A Hybrid Approach to Nonlinear Macromodel Generation for Time-Varying Analog Circuits[†]

Peng Li, Xin Li, Yang Xu and Lawrence T. Pileggi

ECE Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ABSTRACT

Modeling frequency-dependent nonlinear characteristics of complex analog blocks and subsystems is critical for enabling efficient verification of mixed-signal system designs. Recent progress has been made for constructing such macromodels, however, their accuracy and/or efficiency can break down for certain problems, particularly those with high-Q filtering. In this paper we explore a novel hybrid approach for generating accurate analog macromodels for time-varying weakly nonlinear circuits. The combined benefits of nonlinear Padé approximations and pruning by exploitation of the system's internal structure allows us to construct nonlinear circuit models that are accurate for wide input frequency ranges, and thereby capable of modeling systems with sharp frequency selectivity. Such components are widely encountered in analog signal processing and RF applications. The efficacy of the proposed approach is demonstrated by the modeling of large time-varying nonlinear circuits that are commonly found in these application areas.

1. INTRODUCTION

Building compact analog circuit macromodels is a key for enabling complete system verification and high-level design exploration for mixed-signal system design [3]. Various analog macromodels can be categorized according to their intended application (a good taxonomy can be found in [17]). Performance models such as those in [17] and [18] are mainly employed in synthesis for design space exploration; While other research work has focused on models with a primary application of simulation and system verification [8]-[10][12]-[16]. It should be noted, however, that the latter category is not excluded from use in design exploration when appropriate.

This paper addresses the modeling problems that largely fall into the second category. To this end, two requirements exist for any such circuit model: efficiency and accuracy. The latter is often measured by certain performance metrics such as gain, bandwidth and linearity, etc. It is worth emphasizing that modeling system distortion is crucial for most analog applications, since circuit linearity is one of the most important design specifications. This casts the modeling task into a nonlinear model generation or nonlinear model reduction problem, both of which are substantially more challenging than their linear time-invariant model counterparts[8]-[10][12]-[16].

Several nonlinear model generation approaches have been recently reported, and can be largely categorized into two major classes: pruning-based and transformation-based. The first type of approaches work directly on the original circuit structure, and apply pruning-based simplification for model generation. In particularly, symbolic modeling of weakly nonlinear circuits has been used to build system-level models [14]-[16]. The underlying mathematical description employed is Volterra series [1]-[2], and the resulting model is in a block-diagram form suitable for signal-flow type simulation using tools such as Matlab Simulink [19].

The second class of methods, categorized as transformation-based, have been stimulated by the progress made for model order reduction of large linear IC interconnects in the past decade (e.g. [4]-[7]). The recent research effort in this direction was to extend projection-based model order reduction techniques for LTI (linear time invariant) systems to accommodate nonlinear system reduction. The piecewise-linear approximation approach in [12] models a nonlinear system using a collection of weighted linear models based upon a state trajectory generated by a training input. Then each of these linear models are reduced using projection-based techniques for LTI systems [5]-[7]. This approach has the potential capability of handling large nonlinearities, but limited by the training-input dependency. This piecewise-linear approximation was recently extended to a piecewise-polynomial approach to better model nonlinear distortion [24]. Another approach that does not completely fall into either of the above two categories uses model templates for building system-level simulation models [23].

For a broad class of weakly nonlinear systems in analog signal processing and RF communication ICs, the low distortion level, as well as the required modeling accuracy, seemingly makes Volterra series a more suitable choice for system description. The work of [8]-[10] extends the projection-based LTI techniques to weakly nonlinear systems that are described by Volterra series, thereby facilitating a direct transformation-based macromodeling. Recently, several improvements to this nonlinear model reduction framework were proposed in [13]. Based on a careful moment analysis under a general nonlinear system context, the nonlinear model order reduction algorithm NORM [13] uses a minimum set of Krylov subspaces for projection, leading to a significant reduction of model size.

As we move towards modeling larger analog circuit building blocks and subsystems, it becomes increasingly beneficial to exploit the model order reduction techniques that were so successful for LTI systems. This is particularly the case when various filtering blocks, along with the parasitics, are included as part of the system model. However, extending LTI based techniques for analog model generation is nontrivial, as the nonlinear aspects of the problem greatly complicates the situation. For the reduction framework of [8]-[10][13], one difficulty is that the reduced high-order model matrices are extremely dense, and increasingly more expensive to form as the model size grows. This difficulty makes control of the model size growth a critical issue for an effective model reduction [13]. Thus, a full projection-based approach is more appropriate when a low-order approximation is adequate.

In this paper, we propose a hybrid approach for generating nonlinear analog macromodels that can efficiently capture weakly nonlinear effects up to third order. The low-order system nonlinear responses are captured efficiently using recently developed projection-based NORM algorithm of [13], while high-order responses are approximated via direct matrix pruning, in conjunction with the projection-based reduction of a linear adjoint network. The pruning technique is similar to what was used in symbolic modeling [14]-[16], with the difference being that the pruning is applied directly to the high-order system matrices for our approach, and the model order reduction is embedded with the pruning to accelerate this relatively slow procedure. The overall model generation cost is bounded by a time complexity of $O(sElog(E) + kN^3)$, where E is the number of the nonzeros in the second and third

order system matrices, N is the problem dimension, k is number of matrix factorizations for building projection-based reduced models, and s is total number of frequency samples evaluated during the pruning.

By exploiting projection-based model order reduction and the internal circuit structure simultaneously, this hybrid approach allows us to model system nonlinear distortion over a wide range of input frequency of interest, thereby extending the applicability of nonlinear model generation. We demonstrate the efficacy of our approach on several large time-varying circuit examples, where frequency-domain nonlinear characteristics vary dramatically due to active or high Q filtering.

2. BACKGROUND

2.1 Nonlinear System Modeling Via Volterra Series

Volterra series has been widely used to characterize weakly nonlinear systems [1]-[2]. It is usually applicable to cases where the input amplitudes to the system are small and nonlinearities are excited in a weakly nonlinear fashion. For a circuit with input $u(t)$, the response (circuit unknowns) $x(t)$ can be expressed using the expansion

$$x(t) = \sum_{n=1}^{\infty} x_n(t),\qquad (1)$$

where $x_n(t)$ is the nth order response, which is related to the input via convolution in time domain:

$$x_n(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n)u(t-\tau_1)\dots u(t-\tau_n)d\tau_1\dots d\tau_n.\quad (2)$$

In the above equation, $h_n(\cdot)$ is the nth order Volterra kernel, and can be thought as an extension to the impulse response function of a linear system. (2) can also be cast into the frequency domain, where the Laplace transform of the nth order Volterra kernel $H_n(s_1, s_2, \dots, s_n)$, or the nth order nonlinear transfer function, is used.

