight for further improvement. To meet such requirements, Volterra Series (VS) analysis is commonly used. VS approach is capable of analyzing weakly nonlinear circuits in the frequency domain. It plays an important role in the field of distortion analysis.

In the VS approach, each nonlinearity in the original circuit is first substituted with the corresponding Taylor series approximation of the i-th order. Based on the equivalent circuit, the distortion response can be computed by recursively solving the same linear circuit i times with

corresponding equivalent sources. The VS method can be used to separate different distortion contributions exactly in the same way engineers are accustomed to do in noise analysis: the dominant ones can be listed so that designers can target them first.

Dynamic range (DR) is an important figure-of-merit in system level design. It is defined as the ratio between the maximum and minimum detectable signals while maintaining a prescribed performance quality. The upper floor of DR is determined by distortion and the lower floor by noise. It can be maximized either by highly linear design or by reducing the noise floor. Notice that many figure-of-merits, including noise figure and gain, are first order analysis results. Thus, they are essentially a byproduct of distortion analysis.

Finally, since VS analysis is performed entirely in the frequency domain, there is no restriction on the input signal. This makes VS the ideal method for multi-tone distortion analysis, e.g. Ultra Wide Band systems.

However, VS analysis has some limitations: firstly, since a Taylor series expansion is used around a bias point, it is accurate only for small variations around the bias point. In practice, the polynomial approximations are limited to low degrees, e.g. analysis is typical up to the third order. Higher order analysis, e.g. 5th order analysis, is used to study the dependency of IM_3 on the signal amplitude in Power Amplifiers (PA) [21]. This is due to (1), the computational cost grows rapidly as the degree of analysis increases; (2), even if numerically efficient simulators are available, the present compact models are not accurate for high frequency distortion analysis. Lower order VS analysis inevitably leads to truncation error, especially in circuits with large signal excitation, e.g. PA.

VS analysis is based on nonlinear coefficients obtained from a Taylor Series expansion, while circuit designers are accustomed to design circuits in terms of transistor sizing and biasing. To make the matters worse, complicated transistor model makes the relationship between design parameters and the nonlinear coefficients complicated and obscure. Thus, VS is commonly used

for simple circuits with a couple of transistors (e.g. PA or low noise amplifier) or moderate scale analog circuits including 10 to 20 transistors but with simplified models (e.g. op-amps [9]).

In summary, VS analysis is used because of its efficient distortion analysis, it has been implemented in simulators such as SPICE3 [18], HSPICE [16], Voltaire XL [17] and a 5th order electro-thermal multi-tone simulator for PA study [19].

In this thesis, methods to analyze weakly nonlinear circuits are given. A system is defined to be weakly nonlinear if it can be accurately represented by a Volterra series with a small number of terms. Many analog circuits are weakly nonlinear, and they constitute the majority of continuous time analog integrated building blocks, including active filters, RF front ends for telecom systems, analog circuits in audio applications, Op-Amps, etc. In a weakly nonlinear circuit, if sinusoidal signals at different frequencies are applied, the output contains not only the input frequencies and their harmonics but also the linear combinations of the input frequencies called intermodulation components.

1.2   Sensitivity Analysis of Weakly Nonlinear Circuit

In addition to distortion analysis, sensitivity analysis of nonlinear circuits is required. It is used to compare the quality of two designs and in the evaluation of the gradient, Jacobian or Hessian matrices. In currently available RF circuit simulators, sensitivity of distortion can only be approximated by finite difference (FD). If p is the dimension of the design parameter vector, $2p$ additional simulations are required to estimate the two-sided FD. Sensitivity approximation by FD requires significant computational cost and is prone to numerical errors.

In Chapter 3, we present the VS based sensitivity analysis of mildly nonlinear circuits in the

frequency domain, as an extension of the frequency analysis of linear circuits. The application of the Adjoint method improves the numerical efficiency of sensitivity calculations. However, because of multiple mixing effects, the sensitivity calculation may be expensive. To improve efficiency, the Nonlinear Transfer Matrix (NTM) is introduced. Experiments show it is capable of up to two orders of magnitude speedup for the analysis of large analog circuits.

1.3   Per-element Distortion Decomposition

Per-element distortion decomposition determines the contribution to total distortion from an individual nonlinearity, e.g. a single nonlinear coefficient, a group of nonlinear coefficients (e.g. a transistor) or a group of transistors (e.g. an analog building block in a large circuit). In this manner, it is different from the existing VS based simulators, mentioned in Section 1.1[16-19], which determine the overall distortion response only. Per-element distortion decomposition can be used in design optimization, symbolic analysis and nonlinear model reduction.

1.3.1 Previous Approaches

Currently available distortion decomposition methods can be classified into three categories. The first is call "Brute Force" method. In this method, the difference between two distortion analyses is calculated. The first analysis is done by setting the nonlinearity of interest to zero and the second by introducing the nonlinearity in the circuit. The "Brute Force" method is not used due to numerical error and computation cost.

The second method is most intuitive and widely used [8]. First, an analytical expression of

distortion response is calculated based on VS, either by hand or symbolic analysis [34]. Based on the full expression, the contribution from each nonlinearity can then be identified.

The third method is based on PI (polynomial interpolation) [3]. It has been applied in symbolic analysis and component-level behavioral model reduction [4, 9]. The procedure is, firstly, the contribution from each nonlinearity is calculated as a function of frequency s, while circuit elements (small signal and nonlinear coefficients) take on numerical values. The result consists of a product of polynomials in s. Decomposition can thus be achieved by weighting the contribution from each nonlinearity to the total response, both represented as a function of frequency.

PI based methods seem to be the best choice for distortion decomposition so far. However, they are limited by two reasons. Firstly, although partly numerical (circuit elements take numerical values), it is still based on the symbolic representation of frequency s. Secondly, when calculating the contribution to higher order distortion response from lower order nonlinear coefficients, numerical errors in the evaluation of the polynomial functions might not be negligible. In other words, PI based methods are limited because they are partly symbolic.

Generally, symbolic analysis of distortion is complicated and expensive. This limits the size of the circuit and the complexity of the transistor model the above approaches can handle. Simplifications based on previous knowledge (experience with low frequency small signal analysis) are commonly performed [4, 8]. In addition, they are inadequate for multi-tone test [30] or decomposition of higher order distortion response [19]. Decomposition of total distortion at transistor level exists [15]. However, sometimes detailed distortion information at device level is preferred, e.g. RF circuits which contain only a few transistors.

1.3.2 Proposed Method

In Chapter 4, a numerical distortion decomposition technique is introduced based on NTM. It starts

from the original full Spice level equivalent circuit. Because of its numerical efficiency, complex compact models can be used, it is the first distortion analysis tool that uses industry standard compact models, e.g. BSIM3V3 [35]. It is shown to outperform previous approaches in efficiency, accuracy and easy-of-use. We believe the proposed method will leverage the performance of present VS based distortion analysis to the next level.

1.4   Thesis Organization

Chapter 2 reviews Schetzen's nonlinear network analysis method. Next, Schetzen's method is extended to multi-dimensional nonlinearity in order to handle semiconductor device models. **Chapter 3** introduces the sensitivity analysis of mildly nonlinear circuits. Based on VS, for computational efficiency, it uses the Adjoint method of sensitivity calculation. Next, analysis shows that sensitivity analysis can be computationally expensive in some situations due to the complicated multiple mixing effects. To improve efficiency, the Nonlinear Transfer Matrix (NTM) is introduced. For the first time, the complicated multiple mixing effects can be explicitly expressed in analytical form. Numerical examples show that its application in sensitivity calculation is capable of up to a two order of magnitude speedup. **Chapter 4** proposes an advanced distortion decomposition technique based on NTM. It combines the insight of traditional symbolic analysis, and the numerical advantages as well as generality of SPICE like simulators. The decomposition can be achieved at device, transistor and block level, all with device level accuracy. Numerical examples illustrate its applications in design optimization, symbolic simplification and nonlinear model reduction. **Chapter 5** first talks about the applications of the above methods. Then it briefly discusses the extension of present algorithms to a special type of strongly nonlinearity, e.g. mixers. Finally, an interesting question is raised: can numerical simulator provide analytical insight like manual and symbolic analysis?

Chapter 2

Schetzen's Frequency Domain Nonlinear Network Analysis

In this chapter Schetzen's nonlinear network analysis procedure based on n-linear operators and power series expansion is reviewed. Next, the numerical calculation of a nonlinear system's frequency domain response using MNA formulation is introduced. Following that, the frequency domain response of a simple RC circuit is calculated, both by hand and using the proposed method. The agreement of the results from both methods verifies the validity of the method. In Section 2.4, Schetzen's nonlinear network analysis method is extended to multi-dimensional nonlinearity in order to handle semiconductor device models, as an example, a MOS transistor model is developed. Finally, the frequency domain response of a Cascode amplifier is calculated.

2.1 Schetzen's Nonlinear Network Analysis Method

In this section Schetzen's method [1] for the analysis of mildly nonlinear circuits is reviewed. The

method is particularly useful in the analysis of nonlinear circuits, since it gives physical insight into the effect of the nonlinear terms and avoids multiple integrals normally used in the original Volterra analysis [2]. However, it should be mentioned that Schetzen's method is based on Volterra analysis and gives the same results. The first step in Schetzen's method is to expand the nonlinear functions representing the element characteristics into a Taylor series

$$f(x) = f(x_0) + (x-x_0)f'(x_0) + \frac{(x-x_0)^2}{2!}f''(x_0) + ...,$$

where x is the independent variable, $f(\cdot)$ the dependent variable, the primes on the function represent the differential with respect to the independent variable, and $x=x_0$ is the expansion point. For electrical circuits, the independent and dependent variables are typically a voltage, current, charge or flux. For simplicity, initially only elements that depend upon one independent variable are considered. More complex elements, such as semiconductor devices, that depend on more than one independent variable are considered in Section 2.3. Consider a nonlinear capacitor, shown in Fig. 2.1(a), whose charge q is a nonlinear function of the voltage v

$$q = f(v) = f(v_0) + (v-v_0)f'(v_0) + \frac{(v-v_0)^2}{2!}f''(v_0) + + \frac{(v-v_0)^3}{3!}f'''(v_0) + ... \qquad (2.1)$$
$$= q_0 + C_1\hat{v} + C_2\hat{v}^2 + C_3\hat{v}^3 + ...$$

where $q_0 = f(v_0)$ is the quiescent charge on the capacitor, $\hat{v} = v - v_0$ is the change in voltage from the quiescent value v_0, $C_1 = f'(v_0)$ is the first order capacitance term, $C_2 = f''(v_0)/2$ is the second order capacitance coefficient, and so on. The subscript on C determines the order of approximation, thus, C_1 is the linear capacitance term and the remaining coefficients C_2, C_3 ... are the nonlinear terms. Eq. (2.1) can be represented as an equivalent circuit, shown in Fig. 2.1 (b), consisting of a charge source q_0, voltage source v_0 and a nonlinear capacitor, called the incremental nonlinear capacitance, given by $\hat{q} = \hat{f}(\hat{v})$, where
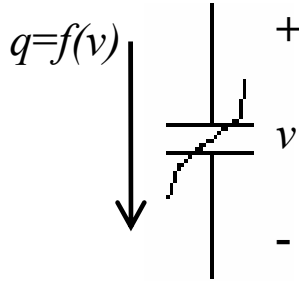
Figure 2.1(a) A Nonlinear capacitor   **Figure 2.1(b) Equivalent circuit model**
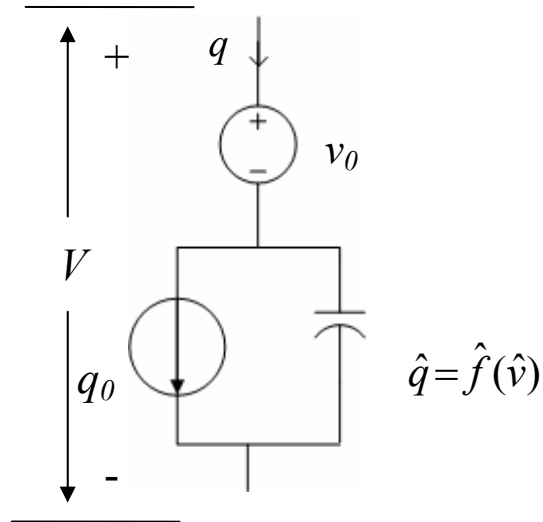
$$\hat{q} = C_1\hat{v} + C_2\hat{v}^2 + C_3\hat{v}^3 + ... = \sum_{n=1}^{\infty} C_n\hat{v}^n \qquad (2.2)$$

and $q = q_0 + \hat{q}$. It is important to manipulate the nonlinear element into the form given by Fig. 2.1(b) and Eq. (2.2), i.e. the constant sources q_0 and v_0 are removed from the representation, to keep the analysis mathematically tractable. A consequence of the series representation is that, if the independent variable v is scaled by a scalar constant α, then the n-th order term in the expansion is scaled by α^n. Eq. (2.2) also shows the mechanism for generation of mixing and harmonics, if the independent variable is a sinusoid $\hat{v} = e^{j\omega t}$, then the dependent variable will contain the harmonics $e^{jn\omega t}$. The same steps can be used for any element that is represented by a nonlinear function of one independent variable.

The next step in Schetzen's method consists of breaking up the complete response into a summation of the response of circuits of different order

$$\hat{q} = \sum_{n=1}^{\infty} \hat{q}_n \quad \text{and} \quad \hat{v} = \sum_{n=1}^{\infty} \hat{v}_n \qquad (2.3)$$

9

where n is a Volterra series order and $n = 1$ corresponds to the first order linearized circuit. Starting with (2.2) and recalling that scaling the variables with a scalar constant α will scale the n-th component by α^n,

$$\hat{q} = \sum_{n=1}^{\infty} \alpha^n \hat{q}_n = \sum_{k=1}^{\infty} C_k \left(\sum_{n=1}^{\infty} \alpha^n \hat{v}_n \right)^k .$$

is obtained. Expanding the double summation and collecting terms with common powers of α produces

$$\begin{aligned}
\sum_{n=1}^{\infty} \alpha^n \hat{q}_n = & \alpha [C_1 \hat{v}_1] \\
& + \alpha^2 [C_1 \hat{v}_2 + C_2 \hat{v}_1^{\,2}] \\
& + \alpha^3 [C_1 \hat{v}_3 + 2 C_2 \hat{v}_1 \hat{v}_2 + C_3 \hat{v}_1^{\,3}] \\
& + ...
\end{aligned} \tag{2.4}$$

By comparing co-efficients of different powers on both sides of (2.4),

$$\begin{aligned}
\hat{q}_1 &= C_1 \hat{v}_1, \\
\hat{q}_2 &= C_1 \hat{v}_2 + C_2 \hat{v}_1^{\,2}, \\
\hat{q}_3 &= C_1 \hat{v}_3 + 2 C_2 \hat{v}_1 \hat{v}_2 + C_3 \hat{v}_1^{\,3}, \\
& \quad ...
\end{aligned} \tag{2.5}$$

can be identified. The general pattern in (2.5) is of the form

$$\hat{q}_n = C_1 \hat{v}_n + g_n (\hat{v}_{n-1}, \hat{v}_{n-2}, ..., \hat{v}_1, C_n, C_{n-1}, ..., C_2) \tag{2.6}$$

where $g_n(\bullet)$ is a function of lower order responses and $g_1(\bullet) = 0$. Thus, the n-th order Volterra representation of the incremental nonlinear element in (2.2) is given by (2.6) and can be represented by the equivalent circuit shown in Fig. 2.2. Note that each Volterra circuit contains the linear capacitance term denoted by C_1, and a charge source that depends on lower order Volterra responses.

To summarize, the steps used in obtaining the Volterra representation of each nonlinear capacitor given by (2.1) and Fig. 2.1(a), is to first extract the quiescent values and the incremental
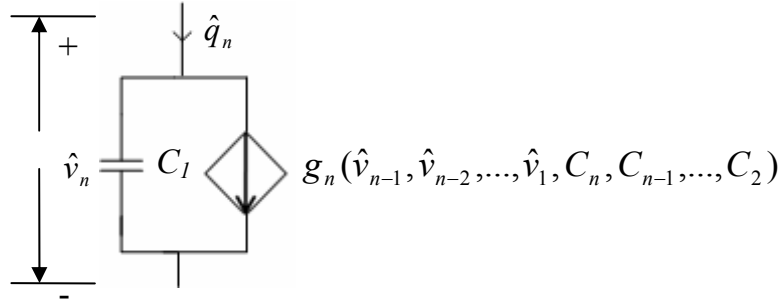
Figure.2.2   Equivalent circuit for nonlinear capacitance given by (2.2)

nonlinear capacitor given by (2.2), as shown in Fig. 2.1(b). The next step is to replace the incremental nonlinear capacitor defined by (2.2) by the linear capacitor and charge sources, as given in Fig. 2.2. In the first order Volterra circuit the quiescent sources appear, along with the linear capacitance term, and there are no additional sources related to the nonlinear element. In Volterra circuits of orders greater than 1, the linear capacitance term along with a charge source that depends on lower order Volterra responses are present, the quiescent sources are not present. The circuits are solved sequentially, starting with the first order circuit and proceeding to higher order circuits. Since each order Volterra circuit is linear, linearity and superposition can be used to analyze the circuit. However since higher order circuits depend upon lower order responses in a complicated way, linearity and superposition can not be used across circuits of different orders.

For mildly nonlinear circuits it is expected that the Taylor expansion (2.1) converges rapidly, so only a few terms are required to accurately approximate the nonlinear function. Typically, the expansion is truncated after N terms. The term given by $g_n(\bullet)$ in (2.5) and (2.6) depends upon all lower order responses and the Taylor series co-efficients. Thus, even if the Taylor series is truncated after order N, the Volterra circuits of order greater than N will still produce a response. It is expected that the Volterra circuits will eventually converge based on converging values for $g_n(\bullet)$ and can also be truncated after a few terms.

The development of this method is compared with the commonly used small signal analysis of mildly nonlinear circuits. In small signal analysis only the first two terms in Taylor series

expansion are used and

$$q \cong q_0 + C_1 \hat{v} = q_0 + C_1 (v - v_0).$$ (2.7)

Eq. (2.7) is the equation of a straight line in the q-v plane, with slope C_1 and intercept q_0. Further, since the circuit is linear, superposition applies and the complete response q can be broken up into the quiescent (DC) response q_{DC} and the change from AC, or AC response q_{AC}

$$q = q_{AC} + q_{DC}$$ (2.8)

Comparing (2.7) and (2.8) and noting that $\hat{v} = v - v_0 = v_{AC}$,

$$q_{AC} = C_1 v_{AC}$$ (2.9)

is identified.

To summarize, linearity and superposition simplifies the analysis considerably. The AC and DC analysis can be separated from each other. Eq. (2.9) is used to replace the nonlinear capacitor with a linear capacitor. In this way, AC analysis of the original nonlinear circuit turns into the frequency domain analysis of a linear circuit. Thus, small signal analysis is equivalent to taking only the first order term in Volterra series analysis.

The same procedure is used to obtain the Volterra representation of any element. These are collected in Fig. 2.3 for the common one dimensional elements used in circuits. Nonlinear admittance and transformer can be derived in a similar way. By definition, independent sources are first-order elements.

Using Fig. 2.3, an n-th order equivalent circuit is obtained in which the branch current and the nodal voltages are the n-th components of the original nonlinear circuit and the embedded independent sources are known function of circuit response of orders less than n. All circuits of different orders satisfy KCL and KVL laws. Based on these two points, the complete nonlinear
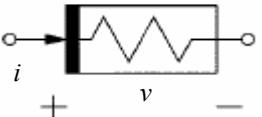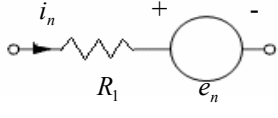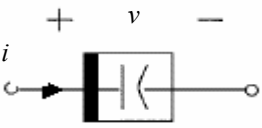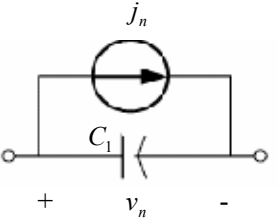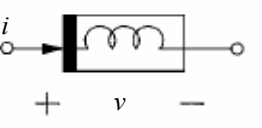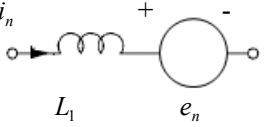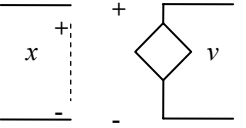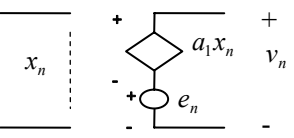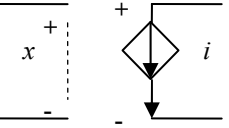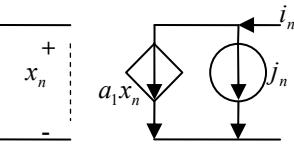
Nonlinear Elements	Equivalent Volterra Circuits

$$v = R_1 i + R_2 i^2 + R_3 i^3$$

$$e_1 = 0$$
$$e_2 = R_2 i_1^2$$
$$e_3 = 2R_2 i_1 i_2 + R_3 i_1^3$$

$$i = \frac{d}{dt}(C_1 v + C_2 v^2 + C_3 v^3)$$

$$j_1 = 0$$
$$j_2 = \frac{d}{dt}(C_2 v_1^2)$$
$$j_3 = \frac{d}{dt}(2C_2 v_1 v_2 + C_3 v_1^3)$$

$$v = \frac{d}{dt}(L_1 i + L_2 i^2 + L_3 i^3)$$

$$e_1 = 0$$
$$e_2 = \frac{d}{dt}(L_2 i_1^2)$$
$$e_3 = \frac{d}{dt}(2L_2 i_1 i_2 + L_3 i_1^3)$$

$$v = a_1 x + a_2 x^2 + a_3 x^3$$

$$e_1 = 0$$
$$e_2 = a_2 x_1^2$$
$$e_3 = 2a_2 x_1 x_2 + a_3 x_1^3$$

$$i = a_1 x + a_2 x^2 + a_3 x^3$$

$$j_1 = 0$$
$$j_2 = a_2 x_1^2$$
$$j_3 = 2a_2 x_1 x_2 + a_3 x_1^3$$

Figure.2.3 Nonlinear elements and their Volterra circuits

circuit analysis is reduced to sequential analyses of linear circuits. All the analysis methods and insights applicable to linear circuit can be used.

