A Linear-Centric Modeling Approach to Harmonic Balance Analysis†

Peng Li and Lawrence Pileggi
Department of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{pli, pileggi}@ece.cmu.edu

Abstract

In this paper we propose a new harmonic balance simulation methodology based on a linear-centric modeling approach. A linear circuit representation of the nonlinear devices and associated parasitics is used along with corresponding time and frequency domain inputs to solve for the nonlinear steady-state response via successive chord (SC) iterations. For our circuit examples this approach is shown to be up to 60x more run-time efficient than traditional Newton-Raphson (N-R) based iterative methods, while providing the same level of accuracy. This SC-based approach converges as reliably as the N-R approaches, including for circuit problems which cause alternative relaxation-based harmonic balance approaches to fail[1][2]. The efficacy of this linear-centric methodology further improves with increasing model complexity, the inclusion of interconnect parasitics and other analyses that are otherwise difficult with traditional nonlinear models.

1 Introduction

As circuits increase in complexity and models of high-frequency parasitics become more detailed, there is an ever-increasing demand for more efficient steady-state circuit response analyses in the analog and RF design community. Harmonic balance methods provide a more efficient frequency domain steady-state analysis than traditional time domain transient analysis for such problems, particularly when the circuit has widely spread time constants[1][2][13]. However, accuracy requirements and aliasing tolerances require that strongly nonlinear circuits are simulated at a large number of frequencies. Consequently, the problem size can render harmonic methods to be impractical, particularly when the nonlinearities are severe.

Krylov-subspace iterative methods have been proposed to solve large linear equations at each iteration to improve the overall efficiency in harmonic balance methods [11][13], and also for time domain shooting methods [12]. In [11] and [13] a preconditioning matrix is required to achieve good convergence speed on the solutions of linear equations. However, simple low-cost block diagonal preconditioners become ineffective when the circuits contain severe nonlinearities and a more complicated adaptive preconditioning scheme is needed [14].

In this paper we outline a methodology for significant improvement in harmonic balance simulation efficiency via a simple linear-centric modeling approach. Each nonlinear element is modeled by a constant linear representation in parallel with a time-varying current source. The successive chord iterative method is then used to solve the harmonic balance nonlinear system of equations that are formulated by these linear-centric models. The successive chord iterative method has been recently applied for time domain transistor-level timing analysis (TETA) in [3][4][5][6].

For the simplicity of terminology, we refer to our new harmonic balance approach as *SC balance*. In contrast to Newton-Raphson based method and the relaxation variants, for which a nonlinear circuit is linearized at every iteration step, the nonlinear circuit is linearized once, prior to the start of the analysis, in SC balance. Newton-Raphson is generally preferred to successive chords due to its *theoretical* quadratic convergence rate. However, such a convergence rate is not always guaranteed in practice due to the step-limiting or damping scheme that is generally employed to ensure the stability.

Our simple linear-centric modeling approach facilitates the use of successive chord method, which has a linear rate of convergence. But while reaching the convergence can require more iterations, the complexity of each iteration is much smaller than that of N-R method. The improvement of simulation efficiency is more pronounced for large circuits since the sparse structure of the problem can be further exploited in the adopted SC method. Moreover, adding interconnect parasitic models and other analyses such as noise, are far more readily accommodated by this linear-centric framework.

This paper is organized as follows. A brief review of successive chord iterative method is given in Section 2. In Section 3, we introduce the circuit models used in our approach. The linear-centric harmonic balance approach with comparison to traditional N-R approach is presented in section 4. Experimental results with comparisons of various iterative methods are presented in Section 5. Finally, con-

† This work was supported by the Semiconductor Research Corporation (SRC) under contract 2000-TJ-779.

clusions are drawn and future research directions are discussed in Section 6.