Before using Volterra series for analysis, consider the following ordinary differential-algebraic equations for a nonlinear circuit

$$f(x(t)) + \frac{d}{dt}q(x(t)) = bu(t),\qquad (3)$$

where $f(.)$ and $q(.)$ represent resistive and dynamic nonlinearities. To apply Volterra series, first expand circuit nonlinearities about a dc bias x_0 (this can be extended to a varying operating point for time-varying systems) into Taylor series. Correspondingly, we define

$$G_i = \frac{1}{i!}\frac{\partial^i f}{\partial x^i}\bigg|_{x=x_0} \in R^{n \times n^i}, C_i = \frac{1}{i!}\frac{\partial^i q}{\partial x^i}\bigg|_{x=x_0} \in R^{n \times n^i}.\qquad (4)$$

It follows that the first through third order nonlinear small-signal responses, based upon x_0, are computed recursively as

$$\frac{d}{dt}(C_1 x_1) + G_1 x_1 = bu_s(t)$$

$$\frac{d}{dt}(C_1 x_2) + G_1 x_2 = -\frac{d}{dt}(C_2(x_1 \otimes x_1)) - G_2(x_1 \otimes x_1)$$

$$\frac{d}{dt}(C_1 x_3) + G_1 x_3 = -G_2(x_1 \otimes x_2 + x_2 \otimes x_1) - G_3(x_1 \otimes x_1 \otimes x_3)$$

$$-\frac{d}{dt}(C_2(x_1 \otimes x_2 + x_2 \otimes x_1) + C_3(x_1 \otimes x_1 \otimes x_3)),\qquad (5)$$

where u_s is the small-signal input to the circuit.

2.2 Nonlinear Model Order Reduction

Equation (5) can be immediately interpreted as a set of linear systems with different inputs. Following the work in [8][10], this nonlinear system model can be reduced using an extension to pro-

jection-based LTI reduction techniques (e.g. [5]-[7]). The nonlinear model reduction was achieved by successively reducing these linear systems with increasing number of inputs, for approximating the first order response x_1, second order response x_2 and third order response x_3, respectively. However, the efficacy of model order reduction eventually suffers from the rapid increase in the number of inputs to the successive linear networks. In [9], a well structured projection-based moment matching scheme was developed based on a theoretically interesting bilinear form of weakly nonlinear systems. However, a bilinear form of a nonlinear system has significantly more state variables than the original state-equation representation. Therefore, the use of a bilinear form leads to a reduction problem with a much larger size.

In [13], a nonlinear model order reduction algorithm NORM was proposed to improve the reduced model compactness under the projection-based framework by carefully examining the problem under a general nonlinear context. For the expanded MNA formulation of (5), this was accomplished by first deriving a general matrix-form nonlinear transfer function $H_n(s_1, s_2, \dots s_n)$ and the associated moments, and then applying a minimum set of Krylov-subspace projection vectors to match the moments of nonlinear transfer functions. A complete description of the NORM algorithm is outside of the scope of this paper. To provide sufficient background to understand the hybrid approach proposed in this paper, we briefly discuss the algorithm and limit it to single-input multi-output (SI-MO) time-invariant systems.

The first order linear transfer function of (5) is given by

$$(G_1 + sC_1)H_1(s) = b.\qquad (6)$$

Due to the recursive nature of (5), the symmetrized second order nonlinear transfer function is determined by

$$[G_1 + \bar{s}C_1]H_2(s_1, s_2) = -[G_2 + \bar{s}C_2] \cdot \overline{H_1(s_1) \otimes H_1(s_2)},\qquad (7)$$

where $\overline{H_1(s_1) \otimes H_1(s_2)} = \frac{1}{2}(H_1(s_1) \otimes H_1(s_2) + H_1(s_2) \otimes H_1(s_1))$, $\bar{s} = s_1 + s_2$. It can be shown that the symmetrized third order transfer function can be computed using

$$[G_1 + \tilde{s}C_1]H_3(s_1, s_2, s_3) = -[G_3 + \tilde{s}C_3] \cdot$$

$$\overline{H_1(s_1) \otimes H_1(s_2) \otimes H_1(s_3)} - 2 \cdot [G_2 + \tilde{s}C_2] \cdot \overline{H_1(s_1) \otimes H_2(s_2, s_3)}\qquad (8)$$

where $\tilde{s} = s_1 + s_2 + s_3$, $\overline{H_1(s_1) \otimes H_1(s_2) \otimes H_1(s_3)}$ is the arithmetic average of terms over all possible permutations of frequency variables in the Kronecker product, and

$$\overline{H_1(s_1) \otimes H_2(s_2, s_3)} = \frac{1}{6}(H_1(s_1) \otimes H_2(s_2, s_3) + H_1(s_2) \otimes H_2(s_1, s_3)$$

$$+ H_1(s_3) \otimes H_2(s_1, s_2) + H_2(s_1, s_2) \otimes H_1(s_3)\qquad (9)$$

$$+ H_2(s_1, s_3) \otimes H_1(s_2) + H_2(s_2, s_3) \otimes H_1(s_1))$$

To perform a projection-based nonlinear Padé approximation, moments of the corresponding nonlinear transfer functions are required. For this purpose, (6)-(8) are expanded into a power series w.r.t frequency variables at the origin

$$H_1(s) = \sum_{k=0}^{\infty} s^k M_{1,k}, H_2(s_1, s_2) = \sum_{k=0}^{\infty}\sum_{l=0}^{k} s_1^l s_2^{k-l} M_{2,k,l}$$

$$H_3(s_1, s_2, s_3) = \sum_{k=0}^{\infty}\sum_{l=0}^{k}\sum_{m=0}^{k-l} s_1^l s_2^m s_3^{k-l-m} M_{3,k,l,m}\qquad, \quad (10)$$

where $M_{1,k}$, $M_{2,k,l}$ and $M_{3,k,l,m}$ are a kth order moment of $H_1(s)$, $H_2(s_1, s_2)$ and $H_3(s_1, s_2, s_3)$, respectively. We define a reduced order model as a kth order model in $H_1(s)$, $H_2(s_1, s_2)$ or $H_3(s_1, s_2, s_3)$ if and only if the respective moments are preserved in the reduced model up to the kth order. In NORM, to match a specific number of moments, a set of Krylov subspace vectors with a minimum total dimension is used in projection-based reduction for improving the model compactness. For instance, if we define

$A = -G_1^{-1}C_1$, $r_1 = G^{-1}b$, $K_m(A, p) = colspan\{p, Ap, ..., A^{m-1}p\}$, it can be shown that the following Krylov subspaces are the required minimum set containing up to the kth order moment directions of H_2:

$$K2(k) = \bigcup_i K_{m_i}(A, v_i), \qquad (11)$$

where

$$v_i = \begin{cases} G_1^{-1}G_2((A^m r_1) \otimes (A^n r_1) + (A^n r_1) \otimes (A^m r_1)) \\ \forall m \geq 0, \forall n \geq 0, m + n \leq k, m \geq n, with\ m_i = k - m - n + 1 \\ G_1^{-1}C_2((A^m r_1) \otimes (A^n r_1) + (A^n r_1) \otimes (A^m r_1)) \\ \forall m \geq 0, \forall n \geq 0, m + n \leq k - 1, m \geq n, with\ m_i = k - m - n \end{cases} \qquad (12)$$