2.2 Numerical Example—Nonlinear RC Circuit

In this section, as an example, consider a simple nonlinear RC in Fig. 2.4(a). $R = 1\,\Omega, C_1 = 1\,F, C_2 = 0,\ C_3 = 1\,F/V^2$. Refer to Fig. 2.3 for parameter definitions. $A\cos t$ is the AC component in the independent source of the circuit. To make the following deduction easy to understand, the DC component in the independent source is set to zero. However, this does not influence the analysis procedure since DC component can be interpreted as one of a special AC component with zero frequency. The quiescent values of the nonlinear capacitor can easily be found to be $q_0 = 0$ and $v_0 = 0$. Thus, for this specific circuit, no quiescent sources are present in the first order circuit.
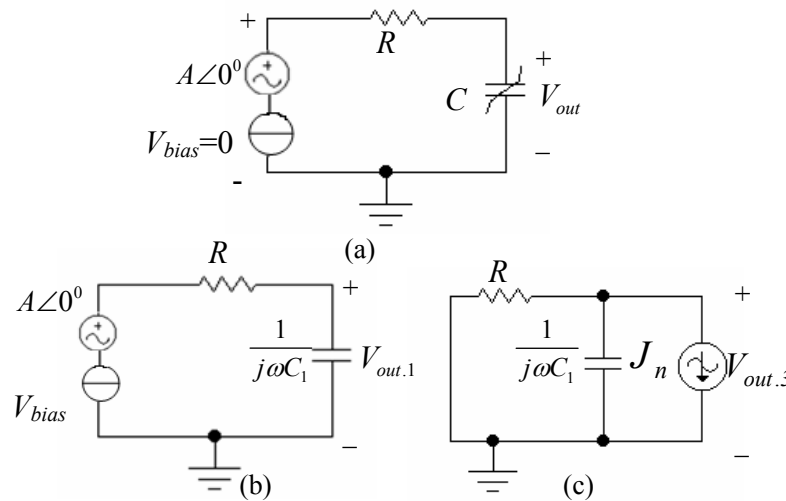
Figure.2.4
(a). Nonlinear RC circuit. (b). First order equivalent circuit (c) Higher-order equivalent circuit

The voltage across the nonlinear capacitor V_{out} is of interest. Since $C_2 = 0$, there is no source in the second order equivalent circuit and $V_{out.2} = 0$. Referring to the capacitor model in Fig. 2.3, a third order equivalent circuit in Fig. 2.4(c) is obtained in which

$$J_3 = C_3 s \cdot (v_{out.1}{}^3) \tag{2.10}$$

14

V_{out} equals the sum of $V_{out.1}$ and $V_{out.3}$, the output of the first and third order Volterra Series. The voltage source is expressed in the complex exponential function form:

$$V_{in} = A\cos t = \frac{A}{2}e^{jt} + \frac{A}{2}e^{-jt}.$$

Since

$$V_{out.1} = V_{in}\frac{1}{1+j\omega\ C_1} \tag{2.11}$$

Substituting $\omega = 1,-1$ into (2.11) produces the first order response:

$$V_{out.1} = \frac{A(1-jC_1)}{2(1+C_1^2)}e^{jt} + \frac{A(1+jC_1)}{2(1+C_1^2)}e^{-jt} \tag{2.12a}$$

For simplicity, in the following derivation, a vector to represent a variable's certain order component is used. In the vector, the response at different frequencies is organized in the form of a sequence of coordinates, separated by semicolons. The first number in the coordinate represents the angular frequencies; the second is the corresponding phasors, from which amplitude and phase information is easily derived. Notice for a given order, the two phasors for frequencies $\pm\omega$ are complex conjugate of each other. Thus only the response corresponding to non-negative frequencies are given to keep the expression concise. In this way, (2.12a) turns into:

$$V_{out.1} = [(1,\frac{A(1-jC_1)}{2(1+C_1^2)})]. \tag{2.12b}$$

Substituting (2.12b) into (2.10) produces the third order equivalent current source of the nonlinear capacitor:

$$J_{C.3} = \left[\left(1, a((C_1 + C_1^3) + j(1+C_1^2))\right); \left(3, a((3C_1 - C_1^3) + j(1-3C_1^2))\right)\right]. \tag{2.13}$$

Where the constant is defined as

15

$$a = \frac{3C_3 A^3}{8(1+C_1^2)^3} \cdot \tag{2.14}$$

Considering

$$V_{out.3} = -J_{C.3} \frac{1}{1+j\omega \ C_1} , \tag{2.15}$$

the third order response is obtained after substituting (2.13) into (2.15):

$$V_{out.3} = [(1, -af_1); (3, -af_3)], \tag{2.16}$$

in which

$$f_1 = \frac{(C_1 + C_1^3) + j(1 + C_1^2)}{1 + jC_1} \quad \text{and} \tag{2.17}$$

$$f_3 = \frac{(3C_1 - C_1^3) + j(1 - 3C_1^2)}{1 + 3jC_1} \cdot \tag{2.18}$$

Substituting parameter values into (2.12b) and (2.16) produces:

$$V_{out.1} = [(1, \ 0.25 - j0.25)]. \tag{2.19}$$

$$V_{out.3} = [(1, \ -9.375e-3); (3, \ 1.875e-3 + 3.75e-3j)]. \tag{2.20}$$

(2.19) and (2.20) calculated by hand are the same as those from software based on the above algorithm.

In [5], Volterra kernels are implemented assuming the standard description of nonlinear circuits by nonlinear algebraic-differential equations. However, as the size of circuit grows, the solution of the algebraic-differential equations may be very complicated. If more than one source exists and/or every source contains more than one frequency, in the "Volterra Kernel" [4] method, the third order Volterra Kernel will be a cumbersome tensor. In "phasor analysis" method [4],

different formulae are required for response at different frequency points, e.g. intermodulation, harmonics, and different computation procedures are required as the number of sources/frequency points change. In the proposed method, unlike the methods in [4,5], Volterra Kernels are not determined explicitly, however, the numerical results derived are the same. Thus, it is equivalent to "Volterra Kernel" and more generalized than "phasor analysis". It is straightforward, easy to program and computationally cheap.

2.3 Multi-dimensional Nonlinearity

In Section 2.1, one dimensional nonlinear elements and their equivalent Volterra circuits were considered. In this section, multi-dimensional nonlinearities, which are used in semiconductor device models, are considered. For example, two-dimensional transconductance is needed to model the collector current of bipolar transistors including Early effect, three-dimensional charge models are used in BSIM3 models. The total drain current i_D of a MOS transistor is a function of V_{GS}, V_{DS} and V_{SB} and can be expanded:

$$
\begin{aligned}
i_D &= i_D(v_{GS}, v_{DS}, v_{SB}) \\
&= i_D(V_{GS} + v_{gs}, V_{DS} + v_{ds}, V_{SB} + v_{sb}) \\
&= I_D(V_{GS}, V_{DS}, V_{SB}) + i_d(v_{gs}, v_{ds}, v_{sb}).
\end{aligned}
$$

(2.21)

Using a power series expansion around the quiescent value, the total value of the current can be split into a quiescent part I_D and an AC part i_d as in (2.21). The AC value is given as follows. Notice that following the definition of power series expansion, all the coefficients in (2.22) are evaluated at the DC operating point V_{GS}, V_{SB} and V_{DS}.

$$i_d = i_d(v_{gs}, v_{ds}, v_{sb})$$

$$= g_m \cdot v_{gs} + K_{2gm} \cdot v_{gs}^2 + K_{3gm} \cdot v_{gs}^3 \ldots$$

$$+ g_o \cdot v_{ds} + K_{2go} \cdot v_{ds}^2 + K_{3go} \cdot v_{ds}^3 \ldots \Bigg\} \begin{array}{l} \text{three} \\ \text{one-dimensional} \\ \text{transconductances} \end{array}$$

$$+ g_{mb} \cdot v_{sb} + K_{2gmb} \cdot v_{sb}^2 + K_{3gmb} \cdot v_{sb}^3 \ldots$$

$$+ K_{gm\&gmb} \cdot v_{gs} \cdot v_{sb} + K_{2gm\&gmb} \cdot v_{gs}^2 \cdot v_{sb} + K_{gm\&2gmb} \cdot v_{gs} \cdot v_{sb}^2 \ldots$$

$$+ K_{gm\&go} \cdot v_{gs} \cdot v_{ds} + K_{2gm\&go} \cdot v_{gs}^2 \cdot v_{ds} + K_{gm\&2go} \cdot v_{gs} \cdot v_{ds}^2 \ldots \Bigg\} \begin{array}{l} \text{three} \\ \text{two-dimensional} \\ \text{transconductances} \end{array}$$

$$+ K_{gmb\&go} \cdot v_{sb} \cdot v_{ds} + K_{2gmb\&go} \cdot v_{sb}^2 \cdot v_{ds} + K_{gmb\&2go} \cdot v_{sb} \cdot v_{ds}^2 \ldots$$

$$+ K_{gm\&gmb\&go} \cdot v_{gs} \cdot v_{sb} \cdot v_{ds} + \ldots \longrightarrow \begin{array}{l} \text{one three-dimensional} \\ \text{transconductances} \end{array}$$

$$(2.22)$$

The small signal drain current i_d can be split into three distinct parts. The first three power series contain only powers of one voltage. These series correspond to one-dimensional nonlinear transconductances. The next three power series contain only cross-products of two voltages, corresponding to two-dimensional transconductances. The final power series contains only cross-terms of three voltages. Since only nonlinear effects up to third order are considered, only the first term of this power series is taken into account.

The first derivatives of i_D with respect to the controlling voltages v_{GS}, v_{SB} and v_{DS} are the small-signal parameters g_m, g_{mb} and g_o. The symbols of these parameters are used as the subscripts of the nonlinear coefficients.

In general, the terms in (2.22) other than the linear ones can be expressed in the form of "$K_{n_1 gm\&n_2 gmb\&n_3 go} \cdot v_{gs}^{n_1} \cdot v_{sb}^{n_2} \cdot v_{ds}^{n_3}$" in which $n_i = 0, 1 \ldots\ldots, (i = 1, 2, 3)$. The nonlinearity coefficient $K_{n_1 gm\&n_2 gmb\&n_3 go}$ is defined as:

$$K_{n_1 gm\&n_2 gmb\&n_3 go} = \frac{1}{n_1!} \cdot \frac{1}{n_2!} \cdot \frac{1}{n_3!} \cdot \frac{\partial^{n_1+n_2+n_3} i_D(v_{GS}, v_{SB}, v_{DS})}{\partial v_{GS}^{n_1} \partial v_{SB}^{n_2} \partial v_{DS}^{n_3}} \tag{2.23}$$

Since the drain current is linearly dependent on v_{DS} in (A.4), the coefficients in (2.22) which are

second or higher order derivatives of v_{DS} are zeros.

Nonlinear transconductances of two and three dimensions and their equivalent Volterra Circuits are given in Fig. 2.5. For comparison, one dimensional nonlinearity is included as well.
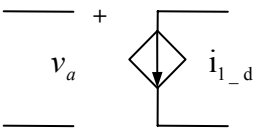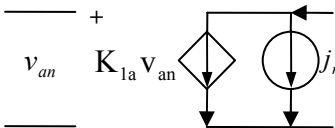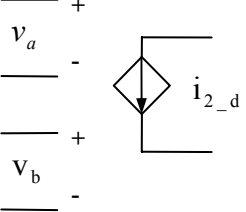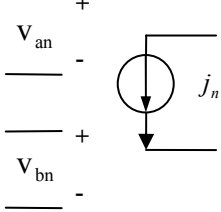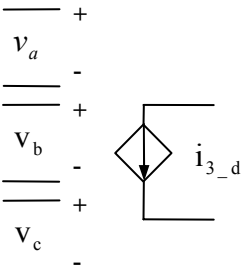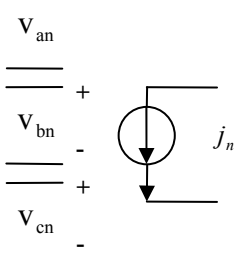
Nonlinear Elements	Equivalent Volterra Circuits

$i_{1\_d} = i_{1\_d}(v_a; K_{1a}, K_{2a}, K_{3a})$
$\quad = K_{1a} \cdot v_a + K_{2a} \cdot v_a{}^2 + K_{3a} \cdot v_a{}^3$

$j_1 = 0$
$j_2 = K_{2a} \cdot v_{a1}{}^2$
$j_3 = 2K_{2a} \cdot v_{a1} \cdot v_{a2} + K_{3a} \cdot v_{a1}{}^3$

$i_{2\_d} = i_{2\_d}(v_a, v_b; K_{1a\&1b}, K_{2a\&1b}, K_{1a\&2b})$
$= K_{1a\&1b} \cdot v_a \cdot v_b + K_{2a\&1b} \cdot v_a{}^2 \cdot v_b$
$\quad + K_{1a\&2b} \cdot v_a \cdot v_b{}^2$

$j_1 = 0$
$j_2 = K_{1a\&1b} \cdot v_{a1} \cdot v_{b1}$
$j_3 = K_{1a\&1b} \cdot v_{a1} \cdot v_{b2} + K_{1a\&1b} \cdot v_{a2} \cdot v_{b1}$
$\quad + K_{2a\&1b} \cdot v_{a1}{}^2 \cdot v_{b1} + K_{1a\&2b} \cdot v_{a1} \cdot v_{b1}{}^2$

$i_{3\_d} = i_{3\_d}(v_a, v_b, v_c; K_{1a\&1b\&1c})$
$\quad = K_{1a\&1b\&1c} \cdot v_a \cdot v_b \cdot v_c$

$j_1 = 0$
$j_2 = 0$
$j_3 = K_{1a\&1b\&1c} \cdot v_{a1} \cdot v_{b1} \cdot v_{c1}$

Figure. 2.5   Transconductances of different dimensions and their Volterra circuits

Notice the equivalent Volterra Circuits in Fig. 2.5 are similar to the Volterra Kernels introduced in [4]. They have the same effect because both are based on [1].

2.4 Frequency Domain Response of Nonlinear Circuit

In this section, the nonlinear system's frequency domain response is calculated using MNA (Modified Nodal Analysis) formulation. Refer to [3] for more details about the entry of linear elements in MNA. The frequency domain response of a simple RC circuit is calculated both by hand and using the proposed method to verify the method.

According to the rules in [3] for MNA, formulate the system of linear equations in the form

$$T\,X_n = (G + sC)X_n = W_n \qquad (2.24)$$

in which $T = (G + sC)$. The only structural difference between first-order and higher-order equivalent circuits is the value and location of sources: independent sources appear only in the first-order circuit, while equivalent sources are added to higher-order ones. Expressed in MNA formulation, since the circuit structure does not change except the sources, G and C matrices are the same for all orders. Secondly, the equivalent sources for the nonlinear elements will have corresponding entries in the right hand side vector W_n and all the higher-order equivalent circuits have the same structure. This means for $n > 1$, g_n which represents the equivalent sources' connections does not change, only the numerical values of V_n change. To summarize, the following algorithm is used to compute the nonlinear circuit's frequency response:

1. MNA formulation of the first order circuit, produces the G and C matrixes. Set n equal to 1. n represents the present order under computation. Calculate the LU factors of $T(s)$

20

corresponding to different frequency points.

2. Solve the system's first order response using the independent sources.

3. Increase n by 1, compute W_n based on the nonlinearity coefficients and known lower-order circuit variables.

4. Substitute W_n and selected frequency, e.g. harmonics, in (2.24) and obtain the n-th order response X_n.

5. If n equals the highest order, predetermined based on the precision requirement, then stop; else go to 3.

2.5   Multidimensional Example—Cascode Amplifier

In this section, the Cascode Amplifier in Fig. 2.6(a) is used to illustrate the frequency domain analysis of nonlinear circuits. The transistor model and technology parameters are given in Appendix A. The transistor sizes are

$$W_A / L_A = 3\mu m / 1\mu m , \quad W_B / L_B = 3\mu m / 1\mu m , \quad W_C / L_C = 6\mu m / 1\mu m .$$

Biasing voltages are selected so the DC component of V_{out} is the midpoint between *Gnd* and V_{DD} :

$$V_{bias1} = 1.7 \ V , V_{bias2} = 3.05 \ V , V_{bias1} = 3.3 \ V , \ V_{DD} = 5 \ V.$$

The DC operating point is calculated to be:

$$V_2 = 1.184 \, V , V_3 = 2.6040 \, V , \ I_D = 157.78 \, \mu A , \ V_{T2} = 0.9144 \, V .$$
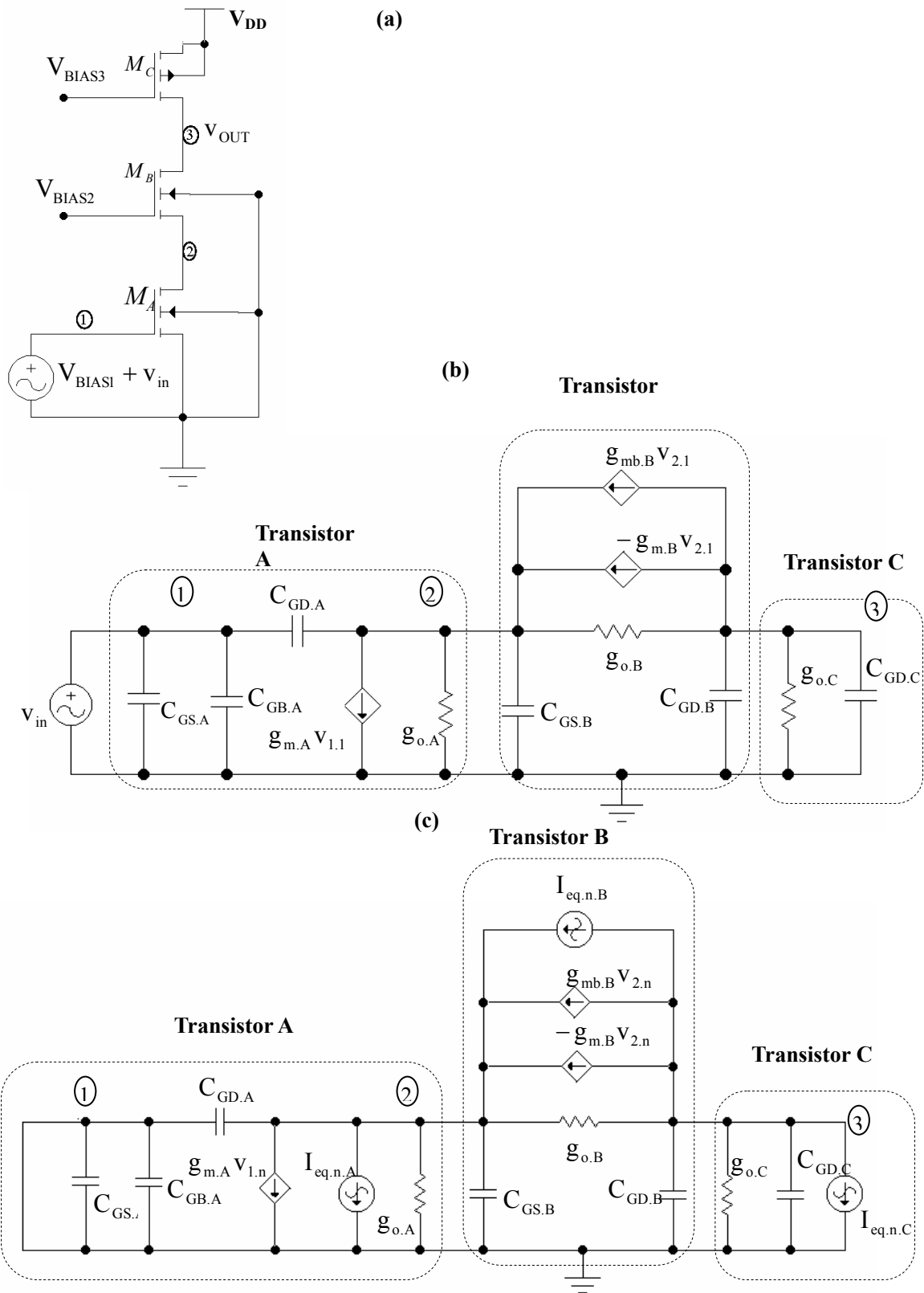
Figure 2.6 (a). Cascode amplifier. (b). First order equivalent circuit (c) Higher-order equivalent circuit

22

With the operating points, the coefficients for Volterra series analysis are easily computed following (2.23).