2 The Basics of Successive Chord Method

Successive chord method is described as a variant of classical Newton-Raphson iterative method in [7][8]. As an example, a simple diode circuit shown in Fig. 1(a) is used to compare the Newton-Raphson method and the successive chord method in Fig. 1(b) and Fig. 1(c) respectively. In Newton-Raphson iterations, nonlinear circuit elements are linearized according to their tangents at each iteration step, while they are linearized using a constant slope in successive chord iterations. These constant linearizations are referred to as *chords*.

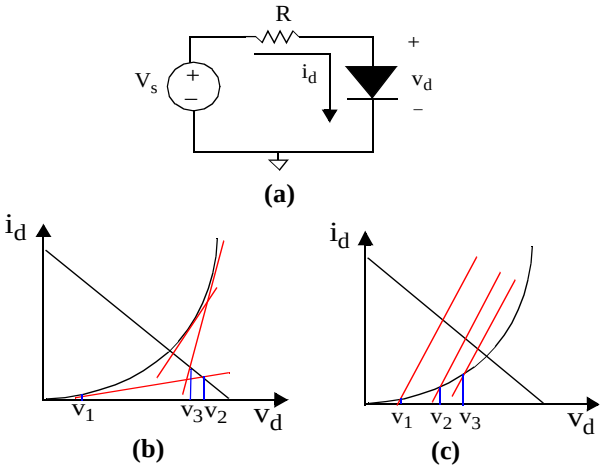


Fig. 1: (a) simple diode circuit, (b) Newton-Raphson Method, and (c) successive chord method.

There are a few differences between these two methods as applied to nonlinear circuit analyses that are worth noting here. Firstly, successive chord method implicitly uses a constant linear circuit model for each nonlinear element for all iterations while Newton-Raphson employs a changing linearized model for every nonlinear element from iteration to iteration. Secondly, Newton-Raphson iterative method achieves theoretical quadratic convergence rate while successive chord method converges linearly. However, successive chord may outperform Newton-Raphson in terms of overall runtime by iterating more steps but spending much less time per step. Moreover, for highly nonlinear circuit problems, damping is required for N-R iterations to converge, and the rate of convergence is superlinear, but not quadratic in practice.

3 Linear-Centric Modeling of Nonlinear Circuit Elements

The successive chord iterative method implicitly uses a constant linear circuit model for each nonlinear element. In this section, circuit models of MOSFETs and nonlinear capacitors are used to demonstrate our linear-centric

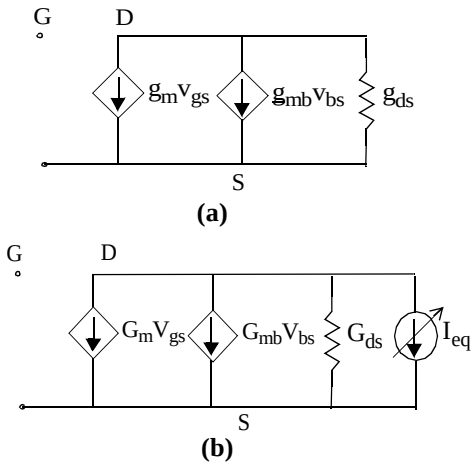


Fig. 2: (a) linearized small-signal MOSFET I_{DS} model, and (b) linearized large-signal MOSFET I_{DS} model

approach. Without loss of generality, this approach can be applied to other nonlinear devices such as nonlinear inductors. Note that we restrict our discussions to nonlinearities which have an admittance representation such as voltage-controlled nonlinear resistors. The linear-centric models for an impedance representation are simply the dual form to the aforementioned linear-centric models.

A typical small-signal low-frequency MOSFET model that is commonly used to perform small-signal AC analysis at an operating point is shown in Fig. 2(a). With the addition of an extra current source this model is used in Newton-Raphson to linearize MOSFETs at each iteration in a nonlinear analysis. Importantly, the linear resistors and controlled sources change with every N-R iteration.