A minimum set of Krylov subspace vectors can also be found for moment matching of the third order transfer function H_3. If we combine Krylov subspaces for moment matching of the first through the third order nonlinear transfer functions and compute a orthonormal basis V for the combined subspaces, the reduced order model can be computed via a projection of the form:

$$\tilde{G}_1 = V^T G_1 V, \tilde{C}_1 = V^T C_1 V, \tilde{G}_2 = V^T G_2(V \otimes V), \tilde{C}_2 = V^T C_2(V \otimes V)$$

$$\tilde{G}_3 = V^T G_3(V \otimes V \otimes V), \tilde{C}_3 = V^T C_3(V \otimes V \otimes V), \tilde{b} = V^T b \quad (13)$$

As seen from (13), the dense third order reduced system matrices are of dimension $q \times q^3$, where q is the number of columns for projection V, which represents the size of the reduced order model. Thus, controlling model size by a proper selection of projection vectors is critical, as addressed in NORM:

- The use of a minimum set of Krylov subspace vectors avoids unnecessary inflation of the model size, thereby leading to a significant improvement in model compactness. Furthermore, it is shown that the order of moment matching for a transfer function at a higher order cannot exceed that of a transfer function of a lower order. In NORM, the order of moment matching for the nonlinear transfer function at each order can be chosen individually, provided that the above condition is satisfied, and the size of overall reduced order model is minimized for each of these choices.

- Unlike the counterpart of LTI system reduction, adopting a multi-point method can improve the model compactness under nonlinear context as recognized in NORM. This stems from the fact that the total dimension of optimal Krylov subspaces used in a single-point method actually exceeds the number of moments matched in the resulting reduced order model. However, the use of a multi-point method, in particular, a zeroth order multi-point NORM, where high order nonlinear transfer function values (*zero*th order moments) are matched simultaneously at several expansion points, can produce a reduced order model with a size equal to the number of moments matched. That means multi-point methods can further increase the model compactness that is essential for nonlinear system reduction.

To compare different methods, we define the size of a reduced order model as the number of state variables that it contains. The worst-case model sizes for the system in (5) are compared in Table 1, where reduced order models match the moment of H_2 or H_2 and H_3 up to certain order. NORM-SP is the single-point version of NORM, and NORM-MP is the "equivalent" *zero*th order multi-point NORM in which the same number of transfer function moments are matched. It is evident that the use of NORM, especial-ly multi-point NORM, dramatically reduces the model size.

Model Order	3	4	5	6	7	8
[8][10]:H2	116	230	402	644	968	1386
[8][10]:H2-H3	3,700	11,480	28,914	63,070	123,848	224,460
NORM-SP:H2	24	40	62	91	128	174
NORM-SP:H2-H3	66	118	194	301	446	636
NORM-MP:H2	16	23	30	39	48	59
NORM-MP:H2-H3	37	56	78	108	141	182

Table 1. Worst-case reduced order model size

2.3 Modeling of Time-Varying Weakly Nonlinear Systems

A large class of networks, such as clocked switching networks and RF current switching mixers, can be characterized as time-varying weakly nonlinear systems with respect to the small input signal being processed. This would require us to treat the large excitations to the system, such as clock and LO, as part of the system under modeling, and analyze the small-signal input upon the time-varying operating point caused by the large excitations. One special case of practical interest is that the system has a periodically time-varying operating point introduced by a periodic clock or LO.

Assume that the large-signal periodic excitation to the system has a period $T_0 = (2\pi)/\omega_0$, then the weakly nonlinear system response to the small-signal input can be characterized using periodically time-varying Volterra series. Now the nonlinear transfer function $H_n(t, s_1, ...s_n)$ becomes a periodic function of time with the same period T_0, and can be expanded into Fourier series

$$H_n(t, \omega_1, ...\omega_n) = \sum_{k=-\infty}^{\infty} H_{n,k}(\omega_1, ...\omega_n) \cdot e^{jk\omega_0 t}. \qquad (14)$$

Using (14), the nth order response is in the form

$$x_n(t) = \left(\frac{1}{2\pi}\right)^n \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} H_n(t, \omega_1, ...\omega_n) \cdot U(\omega_1) \cdot ... \cdot U(\omega_n)$$
$$\cdot e^{j(\omega_1 + ... + \omega_n)t} d\omega_1 ... d\omega_n$$
$$= \left(\frac{1}{2\pi}\right)^n \cdot \sum_{k=-\infty}^{\infty} e^{jk\omega_0 t} \cdot \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} H_{n,k}(\omega_1, ...\omega_n) \qquad (15)$$
$$\cdot U(\omega_1) \cdot ... \cdot U(\omega_n) \cdot e^{j(\omega_1 + ... + \omega_n)t} \cdot d\omega_1 ... d\omega_n$$

It can be seen from the above equation that the kth harmonic $H_{n,k}$ of $H_n(t, \cdot)$ specifies the nth order nonlinear effect that is frequency translated by $k\omega_0$ due to the time variation of the system.

By extending the formulation for linear periodically time-varying systems in [20][11][8] via either frequency-domain collocation or time-domain backward difference, the periodically time-varying nonlinear transfer functions can be formulated in a matrix form similar to that of (6)-(8). In this paper, a time-domain backward Euler is employed to discretize $H_n(t, \cdot)$ over M sample points $\overline{H}_n = [H_n(t_1, \cdot)^T, H_n(t_2, \cdot)^T, ..., H_n(t_M, \cdot)^T]^T$ within a period of T_0. Analogous to the use of $G_1, C_1, b_1, G_2, C_2, G_3, C_3$ in (6)-(8), a set of matrices $\overline{G}_1, \overline{C}_1, \overline{b}_1, \overline{G}_2, \overline{C}_2, \overline{G}_3, \overline{C}_3$ can be formulated based on discretization of the system nonlinear characteristics over a period of T_0 for time-varying systems. Based on these matrices, (6)-(8) can be reformulated in terms of $\overline{H}_1, \overline{H}_2, \overline{H}_3$. Each of these equations now has $M \times N$ unknowns, where N is the number of physical circuit unknowns of the system. To accurately capture the time-varying operating point due to the large-signal excitation, many sample points are often required, leading to a set of large time-varying system equations. These equations must be reduced to be used efficiently in simulation. In the remainder of this paper we focus on the time-varying systems as the results for time-invariant systems

can be easily deduced from this more general description.

3. HYBRID APPROACH

3.1 An Illustrative Example

To motivate the need for a hybrid approach, consider the simple nonlinear circuit shown in Fig. 1(a). This circuit has only two non-linearities, resistors R_1 and R_2. Each of them is modeled as a third order polynomial $i = a_1v + a_2v^2 + a_3v^3$. The small-signal current source drives a high Q resonator (tuned at 1GHz), which is followed by 500 RC segments. A cross-section of the third order nonlinear transfer function at the output is plotted in Fig. 1(b). It is evident from Fig. 1(b) that when coupled with high Q resonator (part of linear network), the two nonlinearities create a sharp transition for the third order nonlinear transfer function within a neighborhood of the resonant frequency. To capture accurately the third-order nonlinear effects around the resonant frequency using a projection-based implicit moment-matching, many projection vectors may be required. In the reduced model, although the number of states can be significantly reduced, a large and dense reduced third order matrix $\tilde{G}_3$ may result(q^4 nonzeros, q is the number of states in the reduced model). This is in contrast to the original third order matrix G_3 that is of much higher dimension but only has 2 nonzeros.