The flow diagram for simulation is shown is Fig. 2.7, which follows the same manner as SPICE like simulators. In SPICE, small signal analysis follows DC analysis; here Volterra Analysis follows
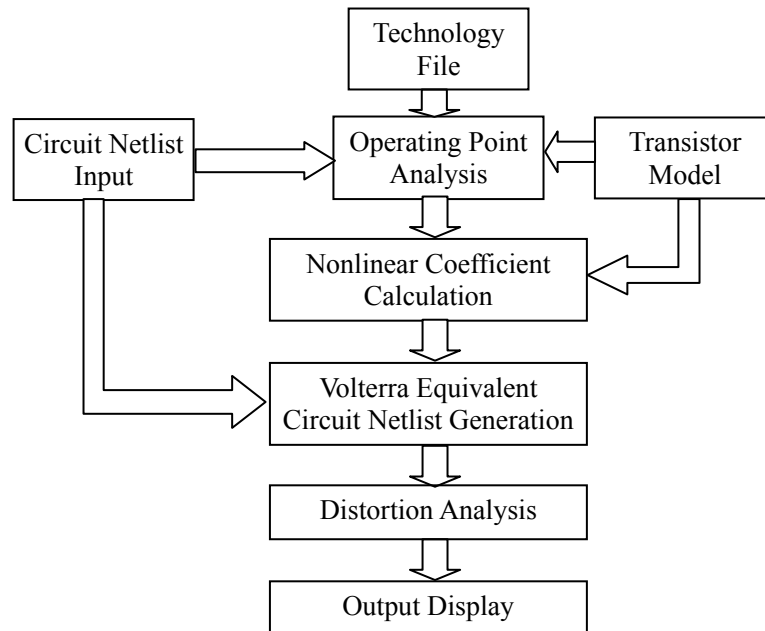
Figure 2.7 Nonlinear circuit analysis procedure

operating point analysis. Operating point analysis provides an appropriate starting point for the Taylor Series expansion to calculate nonlinear coefficients. As mentioned in Section 2.2, the DC sources need not be taken into account in operating point analysis alone. If there is a difference between the operating point analysis result and the response from DC sources, the difference can be thought of as the "AC" input signal with zero frequency. It can thus be included in the first order Volterra circuit. In other words, on the nonlinear function curves, the expansion does not have to be performed exactly at the quiescent point; it can be any point in the surrounding region, so long as convergence is guaranteed. However, in practical applications, the response from quiescent DC analysis is usually taken as the starting point. The general analysis procedure of "operating point + Volterra" degenerates

into "DC+Volterra". This is based on two reasons:

1. In some circuits, strong nonlinearities exist, e.g. the exponential current-voltage relationship in diodes or bipolar transistors. In these circuits, if the operating point of one variable deviates even slightly from the quiescent point, a large deviation can be caused in other variables. This can cause convergence problem in Volterra analysis.

2. Unless the operating point is chosen at the quiescent point, higher order Volterra analysis might be required to achieve high accuracy. However, this is impractical since for the commonly used transistor compact models, (a). higher order derivatives are extremely complicated to evaluate　(b). the accuracy of higher order derivatives is usually unsatisfactory. Besides, it usually takes a couple of extra Newton-Raphson iterations to reach the quiescent point. It follows that, generally, it will be computationally cheaper to choose the quiescent point as the operating point.

However, certain situations exist when it might not be a good idea to choose the DC quiescent point as the operating point for Volterra analysis:

1. Theoretically, the best accuracy is achieved through Volterra analysis at the transient quiescent point. In general, the transient quiescent point is not the DC quiescent point, because of the DC components generated from even order nonlinearities of the circuit. In most situations, the difference is small and can be neglected. However, in the cases that the nonlinearity is strong, high precision is desired or the input signal is large, such differences can not be neglected and the transient quiescent point is required. It can be achieved through a high-precision transient analysis long enough to ensure the steady state is reached. It can also be calculated from a preceding Volterra analysis: the transient quiescent point is the sum of the DC quiescent point and the DC components from all the even order Volterra analyses.

2. Consider the case when distortion response is of interest when one or all of the DC sources are swept. If the change in DC bias is kept within certain limits, Volterra analyses can be performed at the same bias point for different DC sources. The difference in DC sources will be taken into account in the first order Volterra analysis. The reason is: model function evaluation usually represents the

majority of the simulation time and it is generally expensive to calculate nonlinear coefficients at each bias point. Thus, if precision is not of utmost concern, efficiency and precision can be traded off.

For the cascode amplifier in Fig. 2.6 (a), following the general guideline, the operating point is calculated through DC analysis. Then the small signal coefficients such as those in (2.22) are extracted based on the transistor model, technology parameters and the DC operating point. These coefficients can be used to generate the small signal netlist based on circuit topology. The netlist is then fed to the simulator for distortion analysis. The equivalent circuits of the Cascode amplifier are shown in Fig. 2.6 (b, c). Notice, since the DC quiescent point is used as the operating point, there is no DC term in the first order Volterra circuit.

The MNA formulation coefficients in (2.24) are as follows:

$$
CC = \begin{bmatrix} C_{GS.A} + C_{GB.A} + C_{GD.A} & -C_{GD.A} & & \\ -C_{GD.A} & C_{GD.A} + C_{GS.B} & & \\ & & C_{GD.B} + C_{GD.C} & \\ & & & 0 \end{bmatrix} \tag{2.25}
$$

$$
GG = \begin{bmatrix} & & & 1 \\ g_{m.A} & g_{o.A} + g_{o.B} + g_{m.B} - g_{mb.B} & -g_{o.B} & \\ & -g_{o.B} - g_{m.B} + g_{mb.B} & g_{o.B} + g_{o.C} & \\ 1 & & & \end{bmatrix} \tag{2.26}
$$

$$
X = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ i_1 \end{bmatrix} \qquad RHS_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ v_{in} \end{bmatrix} \qquad RHS_n = \begin{bmatrix} -I_{eq.A} + I_{eq.B} \\ -I_{eq.B} + I_{eq.C} \end{bmatrix} \tag{2.27}
$$

In which,

$$
I_{eq.A} = i_{1\_d}(v_1; g_{m.A}, K_{2gm.A}, K_{3gm.A}) + i_{2\_d}(v_1, v_2; K_{gm\&go.A}, K_{2gm\&go.A}, K_{gm\&2go.A}) \tag{2.28}
$$

$$
\begin{aligned}
I_{eq.B} = & i_{1\_d}(-v_2; g_{m.B}, K_{2gm.B}, K_{3gm.B}) + i_{1\_d}(v_2; g_{mb.B}, K_{2gmb.B}, K_{3gmb.B}) + i_{2\_d}(-v_2, v_2; K_{gm\&gmb.B}, \\
& K_{2gm\&gmb.B}, K_{gm\&2gmb.B}) + i_{2\_d}(-v_2, v_3 - v_2; K_{gm\&go.B}, K_{2gm\&go.B}, K_{gm\&2go.B}) + i_{2\_d}(v_2, v_3 - v_2; \\
& K_{gmb\&go.B}, K_{2gmb\&go.B}, K_{gmb\&2go.B}) + i_{3\_d}(-v_2, v_2, v_3 - v_2; K_{gm\&gmb\&go.B}) \tag{2.29}
\end{aligned}
$$

$$I_{eq.C} = 0 \qquad\qquad (2.30)$$

From (2.28- 2.30), there are 16 nonlinear coefficients in this Cascode Amplifier, 4 from M_A, the rest from M_B; none comes from M_C because its source and gate are DC connected and output resistance is linear in this transistor model.

Following the Volterra analysis procedure in Fig. 2.7, the plot of the third order harmonic (HD_3) at the output node, under the input AC excitation with amplitude of 1mV, is plotted in Fig. 2.8.
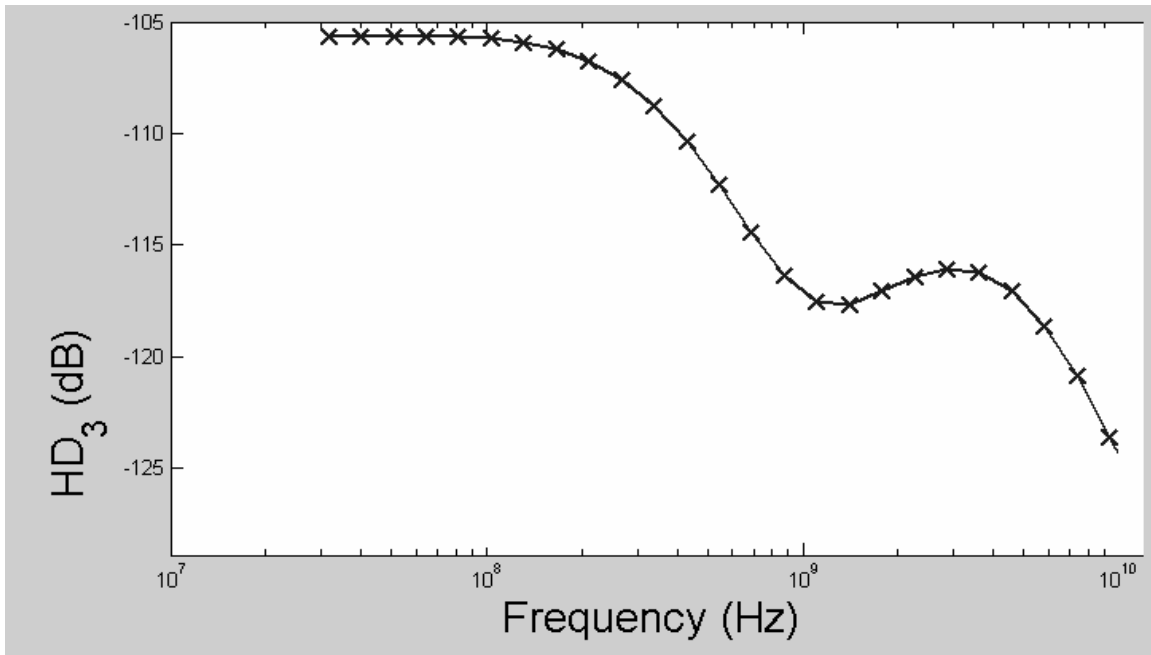
Figure. 2.8 Plot of HD_3 w.r.t. frequency

Chapter 3

Sensitivity Analysis of Mildly Nonlinear Circuits

Sensitivity analysis plays an important role in analog Computer Aided Design (CAD). Based on the Volterra analysis introduced in Chapter 2, the sensitivity analysis of mildly nonlinear circuits is introduced in Section 3.1, and its computational cost is analyzed. To maintain accuracy while improving efficiency, a novel sensitivity calculation technique based on Nonlinearity Transfer Matrix (NTM) is proposed in Section 3.2, and its physical significance is also examined. Finally, two examples show over 100 times speedup in sensitivity calculations based on NTM, with the same accuracy. More savings are expected in higher order distortion analysis with multi-tone excitation. Application of the proposed method is expected to produce improvement in efficiency, capability and accuracy in different areas of electronic design automation.

3.1   Sensitivity Analysis

In this section, we introduce a straightforward numerical method to calculate sensitivity w.r.t. parameters of a mildly nonlinear circuit in frequency domain. It is a natural extension of the Volterra analysis approach from Chapter 2. The validity of the algorithm is proven by comparing

results derived from hand calculation and from the simulation software.

Firstly, the sensitivity of a variable V_{out} w.r.t. any parameter h, defined by $\dfrac{dV_{out}}{dh}$ is calculated. For the circuit in Figure. 2.4 (a), the parameters are C_1, C_2 and C_3. Theoretically, there are nine sensitivities to calculate since V_{out} is the sum of three terms: $V_{out.n}$, $n = 1,2,3$ and the capacitor has three coefficients C_j, $j = 1,2,3$. However, $\dfrac{dV_{out.n}}{dC_j}$ is zero for $n < j$ since a higher-order nonlinearity has no influence on lower-order response, so only six sensitivities are left: $dV_{out.3}/dC_1$, $dV_{out.2}/dC_1$, $dV_{out.1}/dC_1$, $dV_{out.3}/dC_2$, $dV_{out.2}/dC_2$, $dV_{out.3}/dC_3$. Although C_2 and $V_{out.2}$ are zero, the sensitivities associated with them are usually nonzero. The capacitor's influence on third-order response in the output is taken as an example. The analytical expression for $dV_{out.3}/dC_2$ is nonzero and complicated, so it is not given here. Differentiate Eq. (2.16) w.r.t C_j, $j = 1, 3$, results in:

$$\frac{dV_{out.3}}{dC_j} = [\, (\, 1, -(\frac{da}{dC_j} f_1 + a \frac{df_1}{dC_j})\,)\,);\, (\, 3, -(\frac{da}{dC_j} f_3 + a \frac{df_3}{dC_j})\,)\,].\tag{3.1}$$

Where

$$\frac{da}{dC_1} = -\frac{9 C_1 C_3 A^3}{4(1 + C_1^2)^4},\tag{3.2}$$

$$\frac{da}{dC_3} = \frac{3 A^3}{8(1 + C_1^2)^3}.\tag{3.3}$$

Since C_3 does not show up in f_1 or f_3, df_1/dC_3 and df_3/dC_3 are zero, and

$$\frac{df_3}{dC_1} = \frac{6(1 + C_1^2) - j6 C_1(1 + C_1^2)}{(1 - 9 C_1^2) + j6 C_1},\tag{3.4}$$

$$\frac{df_1}{dC_1} = \frac{2(1+C_1{}^2) + j2C_1(1+C_1{}^2)}{(1-C_1{}^2) + j2C_1}.$$ (3.5)

Substituting Eq. (3.2-3.5) into Eq. (3.1) and using the parameter values given in Section 2.3 results in:

$$\frac{dV_{out.3}}{dC_1} = [(1, (18.75 + 9.375j) \cdot 10^{-3}) ; (3, (2.25 - 12.375j) \cdot 10^{-3})];$$ (3.6)

$$\frac{dV_{out.3}}{dC_3} = [(1, -9.375 \cdot 10^{-2}) ; (3, (1.875 + 3.75j) \cdot 10^{-2})].$$ (3.7)

Consider $\omega=3$ in Eq. (3.7) to explain the physical meaning of sensitivity. As given in Section 2.2,

$$V_{out.3} = [(1, \ -9.375 \cdot 10^{-3}); (3, \ 1.875 \cdot 10^{-3} + 3.75 \cdot 10^{-3} j)].$$

If C_3 undergoes a small perturbation, for example 1% from 0.1 to 0.101, then

$$
\begin{aligned}
V_{out.3}\big|_{C3=0.101} &\cong V_{out.3}\big|_{C3=0.1} + \Delta C_3 \cdot \frac{dV_{out.3}}{dC_3} \\
&= (1.875 + 3.75j) \cdot 10^{-3} + 0.001 \cdot (1.875 + 3.75j) \cdot 10^{-2} \\
&= (1.89375 + j3.7875) \cdot 10^{-3}.
\end{aligned}
$$

The phasor $V_{out.3}$ changes from $(1.875 + j3.75) \cdot 10^{-3}$ to $(1.89375 + j3.7875) \cdot 10^{-3}$. As shown above, although it is a simple circuit with one nonlinear element, one independent source and one input frequency, the hand analysis is tedious. This motivates the search for computer methods to calculate sensitivity in frequency domain.

3.2 Adjoint Method of Sensitivity Calculation

In this section, the numerical computation of nonlinear circuit coefficient sensitivity in frequency

domain, using Adjoint system method [3] is introduced. Based on this, the sensitivities of any objective function w.r.t. all the parameters can easily be computed.

In this work, h_j represents j-th order nonlinear coefficient, e.g. h_3 represents some third order nonlinear coefficient. Subscript n is added if necessary to variables to differentiate different orders, e.g. X_3 represents the second order circuit response. N is the highest order considered. Let the output be a scalar variable $\phi(X)$. For simplicity, we restrict $\phi(X)$ to be a linear combination of components in X in the following way:

$$\phi = d^t X \,, \tag{3.8}$$

where d is a constant vector. Differentiate Eq. (2.24) of the n-th order w.r.t. h_j,

$$T \frac{dX_n}{dh_j} = -\frac{dT}{dh_j} X_n + \frac{dW_n}{dh_j} \,, \tag{3.9}$$

Rearrangement of terms results in:

$$\frac{dX_n}{dh_j} = -T^{-1}(\frac{dT}{dh_j} X_n - \frac{dW_n}{dh_j}) \,. \tag{3.10}$$

Combining Eq. (3.8) and (3.10) results in:

$$\frac{\partial \phi_n}{\partial h_j} = -d^t T^{-1}(\frac{\partial T}{\partial h_j} X_n - \frac{\partial W_n}{\partial h_j}) \,. \tag{3.11}$$

Define Adjoint vector X_n^a as

$$(X_n^a)^t = -d^t T^{-1} \,. \tag{3.12}$$

where X_n^a is the solution of the equation:

$$T^t X_n^a = -d \,. \tag{3.13}$$

Substitute Eq. (3.12) into Eq. (3.11) produces the final form for the sensitivity calculation

$$\frac{d\phi_n}{dh_j} = (X_n{}^a)' \frac{dT}{dh_j} X_n - (X_n{}^a)' \frac{dW_n}{dh_j} . \tag{3.14}$$

There are some comments about Eq. (3.14) to simplify computation:

(1) If $n<j$, $d\phi_n / dh_j$ is zero since higher-order nonlinearity has no influence on lower-order response.

(2) If $j>1$, the first term in the RHS of Eq. (3.14) disappears since higher-order nonlinearity h_j does not appear in the T matrix, so $\frac{dT}{dh_j} = 0$. If $j=1$ and all the sensitivities are of interest, except those w.r.t independent sources, the second term disappears since first-order coefficients appear only in the T matrix and $\frac{dW_i}{dh_j} = 0$.

(3) The same LU factors for T can be used for the calculation of X^a in Eq. (3.13).

(4) As mentioned before, the above sensitivity calculation need only be evaluated at the frequency points of interest.

(5) The phasors in X_n and $d\phi_n / dh_j$, corresponding to mirror frequencies, are conjugate of each other. This reduces the computation by 50% in both frequency response and sensitivity calculations.

For the nonlinear RC circuit in Fig. 2.4, sensitivities calculated based on the above algorithm are the same as the hand calculation results Eq. (3.6) and (3.7). As shown above, the frequency response computation is closely related to network topology. Based on this, the sensitivity analysis shows the effect of the network nonlinearities on various circuit variables. It can be used in either gradient based optimization or circuit analysis.

3.3 Computation Cost Analysis

In this section, the computation cost of the sensitivity calculation technique, from the previous section, is analyzed. This will point out the bottleneck in the Adjoint method based sensitivity calculation.

The n-th order Volterra analysis is as follows:

$$TX_n = W_n . \tag{3.15}$$

The computation of Eq. (3.15) includes two parts: the generation of equivalent source W_n, and the solution of system equation. The complexity of the latter is determined by the size of T, or the number of total nodes. On the other hand, the complexity in generating W_n is determined by two parts: firstly, the number of nonlinear coefficients and, secondly, how each nonlinear coefficient contributes to W_n. As shown below, system equation solution dominates in the first order Volterra analysis; while the formulation of W_n dominates in the higher order ones.

First, compare the number of nodes and nonlinear coefficients. Suppose the circuit is composed of T transistors, each transistor has K nodes, and the analysis is N-th order. Then the number of nodes and nonlinear coefficients are $O(TK)$ and $O(TN^K)$, respectively. Consider the simple example of a MOS transistor in common-source configuration, $T=1$. The nonlinear elements considered are the drain current I_{DS} and terminal charges Q_B, Q_G and Q_D. $K=3$ because the controlling variables are v_{GS}, v_{DS} and v_{BS}. In a third order Volterra analysis, $N=3$. In this example, there are 3 nodes (except ground) and 64 nonlinear coefficients. In general, nonlinear coefficients outnumber nodes in the Volterra analysis of a practical circuit.

Secondly, each nonlinear coefficient contributes to W_n by generating a polynomial based on lower order circuit response, from X_1 to X_{n-1}. The polynomial usually contains a few terms which are also frequency dependent. For example, $K_{m\&d\&b}$ contributes to W_5 by generating

$$K_{m\&d\&b}(v_{gs.2}v_{ds.2}v_{bs.1}+v_{gs.2}v_{bs.2}v_{ds.1}+v_{ds.2}v_{bs.2}v_{gs.1}+v_{gs.3}v_{ds.1}v_{bs.1}+v_{gs.1}v_{bs.3}v_{ds.1}+v_{ds.3}v_{bs.1}v_{gs.1}).$$

In general, the total number of operations required to generate the polynomial is the product of the number of terms, variables per term and the number of frequency components in each variable. For example, it takes 176 additions and 192 multiplications to compute the above entry in W_5 in a two-tone analysis. This analysis shows, the generation of W_n is the bottleneck in VS analysis, due to the large number of nonlinear coefficients and the corresponding polynomial entries.

Next, analyze the computational cost of the Adjoint method based sensitivity calculation technique. Eq. (3.9) shows dX_n/dh_j can be interpreted as the response to an equivalent sensitivity circuit, which has the same structure as the corresponding Volterra analysis. The only difference is the right hand side. At the same time, through reusing the same LU factors, the solution of system equation in the sensitivity analysis of Eq. (3.9) is cheaper than that of the Volterra analysis of Eq. (3.15).

Next compare the complexity in the generation of the RHS of Eq. (3.9) and Eq. (3.15). The first term in the RHS of (3.9) is not of a problem since dT/dh_j is nonzero only when $j=1$ and it contains at most 4 nonzero entries. If $j>1$ and $j=n$, the second term is also cheap to evaluate since it contains at most two nonzero entries; otherwise, however, dW_n/dh_j becomes extremely complicated. The reason is that W_n depends on all the lower order circuit responses, some of which are also dependent on h_j. The polynomial entries of nonlinearities in dW_n/dh_j thus get more complicated than their counterparts in W_n. For example, examine how h_1 influences X_3: firstly, h_1 influences X_1, which then affects W_2. W_2 generates X_2 in the second order circuit. Finally, X_1 and X_2 generate W_3 together. That is, dW_3/dh_1 is computationally expensive due to: (1). dX_2/dh_1 and dX_1/dh_1 need to be computed; (2). every polynomial entry in dW_3/dh_1 is more complicated than its counterpart in W_3, as is obvious by comparing the equivalent Volterra

and sensitivity circuits in Fig. 3.1-3.3.

The first problem is not of a significant concern since the repetitive system solution of Eq. (3.9) is greatly simplified by the Adjoint method and the sharing of LU factors. The second problem results from multiple mixing effects. This makes the formulation of $dW_n / dh_j \, (\, n > j \,)$ the computational bottleneck in sensitivity calculation.

3.4   Nonlinear Transfer Matrix

The previous section shows in certain situations, sensitivity calculation is computationally more expensive than the corresponding Volterra analysis of the same order. Recall sensitivity can either be computed by direct calculation or approximated by finite difference, which needs an extra Volterra analysis. This implies that although more accurate, the direct sensitivity calculation technique is more expensive than approximation by finite difference. In order to improve efficiency, the concept of the Nonlinear Transfer Matrix is proposed in this section.

dW_n / dh_j is the bottleneck in sensitivity analysis (3.9) when $n>j$, because of the multiple mixing effects. In order to make the data dependency explicit, dW_n / dh_j is expanded in the time domain. First separate equivalent source W into two parts, W_G and W_C, as in Eq. (3.16):

$$W_n = W_{G.n} + \frac{dW_{C.n}}{dt} . \tag{3.16}$$

W_G is formulated by frequency independent nonlinearity, e.g. transconductances, and W_C is formulated by frequency dependent nonlinearity, e.g. capacitor. Expressed in this way, both $W_{G.n}$ and $W_{C.n}$ can be formulated by performing only basic operations (addition, multiplication) on

34

the nonlinear coefficients and circuit response. Notice h_j influences W_n in two ways: firstly, h_j directly appears in W_n; secondly, h_j influences W_n indirectly through multiple mixing effects: h_j influences lower order circuit response (X_j, X_{j+1} ... X_{n-1}) first, which then influences W_n. To distinguish the "direct" and "indirect" influence, dW_n / dh_j can be expanded:

$$\frac{dW_n}{dh_j} = (\frac{dW_n}{dh_j})_I + (\frac{dW_n}{dh_j})_D \tag{3.17}$$

in which

$$(\frac{dW_n}{dh_j})_D = \frac{\partial W_n}{\partial h_j} \tag{3.18}$$

and

$$(\frac{dW_n}{dh_j})_I = \sum_{k=j}^{n-1} \frac{\partial W_{G.n}}{\partial X_k} \frac{dX_k}{dh_j} + \frac{d}{dt}[\sum_{k=j}^{n-1} \frac{\partial W_{C.n}}{\partial X_k} \frac{dX_k}{dh_j}] \tag{3.19}$$

Notice Eq. (3.18) is zero when $j=1$, otherwise it has at most two nonzero entries. Its formulation from (trans)conductances is given in Fig. 3.1-3.3, the entries for nonlinear capacitors (inductors) can be derived similarly. In Eq. (3.19), $\partial W_{G.n} / \partial X_k$ and $\partial W_{C.n} / \partial X_k$ are vector-vector partial derivatives. It can be shown, $\partial W_{G.n} / \partial X_k$ and $\partial W_{C.n} / \partial X_k$ can be expressed analytically by nonlinear coefficients and circuit response, which are fixed. Thus, $\partial W_{G.n} / \partial X_k$ and $\partial W_{C.n} / \partial X_k$ are constant matrices for a given design and can be pre-computed. They are defined as Nonlinear Transfer Matrix (NTM) as follows,

$$N_{G.n.k} = (\frac{\partial W_{G.n}}{\partial X_k}), \tag{3.20}$$

and

Nonlinear Element

v_a i_{1_d}

$$i_{1\_d} = i_{1\_d}(v_a; K_{2a}, K_{3a})$$
$$= K_{2a}v_a^2 + K_{3a}v_a^3$$

Equivalent Volterra Circuit

v_{ai} j_i

$$j_1 = 0$$
$$j_2 = K_{2a}v_{a1}^2$$
$$j_3 = 2K_{2a}v_{a1}v_{a2} + K_{3a}v_{a1}^3$$

Equivalent Sensitivity Circuit

$\dfrac{\partial v_{ai}}{\partial h}$ $\dfrac{\partial j_i}{\partial h}$

Indirect Influence Component

$$(\frac{\partial j_2}{\partial h_1})_I = 2K_{2a}v_{a1}\frac{\partial v_{a1}}{\partial h_1}$$

$$(\frac{\partial j_3}{\partial h_1})_I = (2K_{2a}v_{a2} + 3K_{3a}v_{a1}^2)\frac{\partial v_{a1}}{\partial h_1}$$
$$+ (2K_{2a}v_{a1})\frac{\partial v_{a2}}{\partial h_1}$$

$$(\frac{\partial j_3}{\partial h_2})_I = (2K_{2a}v_{a1})\frac{\partial v_{a2}}{\partial h_2}$$

Direct Influence Component

$$(\frac{\partial j_2}{\partial K_{2a}})_D = v_{a1}^2 \qquad (\frac{\partial j_3}{\partial K_{2a}})_D = 2v_{a1}v_{a2} \qquad (\frac{\partial j_3}{\partial K_{3a}})_D = v_{a1}^3$$
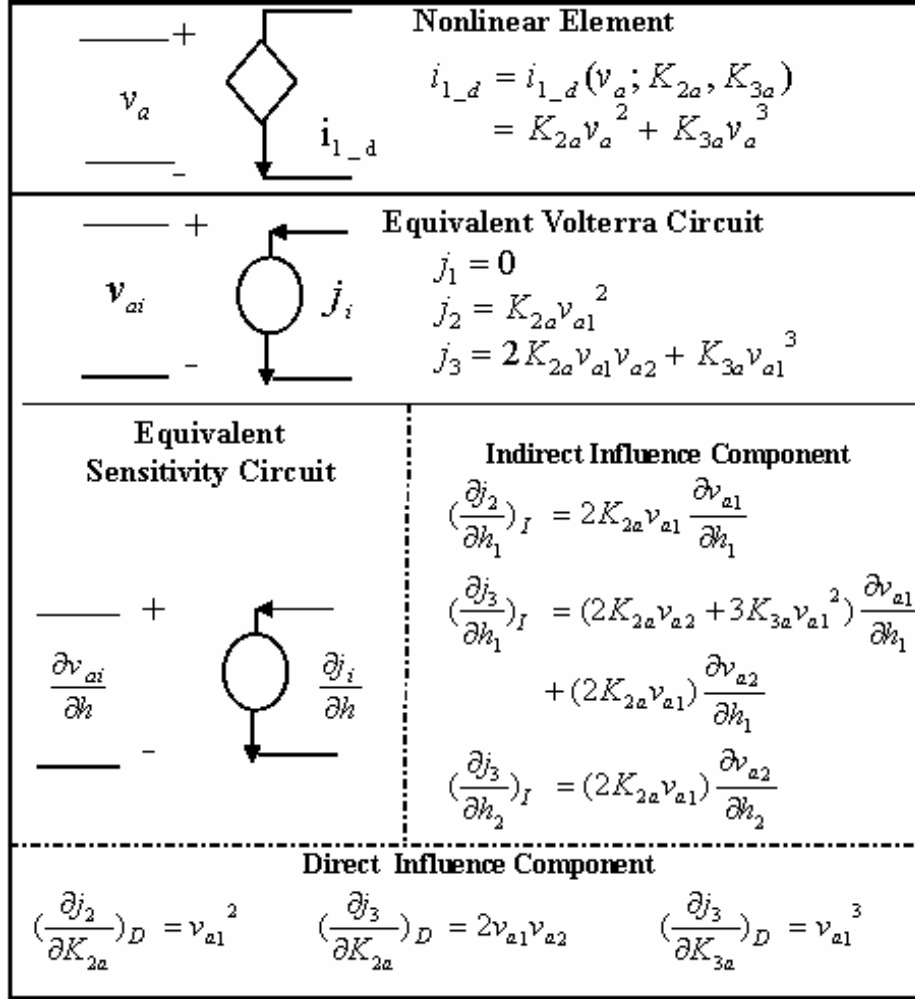
Figure.3.1 Equivalent volterra series and sensitivity circuits of 1-D transconductance

$$N_{C.n.k} = (\frac{\partial W_{C.n}}{\partial X_k}).\tag{3.21}$$

The meanings of the subscripts of NTM are self-explicit by examining the corresponding RHS of Eq. (3.20) and (3.21). Substitute Eq. (3.20) and (3.21) into Eq. (3.19), the indirect influence component $(dW_n / dh_j)_I$ can be expanded as follows:

$$(\frac{dW_n}{dh_j})_I = \sum_{k=j}^{n-1} N_{G.n.k}\frac{dX_k}{dh_j} + \frac{d}{dt}[\sum_{k=j}^{n-1} N_{C.n.k}\frac{dX_k}{dh_j}].\tag{3.22}$$

Nonlinear Element

$$i_{2\_d} = i_{2\_d}(v_a, v_b; K_{1a\&1b}, K_{2a\&1b}, K_{1a\&2b})$$
$$= K_{1a\&1b}v_a v_b + K_{2a\&1b}v_a^2 v_b + K_{1a\&2b}v_a v_b^2$$

Equivalent Volterra Circuit

$$j_1 = 0$$
$$j_2 = K_{1a\&1b}v_{a1}v_{b1}$$
$$j_3 = K_{1a\&1b}v_{a1}v_{b2} + K_{1a\&1b}v_{a2}v_{b1}$$
$$+ K_{2a\&1b}v_{a1}^2 v_{b1} + K_{1a\&2b}v_{a1}v_{b1}^2$$

Equivalent Sensitivity Circuit

Indirect Influence Component

$$\left(\frac{\partial j_2}{\partial h_1}\right)_I = K_{1a\&1b}\left(v_{a1}\frac{\partial v_{b1}}{\partial h_1} + v_{b1}\frac{\partial v_{a1}}{\partial h_1}\right)$$

$$\left(\frac{\partial j_3}{\partial h_1}\right)_I = \left(K_{1a\&1b}v_{a2} + K_{2a\&1b}v_{a1}^2 + 2K_{1a\&2b}v_{a1}v_{b1}\right)\frac{\partial v_{b1}}{\partial h_1}$$
$$+ \left(K_{1a\&1b}v_{b2} + K_{1a\&2b}v_{b1}^2 + 2K_{2a\&1b}v_{a1}v_{b1}\right)\frac{\partial v_{a1}}{\partial h_1}$$
$$+ \left(K_{1a\&1b}v_{b1}\right)\frac{\partial v_{a2}}{\partial h_1} + \left(K_{1a\&1b}v_{a1}\right)\frac{\partial v_{b2}}{\partial h_1}$$

$$\left(\frac{\partial j_3}{\partial h_2}\right)_I = \left(K_{1a\&1b}v_{b1}\right)\frac{\partial v_{a2}}{\partial h_2} + \left(K_{1a\&1b}v_{a1}\right)\frac{\partial v_{b2}}{\partial h_2}$$

Direct Influence Component

$$\left(\frac{\partial j_2}{\partial K_{1a\&1b}}\right)_D = v_{a1}v_{b1} \qquad \left(\frac{\partial j_3}{\partial K_{1a\&1b}}\right)_D = v_{a1}v_{b2} + v_{a2}v_{b1}$$

$$\left(\frac{\partial j_3}{\partial K_{2a\&1b}}\right)_D = v_{a1}^2 v_{b1} \qquad \left(\frac{\partial j_3}{\partial K_{1a\&2b}}\right)_D = v_{b1}^2 v_{a1}$$
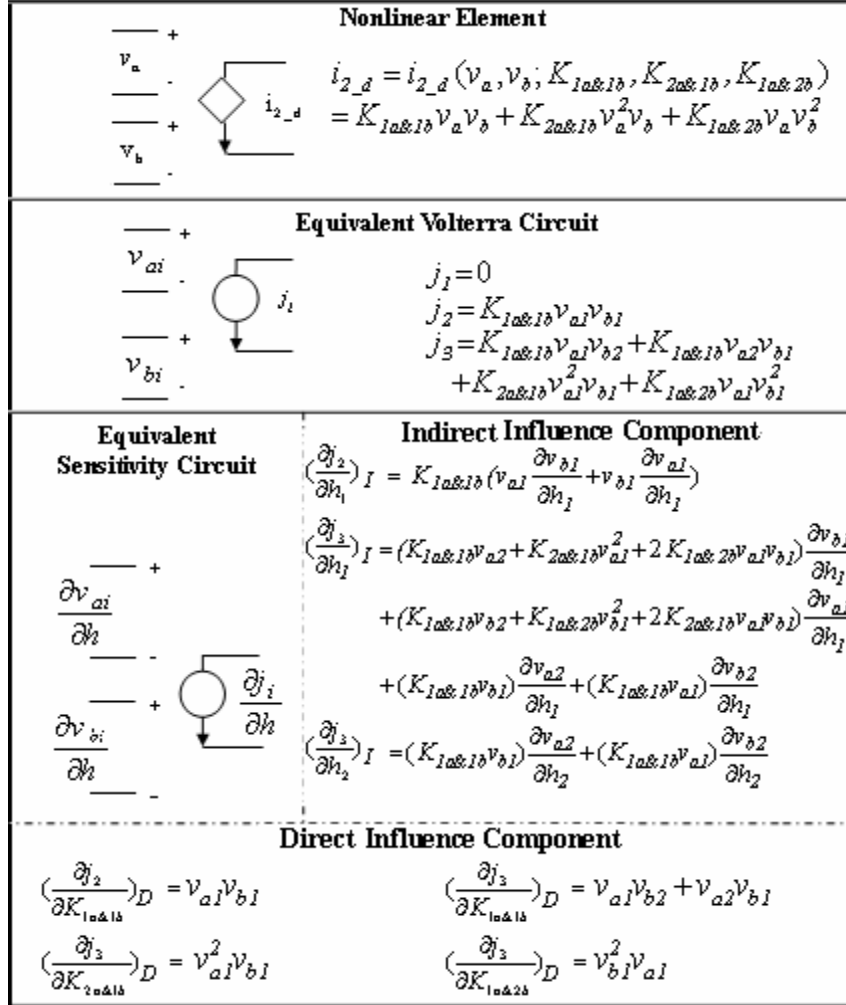
Figure.3.2 Equivalent Volterra series and sensitivity circuits of 2-D transconductance

Take a 1-D nonlinear transconductance for example. Referring to Fig. 3.1, its entry in

dW_3 / dh_1 is $(2K_{2a}v_{a2} + 3K_{3a}v_{a1}^2)\dfrac{dv_{a1}}{dh_1} + (2K_{2a}v_{a1})\dfrac{dv_{a2}}{dh_1}$. The two terms in the brackets,

$(2K_{2a}v_{a2} + 3K_{3a}v_{a1}^2)$ and $(2K_{2a}v_{a1})$, are formulated by nonlinear coefficients and circuit

response. They are entries in NTM $N_{G.3.1}$ and $N_{G.3.2}$, respectively.

The first term in the RHS of Eq. (3.22) explicitly shows how h_j influences $(W_{G.n})_I$ in two

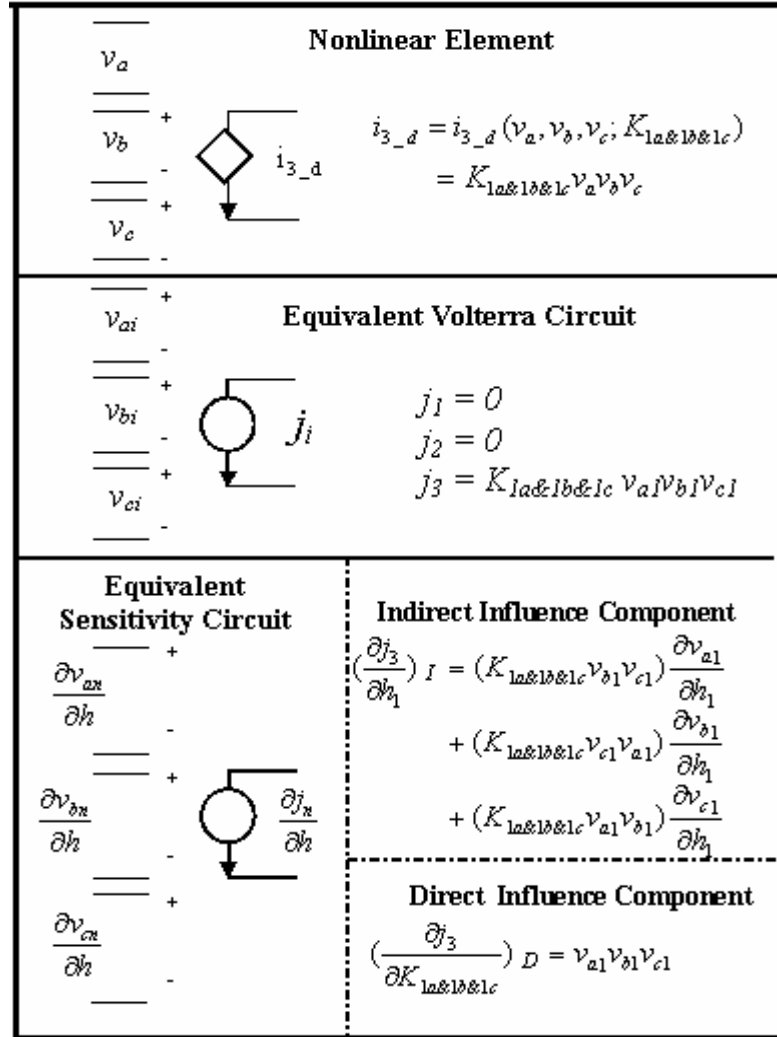steps: firstly, h_j influences circuit response X_k , represented by dX_k / dh_j ; the circuit response

37

Figure.3.3 Equivalent Volterra series and sensitivity circuits of 3-D transconductance

then influences the equivalent source through $N_{G.n.k}$. $N_{G.n.k}$ ($N_{C.n.k}$) thus represents how the k-th order circuit response generates the frequency independent (dependent) part of n-th order equivalent source $W_{G.n}$ ($W_{C.n}$). The analysis so far is performed in time domain. However, in the developed simulator, the above algorithms are realized in the frequency domain. The symbol $F(\cdot)$ is used to represent the Fourier transform. Since multiplication in time domain is transformed to convolution in frequency domain, the formulation of NTM in computer programming can be easily derived based on the time domain descriptions in Table 3.1 and Table

3.2. After performing a Fourier transform on both sides of Eq. (3.22), the following expression results:

$$F\left[(\frac{\partial W_n}{\partial h_j})_I\right] = \sum_{k=j}^{n-1}\left[F(N_{G.n.k}) \otimes F(\frac{\partial X_k}{\partial h_j})\right] + s \cdot \sum_{k=j}^{n-1}\left[F(N_{C.n.k}) \otimes F(\frac{\partial X_k}{\partial h_j})\right], \qquad (3.23)$$

Eq. (3.9), (3.17), (3.18) and (3.23) together constitute the NTM based sensitivity calculation in the frequency domain. With NTM, the formulation of RHS in Eq. (3.9) is now simplified to a few matrix-vector operations, which have the same size as the system equation. The computational cost of $RHS_{n.j}$ is thus reduced to the same level as system equation solution. Examination of Fig. 3.1- Fig. 3.3, Table 3.1 and Table 3.2 also shows the similarity of the entries in W_n and NTM. This implies NTM and W_n can be formulated together. Besides, generally, the number of transistors connected to one node is between one and three. This means, for CMOS

Table.3.1 Entries in N_{G2}

Type	Position	Value
1-d Trans	$e_S e_a{}^T$	$2K_{2a}v_{a1}$
2-d Trans	$e_S e_a{}^T$	$K_{1a\&1b}v_{b1}$
	$e_S e_b{}^T$	$K_{1a\&1b}v_{a1}$

Table.3.2 Entries in N_{G3}

Type	Position	Value
1-d Trans	$e_S e_a{}^T$	$2K_{2a}v_{a2} + 3K_{3a}v_{a1}{}^2$
2-d Trans	$e_S e_a{}^T$	$K_{1a\&1b}v_{b2} + 2K_{2a\&1b}v_{a1}v_{b1} + K_{1a\&2b}v_{b1}{}^2$
	$e_S e_b{}^T$	$K_{1a\&1b}v_{a2} + 2K_{1a\&2b}v_{a1}v_{b1} + K_{2a\&1b}v_{a1}{}^2$
3-d Trans	$e_S e_a{}^T$	$K_{1a\&1b\&1c}v_{b1}v_{c1}$
	$e_S e_b{}^T$	$K_{1a\&1b\&1c}v_{c1}v_{a1}$
	$e_S e_c{}^T$	$K_{1a\&1b\&1c}v_{a1}v_{b1}$

circuits, there are usually between 4 to 10 nonzero entries in each row in NTM. The NTM is thus sparse for circuits of medium to large size.