Projection-based nonlinear model reduction is very effective if a low-order approximation is sufficiently accurate. In such cases, the reduced order model can also be constructed rather efficiently (the dominate cost is no more than a few matrix factorizations). However, for the cases where either a good accuracy is required over a wide range of frequency or the nonlinear frequency-domain characteristics varies dramatically within a band of interest, a complete projection-based approach might be difficult to apply. For practical examples, the primary reason for this difficulty is that projecting high order system matrices (e.g. G_3) becomes less and less efficient as the dimension of the Krylov subspaces increases.
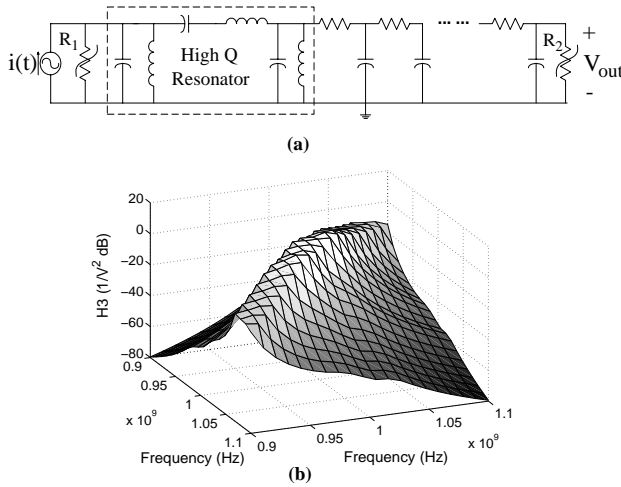

(a)

(b)

Fig. 1. (a) A simple nonlinear circuit, and (b) its third order nonlinear transfer function

3.2 Hybrid Approach

As clearly seen from Table 1, if we use multi-point NORM to only approximate the first and second order nonlinear transfer functions, the size of the reduced order model is much less constrained. More specifically, as the third order nonlinear transfer functions are not considered in the projection-based reduction, the reduced model size now grows much slower as the order of moment-matching increases. For the same reason, forming the reduced third order system matrices is not an issue any more, nei-

ther is it a limiting factor for the model size. Moreover, the adoption of the projection-based approach allows us to efficiently model the first and second order system properties, especially for large circuits with detailed parasitics. From a pruning-based modeling perspective, although numerous nonlinearities can exist in a circuit, there is often a natural tendency for only a few of them to be dominant due to the specific circuit structure. Thus, identifying these dominant nonlinearities under the original circuit structure can be a very useful component of model generation. In the proposed hybrid approach, the low-order (first and second) responses are matched using projection while high-order (third) response are approximated by exploiting both the circuit internal structure and projection-based reduction.

3.2.1 Approximation of low-order responses

We begin with application of multi-point NORM to approximate the first and second nonlinear transfer functions. We denote the corresponding projection matrix as $V \in R^{NM \times q}$, where N is the number of physical circuit unknowns, $M = 2K - 1$ is number of time-domain samples to discretize the time-varying nonlinear transfer functions (assume an odd number of samples for simplicity), and q is the size of reduced order model. We further denote the reduced first and second order matrices produced by NORM as

$$\tilde{G}_1 = V^T\overline{G_1}V \in R^{q \times q}, \tilde{C}_1 = V^T\overline{C_1}V \in R^{q \times q}, \tilde{b} = V^T\bar{b} \in R^q$$
$$\tilde{G}_2 = V^T\overline{G_2}(V \otimes V) \in R^{q \times q^2}, \tilde{C}_2 = V^T\overline{C_2}(V \otimes V) \in R^{q \times q^2} \quad . \quad (16)$$

Substituting (16) into (6)-(7), the first and second order transfer functions $\tilde{H}_1$ and $\tilde{H}_2$ of the reduced model can be computed. The time-sampled first and second order transfer functions of the original system can be approximated as

$$\hat{\tilde{H}}_1 = V\tilde{H}_1, \hat{\tilde{H}}_2 = V\tilde{H}_2 . \quad (17)$$

As the reduced model of (16)-(17) represents a time-invariant system, the approximation of the first and second order responses of the original time-varying system involves the use of discrete Fourier transform. It can be shown that the following q th order system is a time-domain realization of the corresponding reduced model

$$\begin{cases} \dfrac{d}{dt}(\tilde{C}_1\tilde{x}_1) + \tilde{G}_1\tilde{x}_1 = \tilde{b}u_s(t) \\[2mm] \dfrac{d}{dt}(\tilde{C}_1\tilde{x}_2) + \tilde{G}_1\tilde{x}_2 = -\dfrac{d}{dt}(\tilde{C}_2(\tilde{x}_1 \otimes \tilde{x}_1)) - \tilde{G}_2(\tilde{x}_1 \otimes \tilde{x}_1) \\[2mm] \tilde{y}(t) = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix} = \begin{bmatrix} J & 0 \\ 0 & J \end{bmatrix}\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} \end{cases}, \quad (18)$$

where $J = D \cdot \Gamma \cdot V$, Γ is the DFT matrix converting M time samples to the corresponding Fourier coefficients, $D = [e^{-jK\omega_0t}I_{N \times N}, ..., I_{N \times N}, ..., e^{jK\omega_0t}I_{N \times N}]$, $I_{N \times N}$ is the $N \times N$ identity matrix. The first and second order responses of the original system can be approximated in time-domain by $\hat{x}_1 = \tilde{y}_1, \hat{x}_2 = \tilde{y}_2$.

3.2.2 Approximation of the third order response

To approximate the third order transfer function based on the reduced model of (16)-(17) or (18), substitute (16) into (8)

$$[\overline{G_1} + \tilde{s}\overline{C_1}]\hat{\tilde{H}}_3(s_1, s_2, s_3) = g , \quad (19)$$

where

$$g = [\overline{G_3} + \tilde{s}\overline{C_3}] \cdot u_1 + [\overline{G_2} + \tilde{s}\overline{C_2}] \cdot u_2$$
$$u_1 = -(V\tilde{H}_1(s_1)) \otimes (V\tilde{H}_1(s_2)) \otimes (V\tilde{H}_1(s_3)) \quad (20)$$
$$u_2 = -2(V\tilde{H}_1(s_1)) \otimes (V\tilde{H}_2(s_2, s_3))$$

As seen from the above equations, to compute the third order trans-

fer function, one needs to solve a linear system in terms of $\overline{G}_1$ and $\overline{C}_1$, and form the input to the linear system in terms of g that is a function of high-order system matrices $\overline{G}_2, \overline{C}_2, \overline{G}_3, \overline{C}_3$ and low-order transfer functions $\overline{H}_1, \overline{H}_2$. As $\overline{H}_1, \overline{H}_2$ are approximated using (17), to further reduce (19) requires reduction of the dimension of the linear problem in (19) as well as the high-order system matrices $\overline{G}_2, \overline{C}_2, \overline{G}_3, \overline{C}_3$. This goal is accomplished by a combination of projection-based reduction of an adjoint network and direct matrix pruning.