Consider the example of a single common source connected MOS transistor. Suppose both I/V and Q/V nonlinearities are considered and one is interested in the sensitivity of HD_3 w.r.t. g_m. Then it takes 492 multiplications to formulate $(\partial W_3 / \partial g_m)$ in the Adjoint method sensitivity calculation. With the introduction of NTM, the number of multiplications is reduced to 48. Greater efficiency improvement is possible in higher order Volterra analysis, multiple tone excitation or larger circuits, as more computation can be saved through the pre-computation of NTM.

Next, the physical meaning of NTM is examined. By definition, $N_{n.k}$ represents how X_k influences W_n, and thus X_n, from the $(n-k+1)$-th nonlinearity of the circuit. For example, the 5th element on the 4th row of $N_{G.3.1}$ shows how the 4th entry in W_{G3} is exclusively determined by the 5th entry in x_1, from the third order nonlinearity. Generally, the formulation in the form of Eq. (3.21) provides the insight into how lower order responses contribute to higher order response, from the nonlinearity of the circuit. Since nonlinearity is the intrinsic characteristic of the circuit, $N_{n.k}$ is solely determined by $(n-k+1)$, the difference between n and k, instead of their specific values. It can be proven:

$$N_{n.k} = N_{n+i.k+i}, \quad \forall n > k \geq 1 \,\&\, k+i \geq 1. \tag{3.24}$$

Thus, the two subscripts in the NTM can be reduced to one, e.g. $N_{n.k}$ can be simplified to N_{n-k+1}. For example, N_2 ($N_{3.2}$) shows how X_2 contributes to W_3 from the second order nonlinearity; similarly, N_2 ($N_{2.1}$) shows how the second order nonlinearity generates W_2 from X_1. This means only $2(N-1)$ NTM need to be pre-computed in N-th order distortion analysis. The speedup is achieved at the cost of the pre-computation and storage of the extra $2(N-1)$ matrices.

Numerical examples in the next section show, the overhead is negligible compared to the savings.

3.5   Numerical Examples

3.5.1 Nonlinear RC circuit

A simple example is used to illustrate the details in the sensitivity calculation based on NTM. The nonlinear RC circuit in Fig. 3.4 is used. The G and C matrices in Eq. (3.15) and (3.9) are:

$$G = \begin{bmatrix} g_1 & -g_1 & 1 \\ -g_1 & g_1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.25}$$

First consider distortion analysis, the RHS in Eq. (3.15) is given below. Notice for $n>1$, W_n has been decomposed based on frequency dependency.

$$W_1 = \begin{bmatrix} 0 \\ 0 \\ Ae^{j\omega t} \end{bmatrix}; \tag{3.26}$$

$$W_n = W_{Gn} + \frac{d}{dt}W_{cn}, \text{ for } n>1, \tag{3.27}$$

in which

$$W_{G.2} = \begin{bmatrix} -g_2 v_{a1}^{\,2} \\ g_2 v_{a1}^{\,2} \\ 0 \end{bmatrix} \text{ and } W_{C.2} = \begin{bmatrix} 0 \\ -c_2 v_{b1}^{\,2} \\ 0 \end{bmatrix}; \tag{3.28}$$
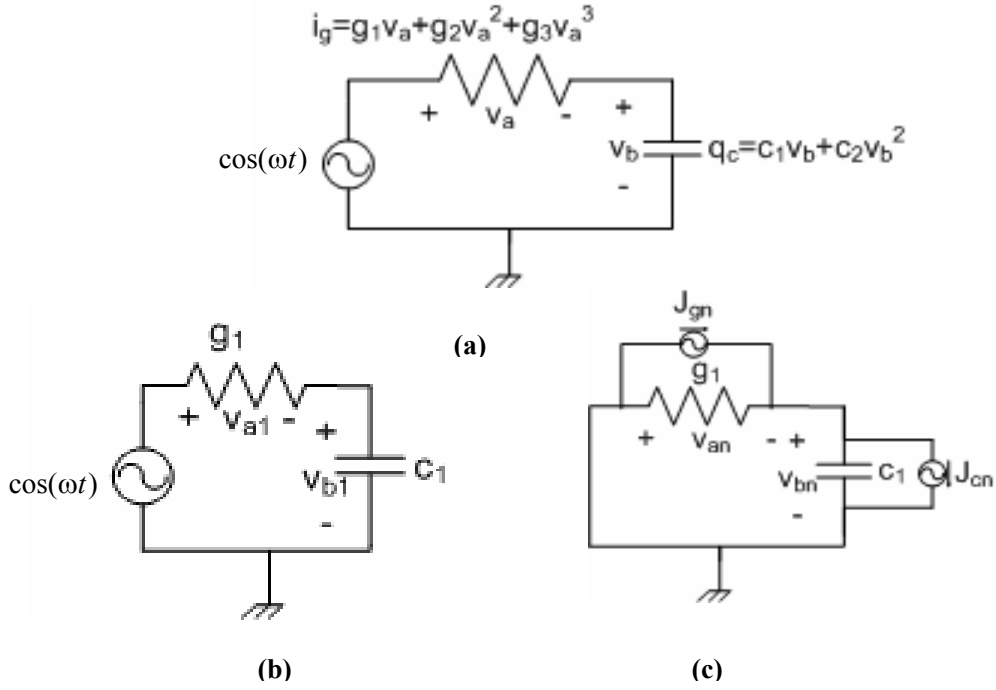
Figure3.4. (a). Nonlinear RC circuit. (b). First order equivalent circuit (c). *n*-th order equivalent circuit

$$W_{G.3} = \begin{bmatrix} -(2g_2 v_{a1} v_{a2} + g_3 v_{a1}{}^3) \\ 2g_2 v_{a1} v_{a2} + g_3 v_{a1}{}^3 \\ 0 \end{bmatrix} \text{ and } W_{C.3} = \begin{bmatrix} 0 \\ -2c_2 v_{b1} v_{b2} \\ 0 \end{bmatrix}. \tag{3.29}$$

The above equations are rewritten in a more compact forms, which reflects the formats used in computer implementation, as follows:

$$W_{G.2} = e_{sg} \cdot g_2 v_{a1}{}^2, \quad W_{C.2} = e_{sc} \cdot c_2 v_{b1}{}^2, \tag{3.30}$$

$$W_{G.3} = e_{sg} \cdot (2g_2 v_{a1} v_{a2} + g_3 v_{a1}{}^3) \text{ and } W_{C.3} = e_{sc} \cdot 2c_2 v_{b1} v_{b2}, \tag{3.31}$$

in which

$$v_a = e_a{}^T X \tag{3.32}$$

$$v_b = e_b{}^T X, \tag{3.33}$$

with

$$e_a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad e_b = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_{sg} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad e_{sc} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}. \tag{3.34}$$

Substituting the above equations into Eq. (3.15), after sequential solutions of system equations, circuit response X_1, X_2 and X_3 can be calculated. Next look at NTM based sensitivity calculation. Following Table 3.1 and Table 3.2, NTM are formulated based on the known circuit response and nonlinear coefficients as follows:

$$N_{G.2} = \frac{\partial W_{G.2}}{\partial X_1} = \frac{\partial W_{G.3}}{\partial X_2} = e_{sg} \cdot (2g_2 v_{a1}) \cdot e_a{}^T, \tag{3.35}$$

$$N_{C.2} = \frac{\partial W_{C.2}}{\partial X_1} = \frac{\partial W_{C.3}}{\partial X_2} = e_{sc} \cdot (2c_2 v_{b1}) \cdot e_b{}^T, \tag{3.36}$$

$$N_{G.3} = \frac{\partial W_{G.3}}{\partial X_1} = e_{sg} \cdot (2g_2 v_{a2} + 3g_3 v_{a1}{}^2) \cdot e_a{}^T, \tag{3.37}$$

$$N_{C.3} = \frac{\partial W_{C.3}}{\partial X_1} = e_{sc} \cdot (2c_2 v_{b2}) \cdot e_b{}^T. \tag{3.38}$$

Recall the formulation of $(dW_n / dh_j)_I$ is the computational bottleneck in sensitivity calculation Eq. (3.9), now with NTM available, $(dW_n / dh_j)_I$ can be computed from Eq. (3.23) efficiently. Notice Eq. (3.35)-(3.38) all take the form of sparse vector-scalar-vector multiplication. Because the present circuit is very simple (one nonlinear conductance and one nonlinear capacitor), there is only one term in each NTM. However, in a practical circuit, there are a large number of nonlinearities and the complexity to formulate NTM will greatly increase: Eq. (3.35)-(3.38) will grow into a summation of terms, with the number of terms equal to the number of corresponding nonlinearities. The pre-computation of NTM can prevent its repetitive evaluation in each

sensitivity calculation, and that is where the computation saving comes from, compared to the straightforward sensitivity calculation in Section 3.1.

Next examine other parts in the calculation of RHS of Eq.- (3.9). First look at "$-(dT/dh_j)X_n$", it is nonzero only when j=1. dT/dh_j is given in Eq. (3.39) and (3.40) for $h_1=g_1$ and $h_1=c_1$, respectively:

$$dT/dh_j = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{3.39}$$

$$dT/dh_j = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.40}$$

Next, $(dW_n/dh_j)_D$ is nonzero only when $j>1$, take the example of j=2 and n=3. For $h_2=g_2$, $(dW_n/dh_j)_D$ is

$$(dW_3/dh_2)_D = e_{sg} \cdot 2v_{a1}v_{a2}; \tag{3.41}$$

and for $h_2=c_2$,

$$(dW_3/dh_2)_D = \frac{d(e_{sc} \cdot (2v_{b1}v_{b2}))}{dt}. \tag{3.42}$$

Next consider the situation with Eq. (3.39)-(3.42) in a practical circuit with a large number of nonlinearities: if j=1, "$-(dT/dh_j)X_n$" will remain a sparse matrix-vector multiplication, like Eq. (3.39) and (3.40) [3]; otherwise, $(dW_n/dh_j)_D$ will remain a sparse vector like Eq. (3.41) and (3.42), as shown in Table 3.1 and Table 3.2. In both cases, the formulation is simple.

3.5.2 Cascode Amplifier

The Cascode Amplifier from Section 2.5 will be used to show the accuracy and usefulness of the proposed method in the sensitivity calculation of the distortion response to design, process or environmental parameters.

The expression for the normalized sensitivity of Third order Harmonic, HD_3, is defined as:

$$HD_3 = |\frac{V_{out}(3\omega_o)}{V_{out}(\omega_o)}| = |\frac{V_{out.3}(3\omega_o)}{V_{out.1}(\omega_o) + V_{out.3}(\omega_o)}| \qquad (3.43)$$

The normalized sensitivity of HD_3 with respect to h can be expressed as [3]:

$$S_h^{HD_3} = S_h^{|V_{out.3}(3\omega_o)|} - S_h^{|V_{out.1}(\omega_o) + V_{out.3}(\omega_o)|} . \qquad (3.44)$$

In Eq. (3.44),

$$S_h^{|V_{out.3}(3\omega_o)|} = h * \text{Re}(\frac{1}{V_{out.3}(3\omega_o)} \frac{\partial V_{out.3}(3\omega_o)}{\partial h}) \qquad (3.45)$$

and

$$S_h^{|V_{out.1}(\omega_o) + V_{out.3}(\omega_o)|} = h * \text{Re}[(\frac{1}{V_{out.1}(\omega_o) + V_{out.3}(\omega_o)}) \cdot (\frac{\partial V_{out.1}(\omega_o)}{\partial h} + \frac{\partial V_{out.3}(\omega_o)}{\partial h})] . \qquad (3.46)$$

Following the definition in Eq. (3.44), the normalized sensitivity of HD_3 w.r.t. all the nonlinear coefficients, at the nominal operating point, are given in Table 3.2. Based on that, the calculations of HD_3 due to temperature change from 265 K to 335 K and the input transistor width variation (over the whole range to ensure the transistor remains in saturation) are plotted in Fig. 3.6.

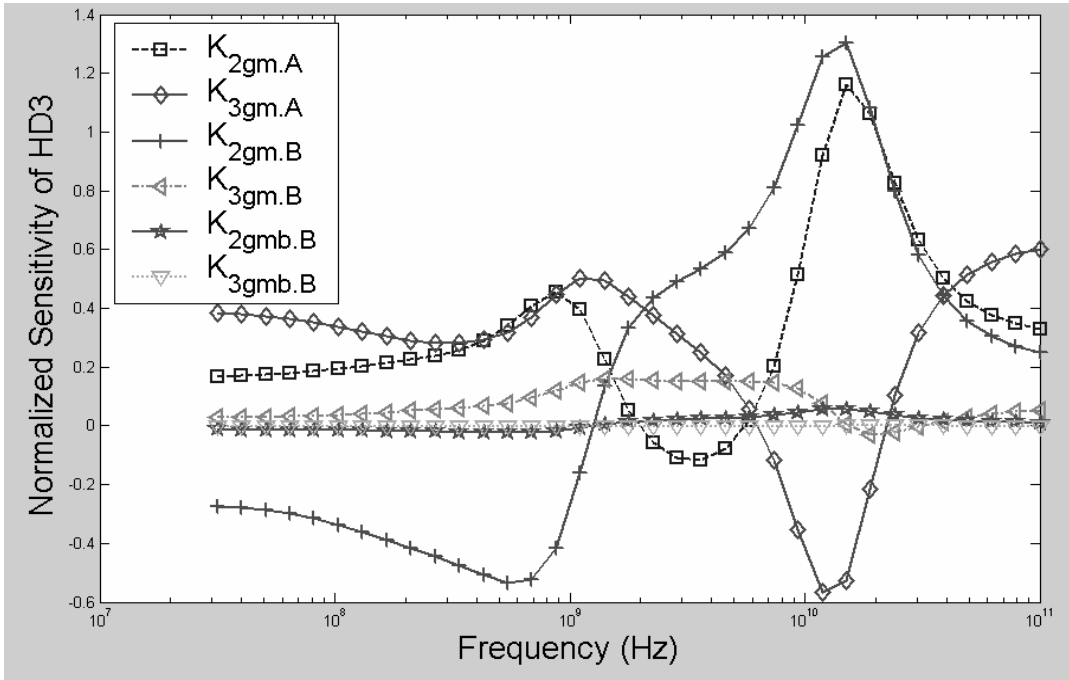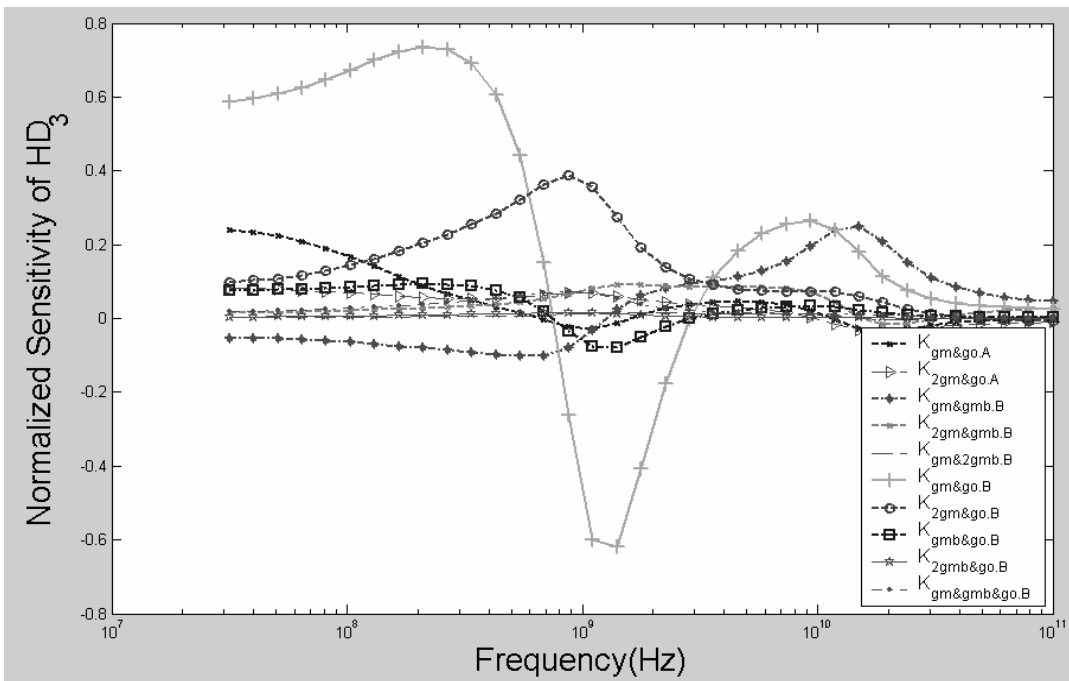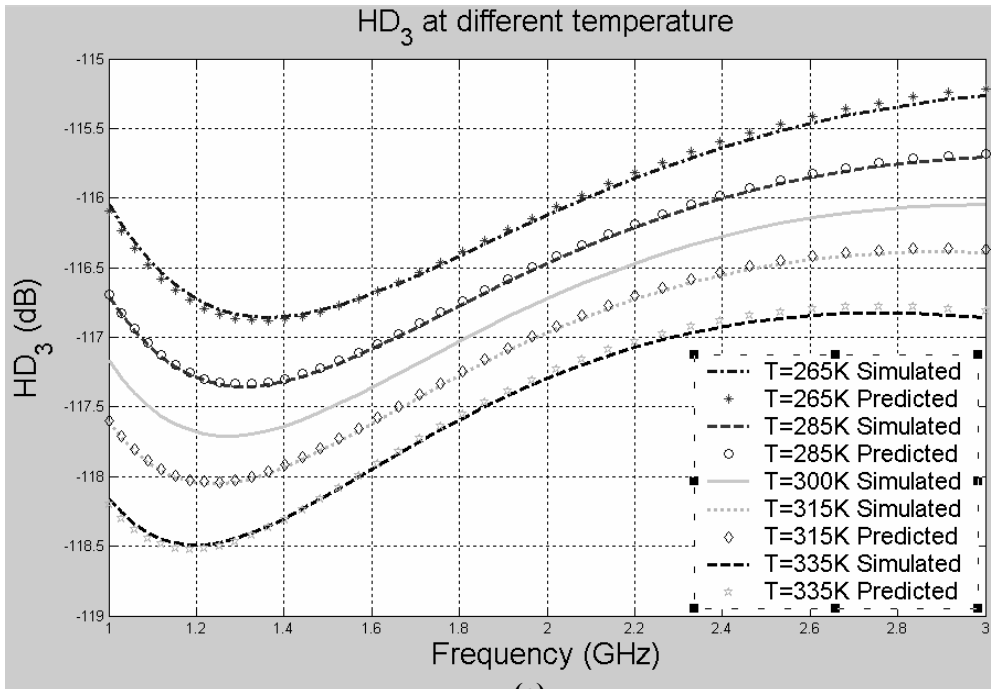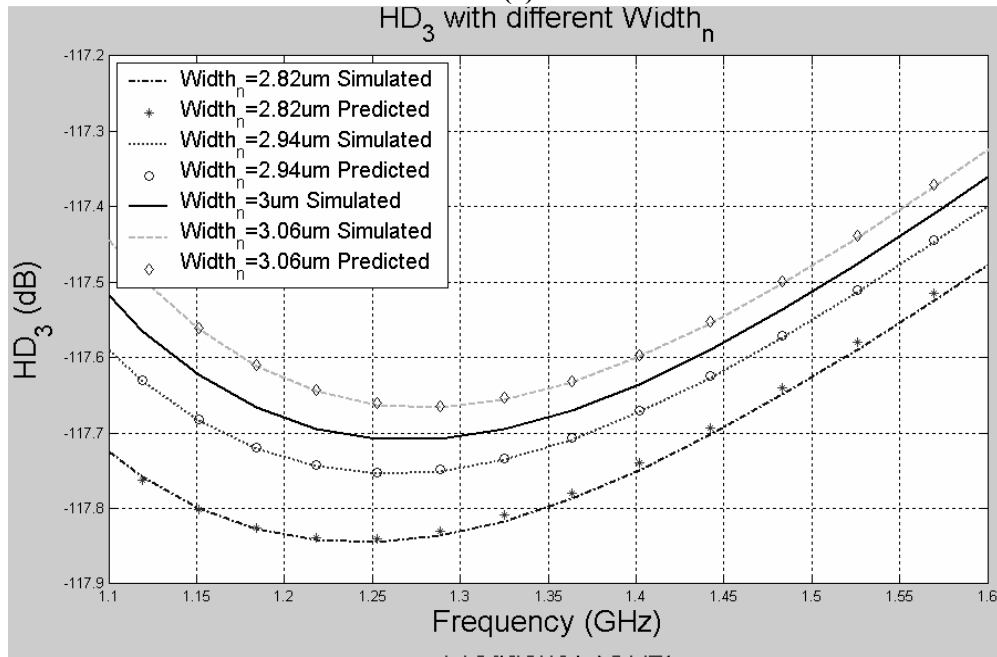**Figure. 3.5 (a) Sensitivity of _HD₃_ w.r.t. 1-D nonlinear coefficients**

Figure. 3.5 (b) Sensitivity of _HD₃_ w.r.t. 2-D and 3-D nonlinear coefficients

(a)

(b)

Figure. 3.6 Prediction of HD_3 due to temperature and width_n change

3.5.3 Third Order Elliptical Filter

The third example is a third order elliptical low pass filter [30] with the Op-Amp realized by the low distortion two stage folded cascode Op-Amp in Fig. 4.4. It contains 104 MOS transistors and the full BSIM3V3 model is adopted. Generally, NTM is expected to bring more computational saving when the analysis order is high or when there are more than one driving frequency. However, only a relatively low order – third order – analysis, in one-tone test is considered here. This demonstrates even for lower order analyses with one input frequency, NTM can still bring significant computational saving.