In analog signal processing and RF applications, a circuit block usually has only one or at most a few output nodes. For time-varying systems, typically only a small number of sidebands (w.r.t. clock or LO) for the outputs are of interest. Thus, (19) can be viewed as a system with potentially many inputs but few outputs, and can be reduced as an adjoint network. Without loss of generality, let us assume only the ith sideband of the voltage response at node p is of interest (e.g., $i = 0, \pm 1$). Define $l = \begin{bmatrix} 0 & 0 & \cdots & 1 \cdots & 0 \end{bmatrix}^T$ as a $N \times 1$ vector that has a 1 at the p-th location and zeros at all other locations, and

$$L = diag(l^T, l^T, ..., l^T) \in R^{M \times NM}$$
$$d = \begin{bmatrix} 1 & e^{-j2\pi i/M} & \cdots & e^{-j2\pi(M-1)i/M} \end{bmatrix}^T / M, b_a = L^T d \tag{21}$$

The ith sideband of the third order transfer function at node p, y_{p_i}, (corresponding to the ith harmonic of the third order transfer function for node p) can be obtained from the adjoint network

$$\begin{cases} [\overline{G}_1^T + \tilde{s}\overline{C}_1^T] z(s_1, s_2, s_3) = b_a \\ y_{p_i} = g^T z \end{cases} \tag{22}$$

We can apply Krylov subspace projection (such as [7]) to reduce the linear adjoint network of (22)

$$\hat{G}_a = V_a^T \overline{G}_1^T V_a, \hat{C}_a = V_a^T \overline{C}_1^T V_a, \hat{b}_a = V_a^T b_a, \tag{23}$$

where V_a is a orthonormal basis of the Krylov subspace $colspan\{r_a, A_a r_a, A_a^2 r_a...\}$ with $A_a = -\overline{G}_1^{-T} \overline{C}_1^T, r_a = \overline{G}_1^{-T} b_a$. As the DFT vector d is absorbed into b_a, to perform a real projection, b_a can be split into real and imaginary parts before reduction [21]. Based on (22)-(23), we can approximate y_{p_i} as

$$\hat{y}_{p_i} = g^T \hat{z}, \hat{z} = V_a[\hat{G}_a + \tilde{s}\hat{C}_a]^{-1}\hat{b}_a \tag{24}$$

To speed up the computation of the third order transfer function or response, high order system matrices $\overline{G}_2, \overline{C}_2, \overline{G}_3, \overline{C}_3$ need to be reduced. To avoid forming dense reduced matrices (particularly for the third order ones) in a projection-based approach, here we exploit the internal structure of the problem by applying a direct matrix pruning technique, where nonzero elements of $\overline{G}_2, \overline{C}_2, \overline{G}_3, \overline{C}_3$ are pruned according to their contributions to the third order transfer function at the output node of interest. Although a nonlinear circuit may contain many nonlinearities, a few of them tend to be dominant. Therefore, retaining the original circuit structural information by avoiding using projection and pruning these high-order matrices in the original coordinates can be very effective. This same idea has been used in symbolic modeling of nonlinear circuits, where dominant circuit nonlinearities are identified and kept for building system-level models [14]-[16].

The pruning process proceeds as follows: A set of sampled frequency points Ω are first selected within the frequency band of interest. Then for each of these frequency samples, the third order transfer function are computed, as well as the individual contributions of nonzeros in $\overline{G}_2, \overline{C}_2, \overline{G}_3, \overline{C}_3$. These contributions are sorted by magnitude, and non-dominant ones are discarded provided that a user-specified error tolerance on H_3 is not exceeded. The end products of the process are a set of pruned matrices $\overline{G}_{2p}, \overline{C}_{2p}, \overline{G}_{3p}, \overline{C}_{3p}$ that satisfy the error tolerance for all the sampled frequency points in Ω. As the computation of H_3 and individ-

ual contributions may take place on many sample points, matrix pruning can be a slow process. However, to speed up this process, we embed the projection-based model reduction step directly into the pruning procedure.

To compute H_3 at a frequency point, H_1 and H_2 need to be computed at related points. As in (20), to form vector g we can use the reduced order model in (16)-(17) to compute approximate first and second order transfer functions. Then, instead of solving a potentially large linear problem in (19), the reduced adjoint network of (24) can be used to find an approximate vector $\hat{z}$. Computation of each individual contribution term is straightforward, based on (20) and (24). For instance, the contribution of the nonzero at (m, n) location of $\overline{G}_3$ can be simply computed as

$$\delta_{g3, m, n} = \overline{G}_3, m, n \cdot u_{1, n} \cdot \hat{z}_m, \tag{25}$$

where $\overline{G}_3, m, n$ is the value of the nonzero, $u_{1, n}$ is the nth element of u_1, and $\hat{z}_m$ is the mth element of $\hat{z}$. Since reduced order models are used, factorizing the original large system matrices at many sampled points are avoided. The cost for constructing projection-based reduced order models of (16) and (24) is dominated by the cost of a few matrix factorizations, therefore it is bounded by $O(kM^3N^3)$, where k is the number of matrix factorizations, M is the number of sampled points used to discretize the time-varying transfer functions, and N is the number of physical circuit unknowns. The cost of the pruning process is dominated by evaluation and sorting of the different contributions, assuming that the use of reduced order models makes other cost much less. The overall cost for model generation is, therefore, bounded by $O(sElog(E) + kM^3N^3)$, where s is the number frequency samples evaluated in the pruning, and E is the number of nonzeros in the high order matrices being pruned.

Using pruned matrices $\overline{G}_{2p}, \overline{C}_{2p}, \overline{G}_{3p}, \overline{C}_{3p}$, (20) can be further approximated as

$$\hat{g} = -[\overline{G}_{3p} + \tilde{s}\overline{C}_{3p}] \cdot \overline{(V\tilde{H}_1(s_1)) \otimes (V\tilde{H}_1(s_2)) \otimes (V\tilde{H}_1(s_3))} \\ -2 \cdot [\overline{G}_{2p} + \tilde{s}\overline{C}_{2p}] \cdot \overline{(V\tilde{H}_1(s_1)) \otimes (V\tilde{H}_2(s_2, s_3))} \tag{26}$$

To derive a time-domain model for generating the desired third order response, we substitute (26) into (24) and transpose both sides of the equation

$$\hat{y}_{p_i} = \hat{b}_a^T [\hat{G}_a^T + \tilde{s}\hat{C}_a^T]^{-1} V_a^T \hat{g}. \tag{27}$$

Defining

$$\overline{x_{111}} = V\tilde{x}_1 \otimes V\tilde{x}_1 \otimes V\tilde{x}_1, \overline{x_{12}} = \frac{1}{2}(V\tilde{x}_1 \otimes V\tilde{x}_2 + V\tilde{x}_2 \otimes V\tilde{x}_1), \tag{28}$$

it is not hard to show that the corresponding time-domain model is

$$\begin{cases} \frac{d}{dt}(\hat{C}_a^T x_a) + \hat{G}_a^T x_a = \\ -V_a^T\Big(\overline{G}_{3p} \cdot \overline{x_{111}} + \frac{d}{dt}(\overline{C}_{3p} \cdot \overline{x_{111}}) + 2\overline{G}_{2p} \cdot \overline{x_{12}} + 2\frac{d}{dt}(\overline{C}_{2p} \cdot \overline{x_{12}})\Big), \\ \hat{y}_{p, 3}(t) = e^{ji\omega_0 t}\hat{b}_a^T x_a \end{cases} \tag{29}$$

where $\tilde{x}_1, \tilde{x}_2$ are given in (18), $\hat{y}_{p, 3}(t)$ is the desired time-domain third order response. Note that since only the ith harmonics of the third order transfer functions are considered at the output, $\hat{y}_{p, 3}(t)$ in the above equation is complex. To recover the corresponding real signal, we can simply add the conjugate component that corresponds to the -ith harmonic. The first and second order responses at node p, $\hat{y}_{p, 1}(t)$ and $\hat{y}_{p, 2}(t)$, can be obtained from (18) by selecting the proper entries from $\tilde{y}$. Also note that when more than one output or set of side bands are of interest, multiple reduced adjoint networks can be incorporated into the model in a straightforward way.

3.2.3 Flow of Hybrid Model Generation

We summarize the model generation flow in Fig. 2. It should be

1. Apply multi-point NORM to compute a projection V for approximating the first and second order transfer functions:

$$\tilde{G}_1 = V^T \overline{G_1} V, \tilde{C}_1 = V^T \overline{C_1} V, \tilde{b} = V^T \bar{b}$$
$$\tilde{G}_2 = V^T \overline{G_2}(V \otimes V), \tilde{C}_2 = V^T \overline{C_2}(V \otimes V).$$

2. Form the adjoint network as in (22) and reduce it using a projection-based reduction:

$$\hat{G}_a = V_a^T \overline{G_1}^T V_a, \hat{C}_a = V_a^T \overline{C_1}^T V_a, \hat{b}_a = V_a^T b_a.$$

3. For each frequency point $(s_1, s_2, s_3) \in \Omega$ (pre-selected):

   3.1. Compute $\tilde{H}_1(s_1)$, $\tilde{H}_1(s_2)$, $\tilde{H}_1(s_3)$, $\tilde{H}_2(s_1, s_2)$, $\tilde{H}_2(s_1, s_3)$, $\tilde{H}_2(s_2, s_3)$ using (6)-(7) based on reduced matrices $\tilde{G}_1, \tilde{C}_1, \tilde{b}, \tilde{G}_2, \tilde{C}_2$.

   3.2. Approximate the first and second order transfer functions of the original system using (17).

   3.3. Compute the desired sideband of $H_3(t, s_1, s_2, s_3)$ at the output node, and individual contributions of nonzeros in $\overline{G_2}, \overline{C_2}, \overline{G_3}, \overline{C_3}$ based on reduced order models using (20), (24) and (25).

   3.4. Prune $\overline{G_2}, \overline{C_2}, \overline{G_3}, \overline{C_3}$ by sorting all the per-nonzero contributions and retaining dominant ones to satisfy an error tolerance ε.

4. Form the time-domain reduced models using (18) and (29) if needed.

Fig. 2. Nonlinear hybrid macromodel generation

noted that the hybrid model can be either realized in the time-domain based on (18) and (29), or the reduced, pruned matrices and associated projections $\tilde{G}_1$, $\tilde{C}_1$, $\tilde{b}$, $\tilde{G}_2$, $\tilde{C}_2$, V, $\hat{G}_a$, $\hat{C}_a$, $\hat{b}_a$, V_a, $\overline{G}_{2p}, \overline{C}_{2p}$, $\overline{G}_{3p}$ and $\overline{C}_{3p}$ can be directly used in a frequency-domain Volterra-based analysis. The corresponding model structure is depicted in Fig. 3.
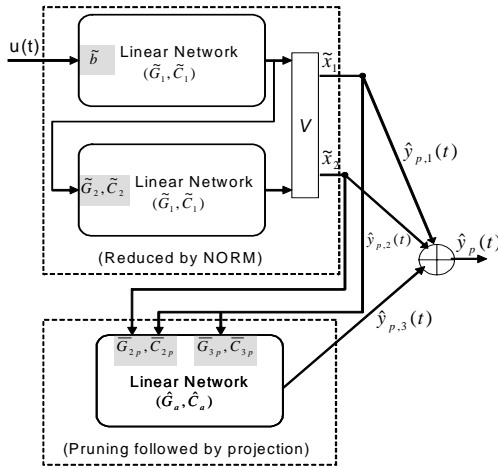

Fig. 3. Hybrid model structure

4. RESULTS

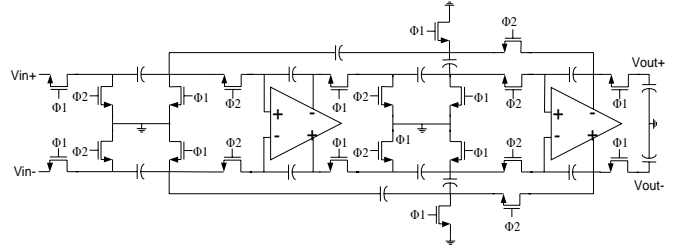### 4.1 A switch-capacitor filter

Fig. 4. A switch-capacitor biquad low-pass filter

Switch-capacitor filters are often found in RF receivers as channel-select filters [22]. If the input signal is small, then these circuits can be characterized by time-varying Volterra series. A Butterworth lowpass switch-capacitor biquad is shown in Fig. 4. The two-phase clock is at 20MHz, and the 3-dB frequency of the filter is about 700KHz. Each circuit nonlinearity in the filter is modeled as a third order polynomial about the periodically varying operating point generated by the clock. The resulting full model has 8142 time-sampled circuit unknowns. To view the third order nonlinear effects within and close to the signal band, we plot the dc component of the time-varying third order nonlinear transfer function $H_3(t, j2\pi f_1, j2\pi f_2, j2\pi f_3)$, where $1Hz \le f_1, f_2 \le 10MHz$, $f_3 = -1Hz$, in Fig. 5(a). It is clearly observable that the third order nonlinear characteristics vary dramatically between passband and stopband. To capture not only the nonlinear distortions due to the
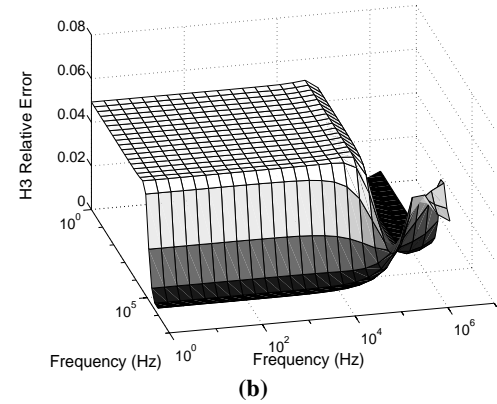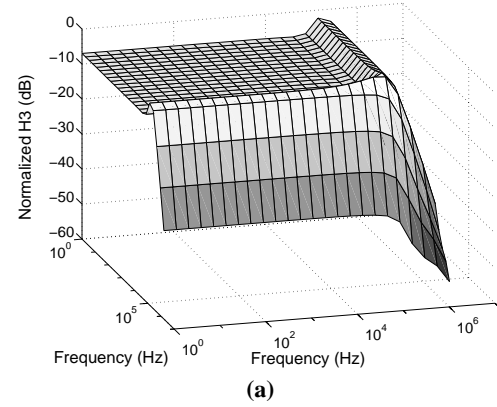