In total, there are 2191 small signal coefficients, 3088 second and 4908 third order nonlinear coefficients. On a P4 computer with 1.6 GHz CPU and 768 MB RAM, one Volterra Series based frequency analysis takes 1.18s and it takes 6 hours and 41 minutes to approximate sensitivities of HD_3 w.r.t. of all the 10187 coefficients by finite difference. The NTM based numerical sensitivity calculation accomplishes the same work in 3 minutes and 27 seconds. It takes 10 seconds to construct NTM. The overhead in storage is a negligible 103 kB. On average, accurate sensitivity calculation is achieved with over 100 times speedup, compared to finite difference approximation. The time and speedup for the sensitivity calculation of different order coefficients are given in Table 3.3.

The above examples show that NTM based nonlinear circuit sensitivity calculation provides both accuracy and efficiency in computation.

Table. 3.3 Time and speedup of NTM based sensitivity calculation

Sensitivity	Time(ms)	Speedup (times)
$\partial X_3 / \partial h_1$	53	45
$\partial X_3 / \partial h_2$	25.5	93
$\partial X_3 / \partial h_3$	2.6	908

Chapter 4

Per-Element Distortion Decomposition

In this chapter, a method to decompose distortion on a per nonlinear element basis is given [48]. It can be used in design optimization, symbolic analysis and nonlinear model reduction. Fully symbolic [8,38,39] and polynomial interpolation [3,7] are the common distortion decomposition methods available so far. However, their efficiency and accuracy is limited since they use full or partial symbolic analysis.

Decomposition based on previous methods is normally made possible by using simplified transistor models [36,42,43], however, even the most popular compact models are inadequate for high frequency (HF) distortion analysis [37] and an accurate HF MOSFET model is essential in distortion analysis. This Chapter proposes a distortion decomposition technique based on Nonlinearity Transfer Matrix, it combines the insight of traditional symbolic analysis, and the handling capability, efficiency and accuracy of commercial numerical simulators.

Section 4.1 briefly reviews the bottleneck in VS based distortion analysis--the complicated mixing effects. Section 4.2 takes a simple nonlinear RLC circuit and performs the distortion decomposition. With the help of examples, section 4.3 demonstrates that the use of NTM leads to an efficient and accurate numerical distortion decomposition method.

4.1　Motivation and Challenge

Based on the equivalent circuit of each nonlinear element as given in Fig. 2.1 to Fig. 2.3, the i-th order equivalent circuit can be formulated as in Eq. (2.24). Frequency response can then be calculated based on VS following the procedure in Section 2.2. This procedure provides the overall nonlinear response. However, it is sometimes desirable for designers to obtain a table, similar to noise and sensitivity analysis, listing the detailed distortion contributions from different nonlinear elements, including both amplitude and phase information.

The desired distortion decomposition technique starts from the original Spice level netlist. Instead of simplified models, it uses the most reliable and, thus, usually complex compact transistor models. After distortion decomposition, a table that gives the contribution from each coefficient is obtained. The dominant entries in this table offer the insight into the origin of distortion. The amplitude and phase information can be used in interactive design optimization. The insignificant entries identify negligible coefficients that can be pruned to simplify the model. The best tradeoff between accuracy and compactness can thus be achieved in symbolic simplification, which is required in symbolic distortion analysis [7]. Based on this, [7] presents a generalization of compact nonlinear modeling for mixed signal co-simulation and system level design, e.g. telecom front-ends.

The ever increasing complexity of RFIC and transistor models requires efficient computation, accuracy and larger handling capability. In general, nonlinear coefficient h_j influences W_n, $n \geq j$, in two ways. Firstly, h_j appears directly in n-th order equivalent circuit. For example, in MNA, h_j appears in at most two locations of W_n; for each entry, h_j contributes only one term. Since this part of contribution is direct, it is called direct contribution and is easy to calculate. Secondly, if $n>j$, h_j also contributes to W_n indirectly by generating lower order circuit responses first, which then

contribute to all the entries in W_n by mixing with other nonlinear coefficients and circuit responses. It is this indirect contribution, resulting from Multiple Mixing Effect, that makes it a challenging task to determinate the contribution from lower order nonlinearity to higher order distortion.

4.2   Per-Element Distortion Decomposition

In this section, a simple example is used to show how numerical distortion decomposition can be achieved. The nonlinear RLC circuit in Fig. 4.1 is used and the object is to:

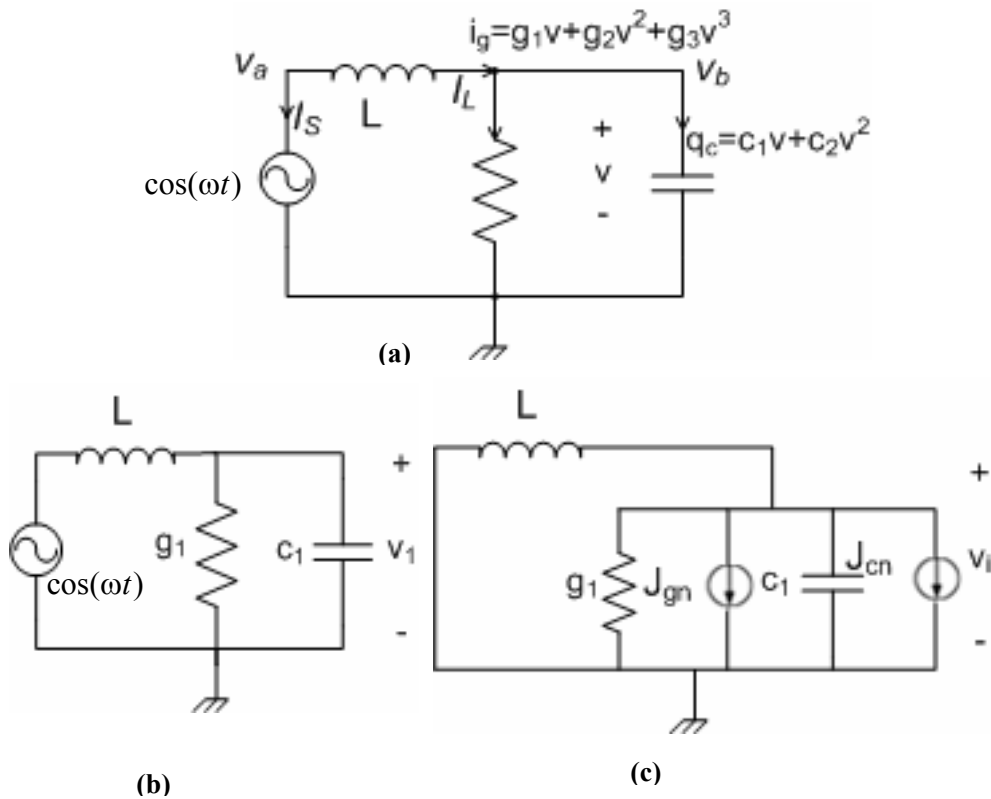(a) Decompose the contributions to X_2 from $h=\{g_2,c_2\}$.

Figure 4.1. (a) Nonlinear RLC circuit. (b) First order equivalent circuit (c) _n_-th order equivalent circuit

(b) Based on (a) and the solution of Eq. (2.24), decompose the contributions to X_3 from $h=\{g_2,c_2,g_3\}$.

The variables in Eq. (2.24), formulated by MNA, are given as follows:

$$G = \begin{bmatrix} & 1 & 1 & \\ & g_1 & -1 & \\ 1 & -1 & & \\ 1 & & & \end{bmatrix}; \quad C = \begin{bmatrix} 0 & & & \\ & c_1 & & \\ & & -L & \\ & & & 0 \end{bmatrix}; \quad X = \begin{bmatrix} v_a \\ v_b \\ I_L \\ I_S \end{bmatrix};$$

$$W_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ A\cos(\omega t) \end{bmatrix}; \quad W_n = \begin{bmatrix} 0 \\ -J_n \\ 0 \\ 0 \end{bmatrix};$$

in which

$$J_n = J_{cn} + J_{gn},$$

$$v_a = e_a^T X, \ v_b = e_b^T X, \ W_n = J_n \cdot e_s \ \text{ with } \ e_s = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, e_a = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, e_b = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

In this chapter, subscript "hj" is added to variables to represent the contribution from nonlinearity h_j. Symbol prime denotes "complement", e.g. $W'_{3.g2}$ is the contribution to W_3 from all nonlinearities except g_2. Start with the decomposition of X_2 with respect to $h=\{g_2,c_2\}$. Based on the above definitions, for $h=g_2$:

$$W_2 = W_{2.g2} + W'_{2.g2}, \tag{4.1}$$

$$X_2 = X_{2.g2} + X'_{2.g2}. \tag{4.2}$$

Substitute Eq. (4.1) and (4.2) into Eq. (2.24), the application of superposition for linear system leads to Eq. (4.3). Eq. (4.3) shows that decomposition of X_n, in general, can be transformed to the decomposition of W_n

$$T \cdot X_{2,g2} = W_{2,g2} = e_s \cdot J_{2,g2},$$
(4.3)

In which

$$J_{2.g2} = J'_{2.c2} = g_2 \cdot v_{b1}^2.$$
(4.4)

Similarly, the expression for $J_{2.c2}$ can be derived:

$$J_{2.c2} = J'_{2.g2} = \frac{d}{dt}[c_2 \cdot v_{b1}^2].$$
(4.5)

Substitute $J_{2.g2}$ in Eq. (4.3) with $J_{2.c2}$ from Eq. (4.5), $X_{2.c2}$ can also be calculated. Based on Eq. (4.2-4.5), it can easily be shown that, in the above decomposition, no overlap exists between $X_{2.c2}$ and $X_{2.g2}$, besides, their sum equals X_2.   As can be shown, this kind of decomposition can always be achieved if the distortion response comes from the same order nonlinearities only, e.g. $X_{n.n}$ with $n>1$. In other words, only direct contribution exists. Besides, since there is no mixing effect, the associated computation cost is low. Next, move to $X_{3.g2}$. As discussed in Section 4.1, for higher order distortion, mixing effects combine the contributions from different lower order nonlinearities. In addition, the mixing effect complicates the formulation of the equivalent sources, and is the computation bottleneck in VS analysis of nonlinear circuit. In the next section it is shown that NTM reduces this bottleneck. Rewrite Eq. (4.3) for X_3 and expand the expression for J_3:

$$T \cdot X_{3,g2} = W_{3,g2} = e_s \cdot J_{3.g2},$$
(4.6)

$$J_3 = 2g_2 v_{b1}(v_{b2.g2} + v'_{b2.g2}) + \frac{d[2c_2 v_{b1}(v_{b2.c2} + v'_{b2.c2})]}{dt} + g_3 v_{b1}^3.$$
(4.7)

Because of mixing effects, decomposition gets much more complicated and different approaches exist:

(1) Direct Contribution Decomposition

Let the nonlinearity of interest be h, and set all other variables and coefficients independent of h to zero, then the contribution from h can be determined. This ignores mixing effects. For example, by setting all the terms in Eq. (4.7) unrelated to g_2 to zero, $J_{3.g2}$ can be identified:

$$J_{3.g2} = 2g_2 v_{b1} v_{b2.g2} . \tag{4.8a}$$

Similarly, $J_{3.c2}$ and $J_{3.g3}$ can be derived:

$$J_{3.c2} = \frac{d(2c_2 v_{b1} v_{b2.c2})}{dt} , \tag{4.8b}$$

$$J_{3.g3} = g_3 v_{b1}{}^3 . \tag{4.8c}$$

Comparison with Eq. (4.7) shows there is still no overlap between different contributions in Eq. (4.8); but the decomposed contributions underestimate the total distortion: the sum of Eq. (4.8) contains 3 terms, compared to 5 in the total distortion of Eq. (4.7). Two terms in Eq. (4.7), $2g_2 v_{b1} v_{b2.c2}$ and $d(2c_2 v_{b1} v_{b2.g2})/dt$, are missing. This is caused by neglecting the mixing effects, e.g. $2g_2 v_{b1} v_{b2.c2}$ is the combined effects from both g_2 and c_2 .

(2) Per-Nonlinearity Decomposition

The inclusion of the mixing effects leads to the second approach: Per-Nonlinearity decomposition. For $J_{3.g2}$, let $h = g_2$, prune all the terms in Eq. (4.7) unless it contains either g_2 or

$v_{2.g2}$. The same procedure leads to $J_{3.c2}$ and $J_{3.g3}$:

$$J_{3.g2} = 2g_2 v_{b1} v_{b2.g2} + \frac{d(2c_2 v_{b1} v_{b2.g2})}{dt} + 2g_2 v_{b1} v'_{b2.g2}, \quad (4.9a)$$

$$J_{3.c2} = 2g_2 v_{b1} v_{b2.c2} + \frac{d(2c_2 v_{b1} v_{b2.c2})}{dt} + \frac{d(2c_2 v_{b1} v'_{b2.c2})}{dt}, \quad (4.9b)$$

$$J_{3.g3} = g_3 v_{b1}{}^3. \quad (4.9c)$$

In Eq. (4.9a), the first two terms are the indirect contribution part and originate from lower order response $v_{2.g2}$; the last term is the direct contribution part and comes from g_2. Comparison of Eq. (4.9) with Eq. (4.7) shows that the sum of individual per-nonlinearity decomposition might overestimate the total distortion. Recall Eq. (4.4-4.5), the third term of Eq. (4.9a) is the same as the first term of Eq. (4.9b) and the third term of Eq. (4.9b) is the same as the first term of Eq. (4.9a). The sum of Eq. (4.9) thus contains 7 terms, compared to 5 in Eq. (4.7). Since mixing effects represents the interaction between different nonlinearities, its inclusion in each relevant entry —derived from physical basis—inevitably introduces the overestimation in Per-Nonlinearity Decomposition.

4.3 Numerical Implementation by Nonlinearity Transfer Matrix

The previous section shows that there are different decomposition approaches based on the inclusion or exclusion of the mixing effect. "Direct contribution" is computationally cheap but neglects mixing effects. "Per-Nonlinearity" includes indirect contribution from mixing effects, but may overestimate the results. Next, consider circuits of practical complexity and deal with the efficiency issue. Because of its physical insight and more computational cost requirement, Per-Nonlinearity decomposition is used for illustration. The BSIM3V3 model contains 88

nonlinear coefficients in third order distortion analysis. For a circuit of practical size with common compact transistor models, the number of 2nd order nonlinearities easily grows to thousands or more. They constitute the indirect contribution component in $W_{3.h2}$. Specifically, the direct contribution part, the last term in each equation of (4.9), remains the same if extended to practical circuit; while the indirect contribution will grow from the first two terms in Eq. (4.9a,b) to an enormously complex expression. However, careful observation of the indirect contribution components in Eq. (4.9a,b) shows similarity in their expressions. In order to use this similarity to save computation, considering $W_n = j_n e_s$, rewrite Eq. (4.9a,b) and after some term rearrangements:

$$W_{3.g2} = [(2g_2 v_{b1})(e_s e_b{}^T)] \cdot v_{2.g2} + \frac{d\{[(2c_2 v_{b1})(e_s e_b{}^T)] \cdot v_{2.g2}\}}{dt} + (2g_2 v_{b1} v_{b2.g2}')e_s, \qquad (4.10a)$$

$$W_{3.c2} = [(2g_2 v_{b1})(e_s e_b{}^T)] \cdot v_{2.c2} + \frac{d\{[(2c_2 v_{b1})(e_s e_b{}^T)] \cdot v_{2.c2}\}}{dt} + \frac{d((2c_2 v_{b1} v_{b2.c2}')e_s)}{dt}. \qquad (4.10b)$$

In Eq. (4.10a-4.10b), the similarity is explicitly revealed: the common part is enclosed in "[]". The expression in the first "[]" comes from mixing effect due to g_2 and the part in the second "[]" comes from mixing effect of c_2. Besides, each term in "[]" is expressed in the form of scalar-matrix multiplication. The scalar is formulated by first order response and second order nonlinearities. The matrix is sparse and has at most 4 nonzero entries for MNA. Eq. (4.10a-4.10b) can be rewritten as:

$$W_{3.h2} = N_{G.2} \cdot X_{2.h2} + \frac{d(N_{C.2} \cdot X_{2.h2})}{dt} + (W_{3.h2})_D. \qquad (4.11)$$

Eq. (4.11) is the generalized decomposition expression. In practical circuits, $(W_{3.2h})_D$, the direct contribution, contains only one term, the same as Eq. (4.10a-4.10b). In computer programming, $(W_{3.2h})_D$ has at most two nonzero entries and can be easily derived. The careful arrangement of terms will result in regularity in both $N_{G.2}$ and $N_{C.2}$: the summation of many terms, each has the same structure as the term in "[]" of Eq. (4.10). $N_{G.2}$ and $N_{C.2}$ are called Nonlinearity

57

Transfer Matrices (NTM). In NTM, the first subscript, G or C, depends on whether NTM is formulated by frequency independent or dependent nonlinearities; the second subscript is the order of nonlinearity. The NTM for the simple nonlinear RLC example are:

$$N_{G.2} = \begin{bmatrix} 0 & & \\ & 2g_2v_{b1} & \\ & & 0 \\ & & & 0 \end{bmatrix}; \qquad N_{C.2} = \begin{bmatrix} 0 & & \\ & 2c_2v_{b1} & \\ & & 0 \\ & & & 0 \end{bmatrix}.$$

For example, $N_{G.2}$ represents how X_2 contributes to the frequency independent part of W_3 from the second order nonlinearity. The details about the definitions of $N_{G.k}$ and $N_{C.k}$ are covered in Section 3.4 and rewritten below. $W_{C.i+k-1}$ and $W_{G.i+k-1}$ are the frequency dependent and independent part of W_{n+k-1}, respectively. Notice the following definitions are given in time domain.

$$N_{G.k} = \frac{\partial W_{G.i+k-1}}{\partial X_i}, k > 1 \ \& \ \forall i \geq 1. \tag{3.19}$$

$$N_{C.k} = \frac{\partial W_{C.i+k-1}}{\partial X_i}, k > 1 \ \& \ \forall i \geq 1. \tag{3.20}$$

As the name implies, NTM represents the complicated mixing effect of how i-th order circuit response X_i indirectly influences $(i+k-1)$-th order equivalent source W_{i+k-1}, from k-th order nonlinearity of the circuit. The formulation of NTM is complicated because it includes all the complexity of multiple mixing effects. However, since nonlinearities are intrinsic characteristics of the circuit, NTM are constant matrixes. NTM is composed of nonlinear coefficients and distortion response of the circuit, which are all determined for a design. The formulation of NTM is thus possible and straightforward. It needs to be done only once for the distortion decomposition of a fixed design at each frequency. Details about its computer programming implementation are given in Table 3.1 and Table 3.2.

$$W_{n.hj} = \sum_{k=2}^{n-j+1} N_{G.k} \cdot X_{n-k+1.hj} + \frac{d(\sum_{k=2}^{n-j+1} N_{C.k} \cdot X_{n-k+1.hj})}{dt} + (W_{n.hj})_D \qquad (4.12)$$

$$T \cdot X_{n,hj} = W_{n,hj} \ , \ n \geq j \geq 2 \qquad (4.13)$$

The sequential solution of Eq. (4.12-4.13) in the increasing order of n forms the generalized "Per-Nonlinearity" distortion decomposition of contributions to n-th order distortion X_n from j-th order nonlinearity h_j. Notice Eq. (4.12) is described in time domain, although the corresponding frequency domain operations are performed in the simulator. If only the last term in Eq. (4.12) is considered, they degenerate into "Direct Contribution" decomposition. The overhead in the pre-computation of NTM is averaged out in its repeated use in the distortion decomposition process, considering the large number of nonlinear coefficients in a practical circuit. Eq. (4.12) shows the use of NTM transforms the bottleneck in traditional VS based nonlinearity analysis, the construction of equivalent source, to cheap sparse matrix-vector operations. Besides, notice Eq. (4.13) is solved repetitively with different right hand sides; instead of the whole vector $X_{n,hj}$, only a few of its entries are of interest, e.g. certain nodal voltage or branch current. These two points suggest further reduction in the linear system solution Eq. (4.13) is possible through sharing LU factors and the use of Adjoint Method.

Finally, it is worth mentioning that there are two kinds of sensitivities: differential and large change sensitivity. Differential sensitivity is defined as the change in output due to infinitesimal change in parameters; while large change sensitivity measures the variation in output when the parameters are subjected to large variations. Specifically, this can be how the inclusion or deletion of some nonlinearity influences the distortion response of the whole system. Distortion decomposition is thus essentially large change sensitivity of nonlinear coefficients. Nonlinearity Transfer Matrix was first proposed in Section 3.3 to overcome the complexity of "multiple mixing effects" and improve the efficiency in differential sensitivity calculation. The similar physical

origin naturally leads to the application of NTM to solve "multiple mixing effects" in distortion decomposition.

4.4   Numerical Examples

4.4.1 5.8Ghz Folded Cascode LNA

In this section a low distortion 5.8GHz folded-cascode LNA [8], shown in Fig. 4.2, is used as an example. Fig. 4.3 plots IM_3, the Third-order Intermodulation over the 3dB band. With full BSIM3V3 model, it takes a P4 computer 90 ms to perform per-nonlinearity distortion decomposition at one frequency point.
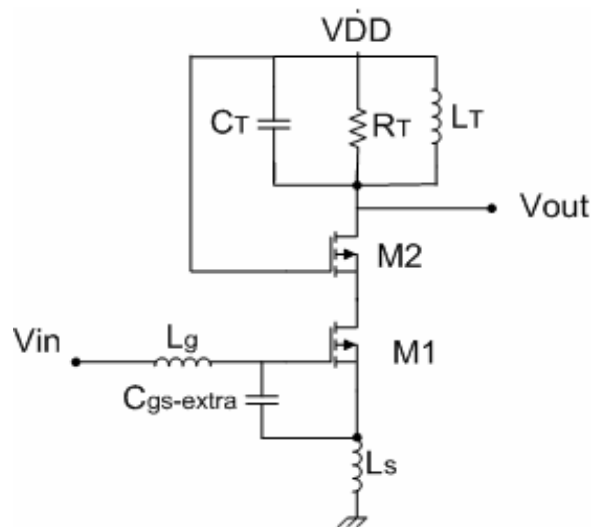

Fig.4.2   5.8GHz folded cascode LNA

Table 4.1summarizes the list of Most Important Contributor (MIC) versus All Contributions

(AC), depending on different error criterion. The first two columns, $\in_{avg}$ and $\in_{max}$, are the average and maximum errors over the whole frequency range. The error is calculated between total distortion of the original circuit including all the nonlinearities, and the approximation by keeping MIC only and pruning all the other nonlinear coefficients. The error reflects how well the

TABLE 4.1 MOST IMPORTANT CONTRIBUTIONS TO IM₃ OF LNA

$\in_{avg}$ (dB)	$\in_{max}$ (dB)	#MIC/#AC	MIC
0.74	1.31	1/176	$K_{m3\text{-}1}$
0.16	0.25	3/176	$K_{m3\text{-}1}$, $K_{m2\text{-}1}$, $K_{Qm3\text{-}1}$
0.08	0.10	4/176	$K_{m3\text{-}1}$, $K_{m2\text{-}1}$ $K_{Qm2\text{-}1}$, $K_{Qm3\text{-}1}$

identified MIC can be used to represent AC in terms of distortion.

Based on Table 4.1, we can create different abstraction levels easily by trading off between accuracy and complexity. For higher accuracy the MIC in the second row of Table 4.1 can be chosen. To capture the third-order nonlinear behavior we have to take into account only three contributions. These three contributions—about 1.7% of the overall quantity of contributions—cover almost 98% of the third-order nonlinear behavior in the observed frequency range. In case compactness is preferred, MIC in the first row can be chosen. It includes only one contributor, $K_{gm3\text{-}1}$, the third order nonlinearity of transconductance of the input transistor. It represents 92% of the total distortion. This means, by targeting one nonlinearity only, designers can estimate IM_3 of the LNA correctly with a 1dB tolerance.

The above results validate the previous assumption that weakly nonlinear behavior is usually due to only a few important contributions. Such compactness and insight can greatly simplify the

analysis and optimization of nonlinear behavior for designers. In addition, it also makes it feasible

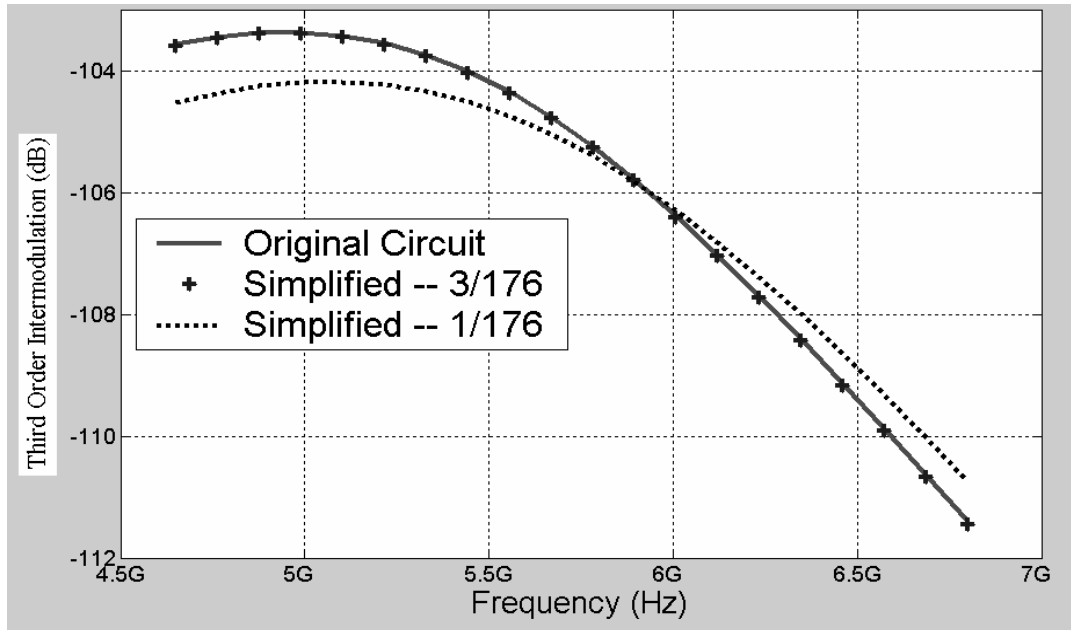to construct a compact and accurate high-level model.

Fig. 4.3 Third Order Intermodulation Plot

Table 4.1 also shows, interestingly, two of the top 4 MIC come from the nonlinearity of

channel charge w.r.t. V_{gs}. This is the first time that Q/V equations have been reported as the major

source of distortion in LNA study. This is because, this is the first work to use full compact model

in CMOS LNA distortion analysis. To achieve the same accuracy of 1dB error, 6 MIC is needed in

a similar design [7], compared to 1 here. Because of the simplified transistor model, 2-4dB error

already exists in the total distortion in [8] for the same design, even before any simplification is

performed.

4.4.2 Two-Stage Folded Cascode Op-Amp

Fig. 4.4 is a two-stage folded Cascode Op-Amp [9,43] in $0.18\,\mu m$ technology, BSIM3V3

model is used in the analysis again. Fig. 4.5 is the plot of third order harmonic distortion, HD_3, vs.

frequency in the whole application frequency range of the Op-Amp (0-150MHz). The transistor
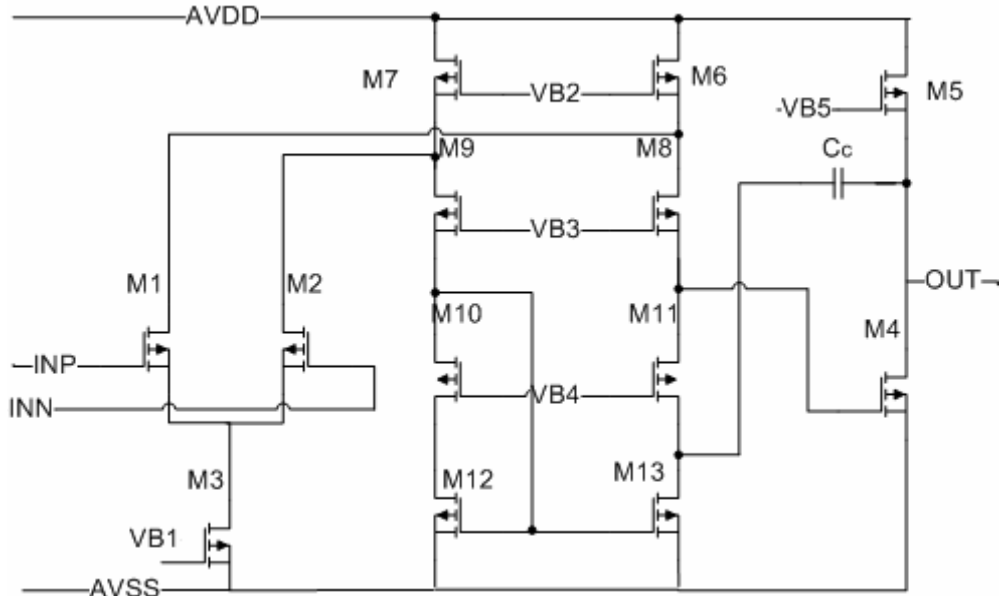

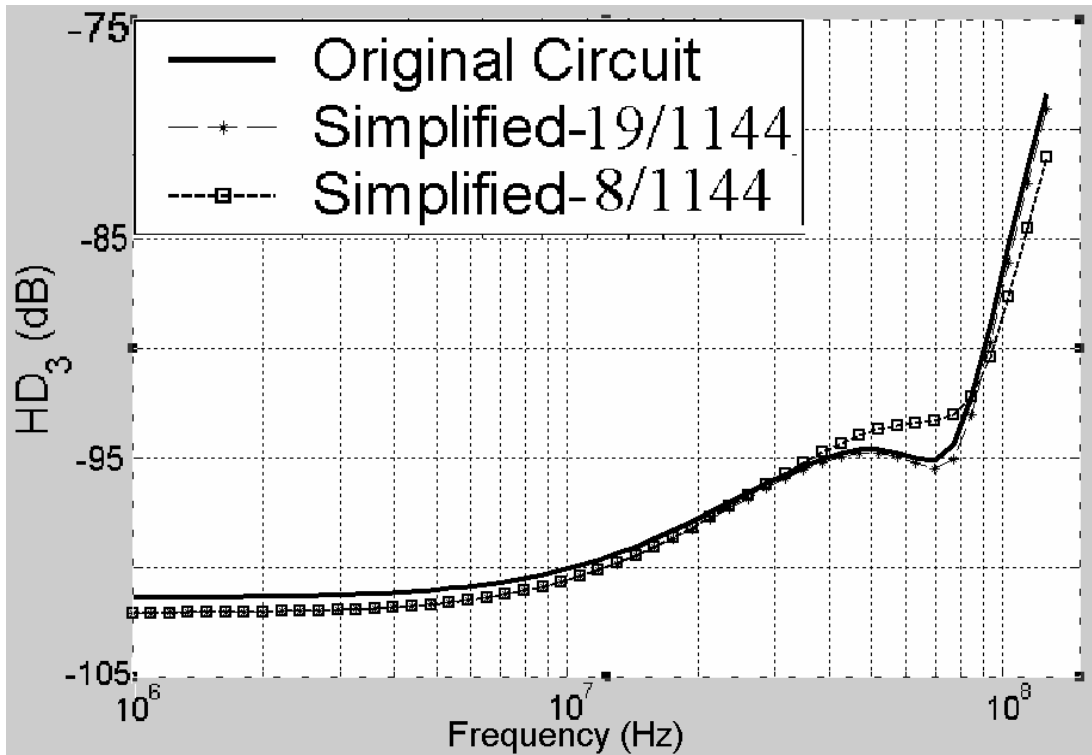Fig. 4.4. Two-stage Folded Cascode Miller Op-Amp.

level

Fig. 4.5. Third Order Harmonic Versus Frequency.

63

distortion decomposition is given in Fig. 4.6. It plots the normalized contributions to the total HD_3 from the three dominant nonlinearity sources. The plot shows, in the low frequency range (0-50 MHz), M4 and M5 in the output stage are the dominant distortion source; between 50 MHz and 100 MHz, contributions from input stage (M1+M2) and Cascode stage (M12+M13) start to get significant and even comparable to the output stage; beyond 100 MHz, contributions from both output and Cascode stages start to diminish and the input stage takes over to be the only dominant distortion source. The accuracy of the approximation, by representing the 1144 AC with the extracted MIC, is shown in Table 4.2. In [9], the "weakly nonlinear model" is used and the circuit includes 208 nonlinear coefficients. Average and maximum errors of 2 dB and 7 dB, respectively, exist in HD3 compared to simulation with full BSIM3V3 model [9]. In Fig. 4.5, by representing

TABLE 4.2 MOST IMPORTANT CONTRIBUTIONS TO HD3 OF OP AMP

Preference I: Higher Accuracy		
ϵ_{avg} =0.51 dB	ϵ_{max} =0.84 dB	#MIC/#AC=19/1144
Most Important Contributions		
K_{m2_1} $K_{m\&b\_1}$ K_{m2_2} K_{m3_2} $K_{2m\&b\_2}$ K_{m2_4} K_{m3_4} K_{d3_4} $K_{m\&d\_4}$ $K_{m\&2d\_4}$ K_{d3_5} K_{m3_12} $K_{m\&d\_12}$ $K_{2m\&d\_12}$ $K_{m\&2d\_12}$ K_{m3_13} K_{m2_13} $K_{2m\&d\_13}$ $K_{m\&2d\_13}$		

Preference II: More Compactness		
ϵ_{avg} =0.74 dB	ϵ_{max} =2.9 dB	#MIC/#AC=8/1144
Most Important Contributions		
K_{m2_1} $K_{m\&b\_1}$ K_{m2_4} K_{m3_4} K_{d3_4} $K_{m\&d\_4}$ $K_{m\&2d\_4}$ K_{d3_5}		

AC with the top 8 MIC, average and maximum errors of only 0.8 dB and 3 dB, respectively, are observed. Again, the improvement results from the accurate full models used.
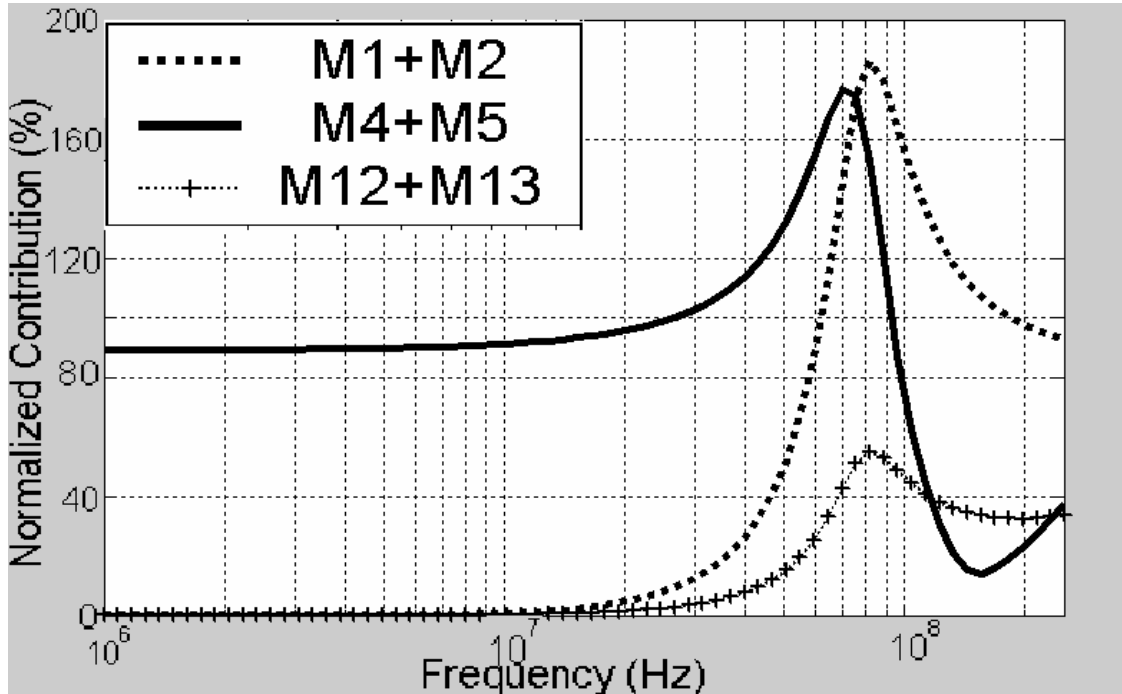

Fig. 4.6. Transistor Level Distortion Decomposition.

The extracted MIC provides significant physical insight for distortion mechanisms, in a very concise way. Besides, the selected MIC can also be used to generate compact and accurate symbolic expressions and behavioral models for Mixed signal co-simulation. Fig. 4.5 shows the local optimum of HD_3 appears at 70 MHz. Fig. 4.7 is the vector diagram of different distortion components at 70 MHz. It reveals the underlying cancellation mechanism due to opposite phase angles.

Finally, Fig. 4.8 shows the distortion decomposition at device level, at frequency points 10 MHz, 40 MHz, 70 MHz and 200 MHz. Since amplitude and phase relationships between different contributions, as well as their dependency on frequency, can be visualized, distortion origins can be pinpointed and exploited for further optimization. For example, Fig. 4.8(a) shows at low frequency, all the dominant nonlinearity comes from M4. This agrees with the transistor level
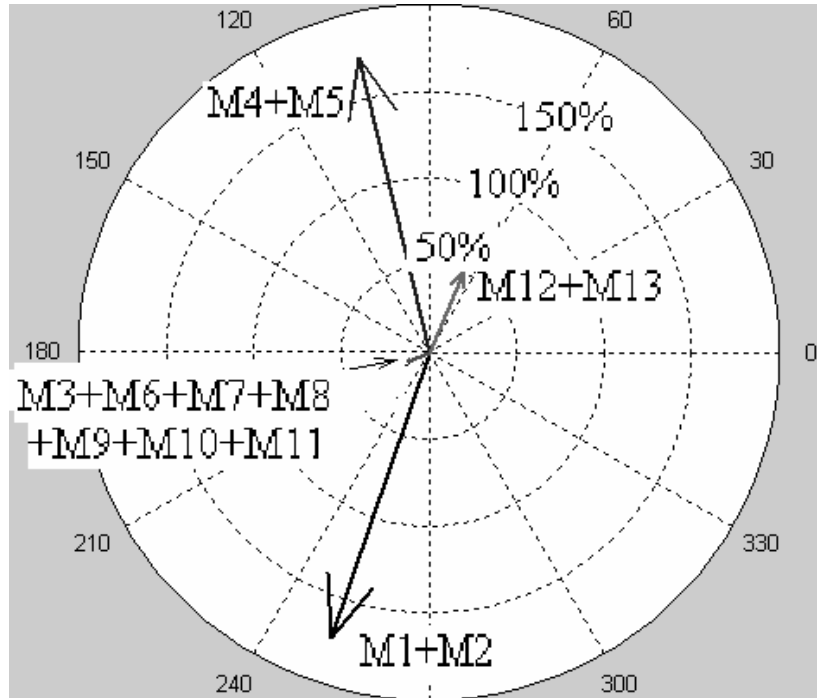
65

Fig. 4.7. Block Level Distortion Decomposition at 70MHz.

decomposition in Fig. 4.6, which shows the output stage contributes nearly 90% of the distortion in lower frequency band (0-20 MHz). Further, it shows most of the distortion in M4 comes from the nonlinearity of output impedance. This is because the voltage at the output node is the controlling variable for the nonlinear output impedance, and it undergoes the largest voltage swing. Fig. 4.8(a) shows, the second dominant nonlinearity source is the nonlinear transconductance of M4. This is because the nonlinear coefficients of transconductance are much larger than those of output impedance. For example, $K_{m3\_4} = -0.3871$ and $K_{d3\_4} = 0.0007$. The small signal analysis shows, at low frequency, the output stage is an inverting amplification stage with gain approximately equal to 11. This means $v_{gs\_4.1} / v_{ds\_4.1} \approx -11$. Remember K_{m3_4} and K_{d3_4} contribute the terms $K_{m3\_4} \cdot v_{gs\_4.1}^3$ and $K_{d3\_4} \cdot v_{ds\_4.1}^3$, respectively, to the equivalent source. Combining the above relationships, it can be estimated at low frequency the contribution from K_{d3_4} is around 2.4 times that of K_{m3_4} in amplitude, with the same phase angle. Referring to
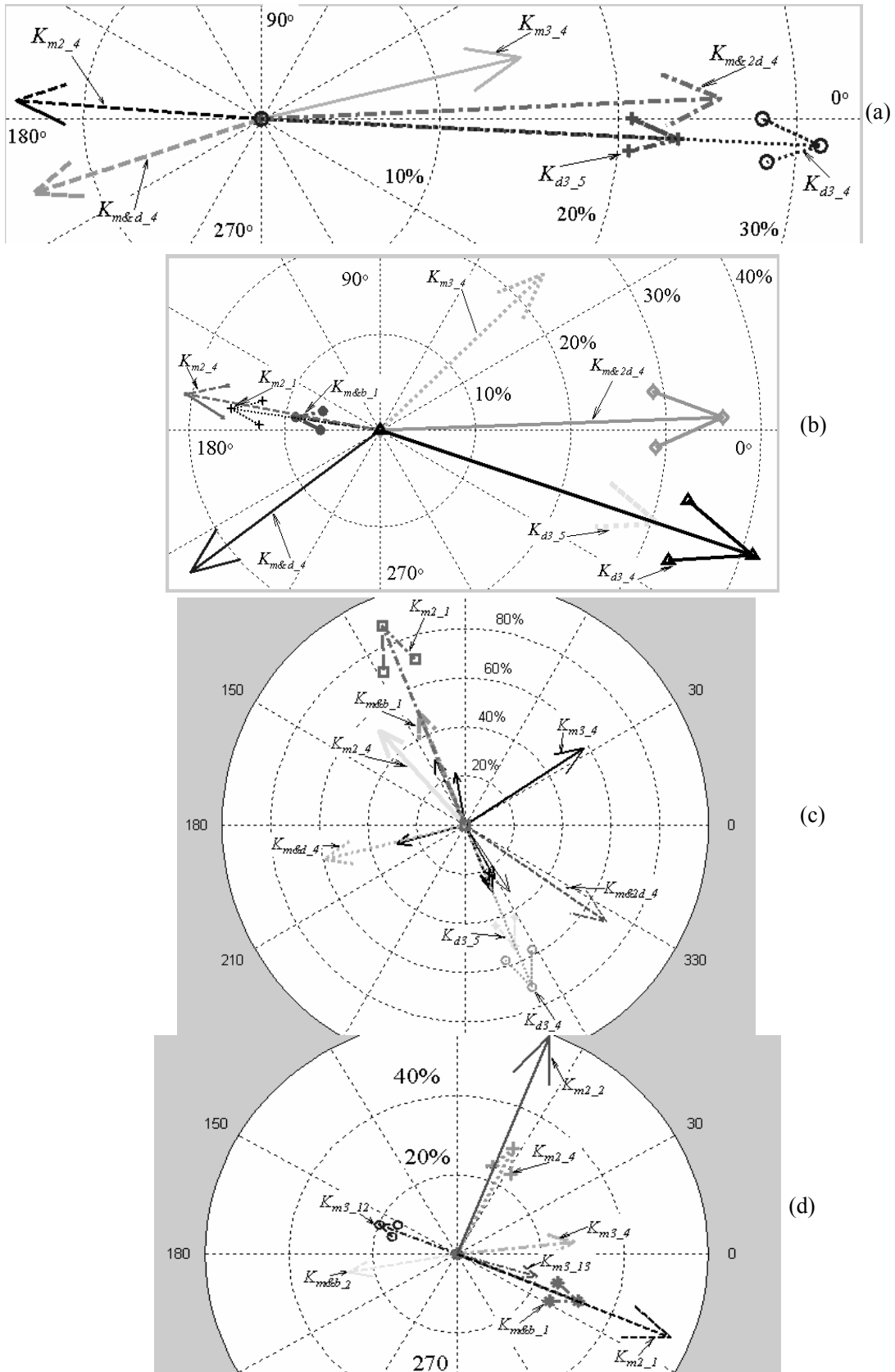
66

Fig.4.8 Two-stage folded cascode Miller opamp with single-ended output

4.8(a), the normalized contribution from K_{d3_4} is 32% while that from K_{m3_4} is 14%. This agrees very well with the estimation (32/14=2.3). Besides, there is a small phase angle difference (around 20°) between the two contributors, this is because at 10 MHz, frequency dependency has already started, especially due to the addition of the Miller capacitor, (the phase of the the output stage gain is -174°).