(a)

(b)

Fig. 5. (a) Original H3 of the biquad, and (b) relative error of H3

signals within the passband, but also the third order mixing of the large out of channel interferers into the passband, the nonlinear frequency-domain characteristics of the filter must be modeled accurately over a frequency range of 7 decades. As such, a complete projection-based approach becomes inefficient, since a sufficient number of moments have to be matched for accuracy, while the resulting model size leads to expensive dense projected third order matrices.

To apply our hybrid approach, we first used the multi-point NORM algorithm to accurately capture the first and second transfer functions using a reduced order model with 27 states (SVD was used to deflate the Krylov subspaces), where 6 moments of H_1 and 24 moments of H_2 were matched. The adjoint network describing the propagation of third order nonlinear effects from various nonlinearities to the output was reduced to an 11th order model. Based on these reduced models, $\overline{G_2}, \overline{C_2}, \overline{G_3}, \overline{C_3}$ were significantly pruned in the original coordinates on a set of sampled frequencies points. The final hybrid model has a maximum relative modeling error of 6% (or about 0.5dB) for H_3, as shown in Fig. 5(b).

This hybrid model was also validated in a frequency-domain Volterra-like analysis, where 6 sinusoidals with various phases were selected from passband, transition band and stopband as input signals. The simulation result was compared against that of the full model, as shown in Fig. 6. As can be seen from the figure, the hybrid model captures the frequency-domain nonlinear characteristics of the filter over a wide range of input band very accurately. In additional to excellent accuracy, for this example, a runtime speed up of 64x over the full model was achieved by using the hybrid model in simulation.
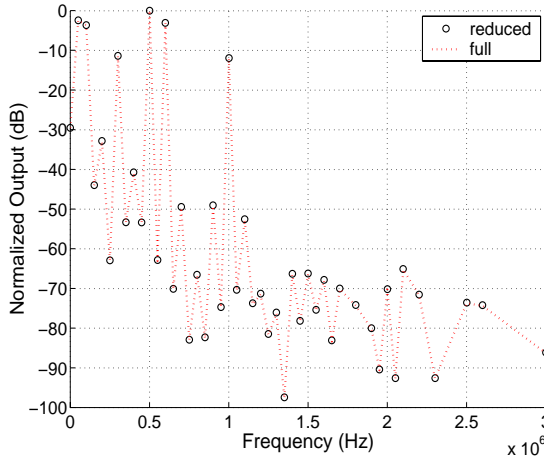

Fig. 6. Biquad output due to a six-tone sinusoidal input

4.2 A 900MHz Receiver Front-End

Next consider the 900MHz receiver frond-end depicted in Fig. 7. It is composed of a LNA, a mixer and an IF amp as well as high-Q bandpass and lowpass filters. To model the weakly nonlinear behavior of the receiver with respect to the RF input, we extracted a time-varying third order polynomial nonlinear model
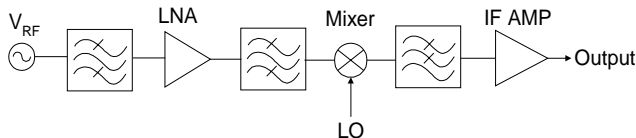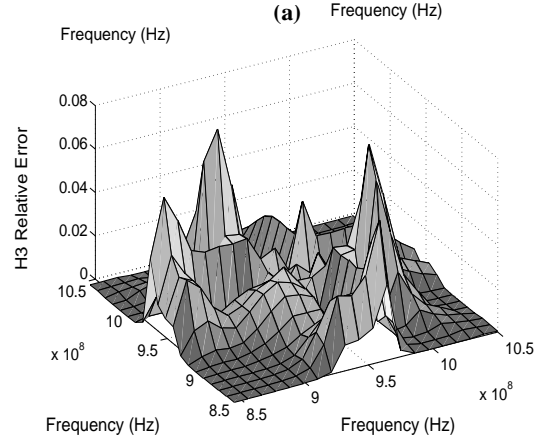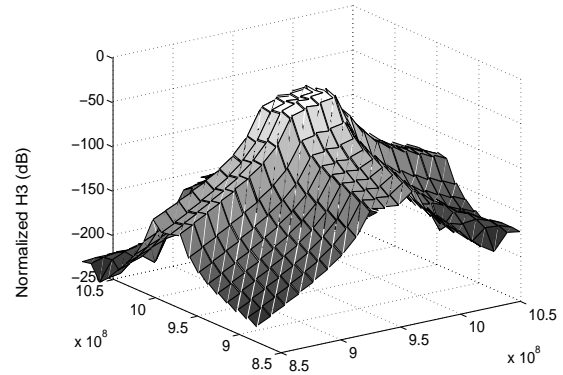

Fig. 7. A 900MHz receiver front-end

(a) Frequency (Hz)

(b)

Fig. 8. (a) Original H3 of the front-end, and (b) relative error of H3

of the system upon the varying operating point due to the 880MHz LO. The extracted full model has 8496 unknowns. To see the third order intermodulation effects of the system, we plot the first harmonic of the time-varying third order nonlinear transfer function $H_3(t, j2\pi f_1, j2\pi f_2, j2\pi f_3)$ of the output about the RF center frequency in Fig. 8(a). This harmonic specifies the amount of downconverted (by one LO frequency) third order intermodulation (IM3) at the output. In the figure, $850Hz \leq f_1, f_2 \leq 1050MHz$, $f_3 = -950MHz$.

As high-Q filterings are included as part of the system, due to their high frequency selectivity, H_3 varies dramatically about the center frequency. To account fully for the receiver distortion performance under the presence of large out-of-band interferers (e.g. large blocking signals), the receiver third order intermodulation must be characterized accurately not only in the channel, but also at frequencies tens or hundreds of MHz away from the center frequency. This requirement makes it very difficult to directly apply a projection-based reduction to the complete receiver, and may necessitate modeling each nonlinear block individually in order to isolate the high-Q filtering blocks.