Comparing Fig. 4.8(a) and Fig. 4.8(b), it can be seen that because of the frequency dependency in the phases of different contributions, there is less cancellation due to opposite phase angles as frequency increases. This partly explains the increase in distortion up till 40 MHz. As frequency further increases to 70 MHz, another "optimum" cancellation point between the significant contributors is reached, as shown in Fig. 4.8(c). This explains the local optimum in Fig. 4.5 at 70 MHz. However, there are many other insignificant distortion contributors in the circuit, for example, those from the cascode stage. Their influence stands out when the significant contributors cancel each other. It is thus difficult to reach the "idealistic" optimum as some might expect based on simplified transistor models. Finally, as frequency increases to 200 MHz, the gain of the output stage drops to half its value at low frequency. The contribution from input stage starts to get dominant, consistent with the results in transistor level distortion decomposition Fig. 4.6.

Fig. 4.8 also illustrates that the influence of certain nonlinearities is limited to a certain frequency range. This suggests the possibility of neglecting coefficients that contribute significantly in a frequency band outside the region of interest. If instead of the whole operating frequency range (0-150 MHz), the Op-Amp is used only in low frequency band, then morecompact MIC set than those in Table 4.2 can be derived. For example, if 0-20 MHz band is of interest, then only the nonlinear coefficients from the output stage are needed.

4.4.3 Third Order Elliptical Filter

The third example is a third order elliptical low pass filter with the Op-Amp taken from Fig. 4.4. It contains 104 CMOS transistors, 280 nodes (including internal nodes in BSIM3V3 models). As shown in the last section, the higher order the nonlinearity, the more driving frequencies, the more computational intensive it is to perform distortion decomposition because of the complicated mixing effect. However, only a third order response in one-tone test is considered here. This will help demonstrate great computation cost saving can still be achieved even for lower order nonlinearity with one exciting frequencies benefiting from NTM. There are in total 3088 second and 4908 third order nonlinear coefficients. One Volterra Series based distortion analysis takes 1.18s and it takes 2 hours and 37 minutes to approximate contributions to HD3 from all the 7996 coefficients by the "brute force" method--finite difference; while it only takes 1 minutes and 32 seconds for the proposed distortion decomposition method. The overhead to calculate NTM is negligible: 3 seconds and 34kB memory for storage. An extra 20 seconds is required to calculate the nonlinear coefficients from DC simulation results. In terms of handling capability, the largest number of nonlinear coefficients that previous methods can handle is of the order of hundred [9], compared to 10,000 here. On average, accurate sensitivity result is achieved with an over 100 times speedup, as compared to crude approximation. The time per decomposition and speedup for the sensitivity calculation of different order coefficients are given in Table 4.3.

TABLE 4.3 EFFICIEINCY PERFORMANCE

Contributor	Time (ms)	Speedup (times)
$X_{3,h2}$	25.5	93
$X_{3,h3}$	2.6	908

Chapter 5

Conclusions

The main contributions of this thesis to the computer aided design of mildly nonlinear circuits are:

(1) A Volterra Series based sensitivity analysis method,

(2) Introduction of Nonlinear Transfer Matrix, which explicitly reveals multiple mixing effects and

(3) Numerical per-element distortion decomposition technique

This chapter discusses the applications of the above algorithms. Some of them are good candidates for future research. The work done in this thesis is applicable to mildly nonlinear circuits. It is possible that the same algorithms can be extended to an important class of strongly nonlinear circuits such as mixers and periodically switching networks. The attempt would be similar to the work in [26, 27], which use time-varying Volterra series. In those work, the strongly nonlinear circuits are treated as periodically time-varying weakly nonlinear systems w.r.t. the small-signal input of interest.

5.1   Application of Distortion Sensitivity analysis in EDA

In this section, the possible applications of the proposed distortion and sensitivity calculation in analog EDA are discussed. Sensitivity calculation can help solve the bottleneck in performance space exploration, the expensive evaluation of Jacobian and Hessian matrix [6,24]. With sensitivity information available at a reasonable price, there are ways to combine global optimization and fast convergence in automatic nominal sizing: introduction of weighted gradient-based moves in the annealing process; following genetic global search by fast local gradient search [10-12]. Sensitivity information is also important in yield optimization, e.g. to calculate yield gradients [6] or to linearize the feasible region [13]. Further, since sensitivity analysis quantifies the impact of layout parasitics on circuit performance, it can be directly applied in layout automation and postlayout "smart extraction" [10, 11, 12, 14].

5.2   Applications of Nonlinear Transfer Matrix

Nonlinear transfer matrix shows how lower order response generates higher order distortion from nonlinearity of different orders. It explicitly represents the obscure and complicated multiple mixing effects. Aside from speedup in sensitivity calculation and distortion decomposition, NTM can also be applied in the distortion study of small circuit, for example, the optimization of out of band terminal impedances for low distortion design of power amplifier (PA) [21]. In a two-tone test, the IM_3 at output is not only the sum of the effects from cubic nonlinearity, but also the cascaded quadratic nonlinearities. Since the power of the second order signal lies well away from the fundamental, filtering can be used to improve linearity by optimizing out-of-band impedances.

5.3   Applications of Distortion Decomposition

It is our genuine hope that designers will adopt the numerical distortion decomposition technique so that they can be liberated from the laborious symbolic/manual distortion analysis. Specifically, the applications in the following areas can serve as motivating success stories.

The proposed distortion decomposition can be applied in Simplification Before Generation (SBG) of symbolic expression as follows: first perform distortion decomposition at device level. Then by weighing the magnitude of different contributions, the original equivalent circuit can be simplified. In this way, we can achieve the best tradeoff between accuracy and compactness in the resultant symbolic expression [20] and component-level Volterra model [30,38].

Low distortion is commonly achieved by either symmetric circuit topology [22] or perfect tracking between the distortion and pre-distortion blocks [23]. In either case, the theoretical optimum design is very sensitive to process variations and parasitics. Because of the numerical advantages, we can use compact models including statistical process variation and parasitics in the distortion decomposition process. In this way, we can not only study the feasibility of the theoretical optimum but also pinpoint the weak point in the circuit under process variations. This can assist the selection of robust topology and layout automation.

Appendix

Simplified CMOS Transistor Model

The simplified CMOS transistor model adopted in this thesis is a combination of Level 1 and Level 3 models in [4, 44-47]. It takes into consideration the first order effects of mobility reduction due to vertical field and velocity saturation, linear variation of the depletion layer along the channel and body effect.

A.1. Mobility Reduction due to Velocity Saturation

Electron mobility μ depends on many effects, including the position in the channel and on the applied voltages. A simple model of mobility reduction is given by:

$$\mu = \frac{\mu_0}{1 + \theta(v_{GS} - V_T)} \tag{A.1}$$

A.2. Mobility Reduction due to Vertical Field

The assumption that the drift velocity of carriers is linearly proportional to the lateral electric field by $v = \mu E_x$, in which the proportionality is the mobility, is not correct when the drift velocity is comparable to the thermal velocity of carriers v_{sat}, which is about $10^7 cm/s$ for silicon at room temperature. As the drift velocity approaches the thermal velocity, the velocity will not increase

much anymore. This saturation effect is referred to as velocity saturation. The relationship between velocity and field reduces to:

$$v = \frac{\mu_{eff} E_x}{1 + \dfrac{E_x}{E_c}}$$

(A.2)

Here μ_{eff} is the effective mobility, E_x is the lateral electrical field and E_c is the critical electric field.

A.3. Variation of the Depletion Layer

If variation of the depletion layer along the channel is taken into account, we assume it changes linearly along the channel and characterize it with a :

$$a = \frac{2\sqrt{\phi + v_{SB}} + \gamma}{2\sqrt{\phi + v_{SB}}}$$

(A.3)

A.4 Transistor Model

The drain current in saturation region:

$$i_D = \frac{\mu_0 C_{ox}}{a2} \frac{W_{eff}}{L_{eff}} (v_{GS} - V_T)^2 (1 + \lambda v_{DS}) mobhot(v_{GS}, v_{SB})$$

(A.4)

In which

$$mobhot(v_{GS}, v_{SB}) = \frac{1}{1 + (\theta + \dfrac{\mu_0}{v_{sat} L_{eff}})(v_{GS} - V_T)}$$

(A.5)

The function *mobhot* indicates how much the drain current is reduced by the combination of mobility reduction due to a vertical field and velocity saturation. This function is always smaller than 1

$$V_T = V_{T0} + \gamma(\sqrt{v_{SB} + \phi} - \sqrt{\phi})$$

(A.6)

$$L_{eff} = L - 2(LD)$$

(A.7)

$$W_{eff} = W - 2(WD)$$

(A.8)

The capacitors in saturation region are:

$$C_{GB} = CGBO(L_{eff})$$

(A.9)

$$C_{GS} = CGSO(W_{eff}) + 0.67C_{ox}(W_{eff})(L_{eff})$$

(A.10)

$$C_{GD} = CGDO(W_{eff})$$

(A.11)

Note that due to the approximations, the drain current and capacitors calculated at the boundary between triode and saturation are discontinuous. Since all the transistors operate in saturation region in the Cascode Amplifier example studied in this thesis, no smoothing function is considered to connect the drain and capacitor equations between triode and saturation regions.

A.5 Technology Parameters

Technology parameters for a typical $0.8\mu m$ Silicon-Gate Bulk CMOS n-well process are used for the above transistor model [4, 44-47]:

Table A.1. Technology parameters for a $0.8\mu m$ **Silicon-Gate Bulk CMOS n-well process**

Parameter	Meaning	Used in Level	NMOS	PMOS	Unit
μ_0	Surface mobility of the channel	1,2,3	0.066	0.021	$m^2/V \cdot$
C_{ox}	Capacitance per unit area of the gate oxide	1,2,3	2.4665×10^{-3}		F/m^2
V_{T0}	Zero-bias gate-source extrapolated threshold voltage	1,2,3	0.7	-0.7	V
γ	Body-effect coefficient	1,2,3	0.4	0.57	$V^{1/2}$
ϕ	Surface inversion potential	1,2,3	0.7	0.8	V
λ	Channel length modulation factor	1,2	0.04	0.05	V^{-1}
θ	Mobility reduction coefficient	3	0.1	0.1	V^{-1}
v_{sat} ***	Saturation velocity	3	2.7×10^5	2.7×10^{5}*	m/s
LD	Lateral diffusion	2,3	0.016	0.015	μm
WD	Delta Width	2,3	0**	0	μm
$CGBO$		2,3	700×10^{-12}		F/m
$CGDO$		2,3	220×10^{-12}		F/m
$CGSO$		2,3	220×10^{-12}		F/m

*Based on the assumption NMOS and PMOS have the same saturation velocity [47]

**Assume to be zero, since unavailable

***Some of the parameters are fit parameters and their value do not necessarily correspond to the corresponding physical parameter. For example, the value of v_{sat} is not equal to the saturation velocity.

Bibliography

[1] M. Schetzen, "Multilinear theory of nonlinear networks," *Journal of the Franklin Institute,* Vol. 320, No.5, pp. 221-247, 1985.

[2] V. Volterra, *Theory of functionals and of integral and integro-differential equations,* New York: Dover, 1959.

[3] J. Vlach and K. Singhal, *Computer Methods for Circuits Analysis and Design,* 2nd ed., New York: Champman & Hall 1993.

[4] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*, Boston: Kluwer Academic Publishers, 1998.

[5] A. Bauer and W. Schwarz, "Circuit analysis and optimization with automatically derived Volterra kernels," *Proc. ISCAS,* 2000, Geneva, Switzerland, Vol I, pp. 491-494.

[6] G. Stehr, *On the Performance Space Exploration of Analog Integrated Circuits*, PhD Thesis, Technische Universität München, Verlag Dr. Hut, Munich,2005.

[7] P. Dobrovolny et al, "Analysis and Compact Behavioral Modeling of Nonlinear Distortion in Analog Communication Circuits," *IEEE Trans. CAD*, VOL. 22, NO. 9, pp. 1215–1227, Sept 2003.

[8] 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.

[9] B. Hernes and T. Saether, *Design Criteria for Low Distortion in Feedback Opamp Circuits,* Norwell, MA: Kluwer, 2003.

[10] G. Gielen and R. Rutenbar, "Computer-aided design of analog and mixed-signal integrated circuits," *Proc. IEEE,* vol. 88, pp. 1825–1852, Dec. 2000.

[11] G. Van der Plas et al., "AMGIE: A synthesis environment for CMOS analog integrated circuits," *IEEE Trans. Computer-Aided Design,* vol. 20, pp. 1037–1058, Sept. 2001.

[12] S. Director, P. Feldmann, and K. Krishna, "Optimization of parametric yield: A tutorial," *Proc. IEEE Custom Integrated Circuit Conference,* 1992, pp. 3.1.1–3.1.8.

[13] J. M. Wojciechowski and J. Vlach, "Ellipsoidal method for design centering and yield estimation," *IEEE Trans. Computer-Aided Design,* vol. 12, pp. 1570–1579, Oct. 1993.

[14] K. Lampaert et al., *Analog Layout Generation for Performance and Manufacturability*, Norwell, MA: Kluwer, 1999.

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

[16] *HSPICE: HSPICE User's Manual Release 96.1*, Meta-Software Inc., 1996

[17] *Microwave Office User's manual II*, Applied Wave Research, Inc. 2000

[18] T. Quarles, A. R. Newton, D. O. Pederson, and A. Sangiovanni-Vincentelli, *SPICE 3 Version 3F5 User's Manual*, Dept. Elec. Eng. Comp. Sci., Univ. of California, Berkeley, CA, Mar. 1994.

[19] A. Heiskanen, J. Aikio, and T. Rahkonen, "A 5th order Volterra study of a 30 W LDMOS power amplifier," *in Proc. IEEE ISCAS*, 2003, pp. 616–619.

[20] X. Li, P. Li, Y. Xu and L. Pileggi, "Analog and RF circuit macromodels for system-level

analysis," *Proc. of ACM/IEEE DAC*, 2003.

[21] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*.Norwood, MA: Artech House, 2003.

[22] A. Kopa and A. B. Apsel, "Common-emitter Feedback Transimpedance Amplifier for Analog Optical Receivers," *ISCAS 2006,* pp.5479-5482, May. 2006.

[23] 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.

[24] http://www.solidodesign.com/

[25] J. C. Pedro and N. B. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Norwood, MA: Artech House, 2003.

[26] F. Yuan and A. Opal, "Distortion analysis of periodically switched nonlinear circuits using time-varying volterra series," *Trans. on Circuits and systems,* vol. 48, no.6, pp. 726-737, June 2001.

[27] M. T. Terrovitis and R. G. Meyer, "Intermodulation distortion in current commutating CMOS mixers," *IEEE J. Solid-State Circuits*, vol. 35, pp.1461–1473, Oct. 2000.

[28] C. L. Recker, R. Braswell, P.G. Drennan and C.C.McAndrew, "A Web Tool for Interactive Exploration of Analog Design Tradeoffs", *Proc. IEEE Custom Integrated Circuit Conference 2006,* pp. 631-634, Sept 2006.

[29] M. E. Schlarmann, R. L. Geiger, "Prototype implementation of a www based analog circuit design tool", IEEE ISCAS, vol. I, pp. 97-100, 2001.

[30] P.Wambacq *et al.*, "Compact modeling of nonlinear distortion in analog communication circuits," in *Proc. Design, Automation, Test in Eur.Conf.*, Mar. 2000, pp. 350–354.

[31] P. Feldmann and J. Roychowdhury. Computation of circuit waveform envelopes using an efficient, matrix-decomposed harmonic balance algorithm. *IEEE International Conference on Computer-Aided Design: Digest of Technical Papers*, November 1996.

[32] K. Kundert, "Simulation methods for RF integrated circuits," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, 1997, pp. 752–765.

[33] K. Kundert, "Introduction to RF simulation and its applications," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1298–1319, Sept. 1999.

[34] http://www.maplesoft.com/

[35] Y. Cheng and C. Hu, *MOSFET Modeling & BSIM3 User's Guide,* Kluwer Academic Publisher, 1999.

[36] M.T. Terrovitis, and R.G. Meyer, "Intermodulation Distortion in Current-Commutating CMOS Mixers," *IEEE Journal of Solid-State Circuits*,vol.35, pp.1461-1473, October, 2000.

[37] S. C. Terry, J. M. Rochelle, D. M. Binkley, B. J. Blalock, D. P. Foty, and M. Bucher, "Comparison of a BSIM3V3 and EKV MOSFET mode for a 0.5 μm CMOS process and implications for analog circuit design," *IEEE Trans. Nucl. Sci.*, vol. 50, no. 4, pp. 915–920, Aug. 2003.

[38] P.Wambacq, G.Vandersteen, P. Dobrovolný, M. Goffioul,W. Eberle, M. Badaroglu, and S. Donnay, "High-level simulation and modeling tools for mixed-signal front-ends of wireless systems," in *Proc. Advances in Analog Circuit Design*, Mar. 2002.

[39] G. Gielen, H. Walscharts and W. Sansen, "ISSAAC: a symbolic simulator for analog integrated circuits," *IEEE J. Solid-State Circuits,* Vol. 24, No.6, pp. 1587-1597 December 1989.

[40] S. A. Maas, *Nonlinear Microwave and RF Circuits,* 2nd ed., Boston: Artech House Microwave Library.

[41] G. Zhu and A. Opal, "Sensitivity Analysis of Nonlinear Circuits using Volterra Series," *Proc. ISCAS 2006,*pp. 5035-5038, May 2006.

[42] G. Tulunay and S. Balkır, "Automatic Synthesis of CMOS RF Front-Ends," *Proc. ISCAS 2006,* pp. 625-628, May 2006

[43] B. Hernes and W. Sansen, "Distortion in single-, two- and three-stage amplifiers," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 52, no. 5, pp. 846–856, May 2005.

[44] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits,* 3rd ed., NY: Wiley, 1993.

[45] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge: Cambridge University Press, 1998.

[46] P. E. Allen and D. R. Holberg, *CMOS analog circuit design,* 2nd ed., New York: Oxford University Press, 2002

[47] J. M. Rabaey, A. Chandrakasan and B. Nikolic, *Digital Integrated Circuit—A Design Perspective,* 2nd ed., Englewood Cliffs, NJ: Prentice Hall, 2002

[48] G. Zhu and A. Opal, "Per-Element Decomposition in Distortion Analysis," *ISCAS 2007*