To apply the proposed hybrid approach to the complete receiver, multi-point NORM was first applied to approximate H_1 and H_2 using a reduced order system with 52 states, where 16 moments of H_1 and 44 moments of H_2 are matched (the Krylov subspaces were compacted using SVD). The adjoint network for propagating third-order nonlinear effects to the first harmonic of time-varying H_3 at the output was reduced to a 40th order system. The second and the third order system matrices were again pruned on a set of

sampled frequency points for meeting the accuracy requirement for H_3 based on these reduced models. The corresponding relative modeling error on H_3 of the hybrid model is less than 8% (0.7dB) as plotted in Fig. 8(b). Since H_3 varies as much as 230dB in the frequency band shown in Fig. 8, the relative modeling error was computed as $\delta = \dfrac{\left| H_3^{ori} - H_3^{app} \right|}{\left| H_3^{ori} \right| + \alpha}$, where α is a small positive number that is added to avoid overestimation of the error for an extremely small value of H_3^{ori}.

To further verify the accuracy of the receiver model, we performed a frequency-domain Volterra-based analysis where the RF input consisted of 12 sinusoidals over a frequency span of 100MHz. To test the receiver performance under the interference of large blocking signals, the largest amplitude difference between the in-band and out-of-band signals was as high as 60dB. The complete output spectrum consisted of 147 frequencies, and the frequency components close to the IF band are plotted in Fig. 9. As clearly seen from the plot, the hybrid model accurately captures the receiver linearity performance over a wide range of input frequency. In addition, the use of the hybrid model brought more than 20x simulation efficiency improvement over the full model. We shall also note that model accuracy can be easily traded off with simulation efficiency for specific cases. For instance, as the accuracy requirement permits, simpler hybrid models can be constructed by adopting lower order projection-based approximations, as well as less stringent error tolerance for pruning.
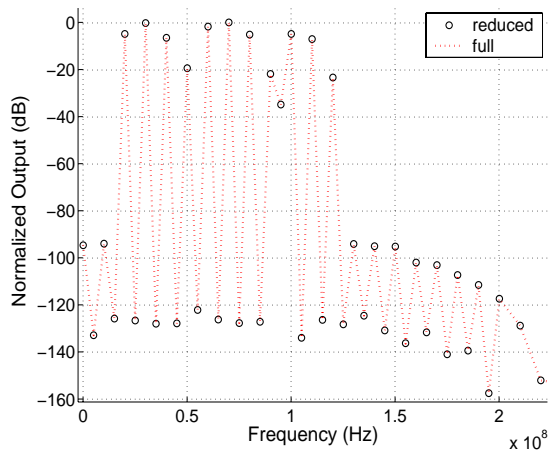

Fig. 9. Front-end receiver output due to a 12-tone sinusoidal input

5. CONCLUSIONS

We have shown that generating fully-encompassing nonlinear models for complex analog circuit blocks or subsystems presents challenging modeling problems. For these cases, the requirements on model accuracy and applicability can often render existing methods ineffective. To address these challenges, a combination of different techniques are required. Our proposed hybrid model generation approach combines the benefits of both pruning-based and transformation-based methodologies into a unifying model. The utility of this approach is demonstrated for modeling circuits in analog signal processing and RF applications.

6. REFERENCES

[1] D. Wiener and J. Spina, *Sinusoidal analysis and modeling of weakly nonlinear circuits*, Van Nostrand Reinhold, 1980.
[2] S. A. Maas, *Nonlinear Microwave Circuits*, Artech House, Norwood, Massachusetts, 1988.
[3] G. Gielen and R. Rutenbar, "Computer-aided design of analog and mixed-signal integrated circuits," *Proc. of IEEE*, vol. 88., no. 12, December 2000.
[4] L. Pillage and R. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. CAD*, vol. 9, pp. 352-366, April 1990.
[5] P. Feldmann and R. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Trans. CAD*, vol. 14, pp. 639-649, May 1995.
[6] L. Silveira, M. Kamon and J. White, "Efficient reduced-order modeling of frequency-dependent coupling inductances associated with 3-d interconnect structures," *Proc. of ACM/IEEE DAC*, 1995.
[7] A. Odabasioglu, M. Celik and L. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. CAD*, vol. 17, no. 8, pp. 645-653, August 1998.
[8] J. Roychowdhury, "Reduced-order modeling of time-varying systems," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 10,Oct., 1999.
[9] J. Phillips, "Projection frameworks for model reduction of weakly nonlinear systems," *Proc. of ACM/IEEE DAC*, 2000.
[10] J. Phillips, "Automated extraction of nonlinear circuit macromodels," *Proc. of IEEE CICC*, 2000.
[11] J. Phillips, "Model reduction of time-varying linear systems using approximate multipoint krylov-subspace projectors," *Proc. of ACM/IEEE ICCAD*, 1998.
[12] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *Proc. of ACM/IEEE ICCAD*, 2001.
[13] P. Li and L. Pileggi, "NORM: compact model order reduction of weakly nonlinear systems," *Proc. of ACM/IEEE DAC*, 2003.
[14] P. Wambacq and W. Sansen, Distortion analysis of analog integrated circuits, Kluwer Academic Publishers, 1998.
[15] P. Wambacq, P. Dobrovolny, S. Donnay, M. Engels and I. Bolsens, "Compact modeling of nonlinear distortion in analog communication circuits," *Proc. of ACM/IEEE DATE*, 2000.
[16] P. Dobrovolny, G. Vandersteen, P. Wambacq, S. Donnay, "Analysis and white-box modeling of weakly nonlinear time-varying circuits," *Proc. of ACM/IEEE DATE*, 2003.
[17] H. Liu, A. Singhee, R. Rutenbar and L. Carley, "Rememberance of circuits past: macromodeling by data mining in large analog design spaces," *Proc. of ACM/IEEE DAC*, 2002.
[18] W. Daemes, G. Gielen and W. Sansen, "An efficient optimization-based technique to generate posynomial performance models for analog integrated circuits," *Proc. of ACM/IEEE DAC*, 2002.
[19] Matlab Simulink Reference, http://www.mathworks.com/.
[20] M. Okumura, T. Sugawara and H. Tanimoto, "An efficient small-signal frequency analysis method of nonlinear circuits with two frequency excitations," *IEEE Trans. CAD*, vol. 9, no. 3, pp. 225-234, March 1990.
[21] A. Ruhe, "The rational Krylov algorithm for nonsymmetric eigenvalue problems III: complex shifts for real matrices,"*BIT*, vol. 34, pp. 165-176, 1994.
[22] P. Chang, A Rofougaran and A. Abidi, "A CMOS channel-select filter for a direct-conversion wireless receiver," *IEEE J. of Solid-state Circuits*, vol. 32. no. 5, pp. 722-729, May 1997.
[23] X. Li, P. Li, Y. Xu and L. Pileggi, "Analog and RF circuit macromodels for system-level analysis," *Proc. of ACM/IEEE DAC*, 2003.
[24] N. Dong and J. Roychowdhury, "Piecewise polynomial nonlinear model reduction," *Proc. of ACM/IEEE DAC*, 2003.