The linearized large-signal MOSFET I_{DS} model used in our SC approach is shown in Fig. 2(b). Unlike the small-signal model, this model uses fixed values for transconductances and conductances for all biasing conditions. The only difference is a varying current source is added such that I_{DS} of the MOSFET is equal to the value dictated by the device model for every operating point. In this paper, we refer to this varying current source in a linearized large-signal model as a *chord current source*.

The linearized large-signal nonlinear capacitance model employed in our linear-centric approach is shown in Fig. 3. Each nonlinear capacitance is simply modeled by a constant linear capacitance in parallel with a varying current source. It should be noted that this model represents a frequency domain perspective for harmonic balance analysis.

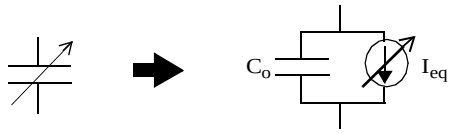


Fig. 3: linearized large-signal nonlinear capacitance model.

A linear-centric treatment for time domain transient analysis will result in a somewhat different *companion model* for nonlinear capacitances[9].

Combining the above MOSFET model and the nonlinear capacitance model, the complete linear-centric MOSFET model that is used in our prototype simulator is shown in Fig. 4, where each nonlinear element is modeled by a linearized element (chord) and a varying chord current source. Some subtle differences exist when a charge conservation gate capacitance model such as the one in [10] is used. In this case, small-signal nonlinear gate capacitances become non-reciprocal, f.g., $\frac{\partial Q_G}{\partial V_B} \neq \frac{\partial Q_B}{\partial V_G}$. Under this situation, capacitive parasitics cannot be simply modeled by two-terminal capacitors in Fig. 4(a), and more general concept of transcapacitances applies. An exact linear-centric approach will generate a more complicated model than the one shown in Fig. 4(b). However, the conceptually more intuitive model in Fig. 4(b) can be used even for charge conservation capacitance models since the exact circuit responses will still be reached upon convergence.

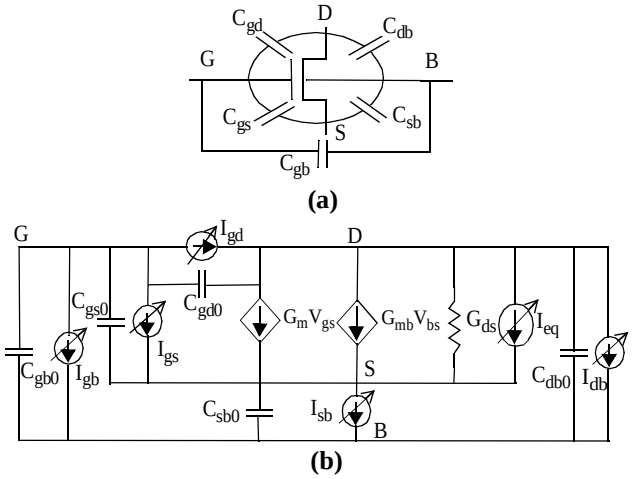


Fig. 4: (a) A MOSFET, and (b) linear-centric MOSFET model.

4 Linear-Centric Harmonic Balance

4.1 Newton-Raphson Method

To solve the steady-state response of a circuit, harmonic balance performs MNA analysis in frequency domain. If we denote the number circuit unknowns as N, then nonlinear system equations of dimension of $MN \times MN$ on a truncated set of frequencies $\{\omega_1, \omega_2, ..., \omega_M\}$ can be expressed as [2]:

$$F(V) = I(V) + \Omega Q(V) + YV + U = 0, \qquad (1)$$

where, in terms of Fourier coefficients, V, $I(V)$, $Q(V)$, and U (all $\in C^{MN}$) are vectors of system unknowns, currents of nonlinear resistors, nodal charges(fluxes), and input excitations, respectively. $Y \in C^{MN \times MN}$ is the admittance matrix composed of linear circuit element. It is worth noting that (1) is usually more efficiently formulated in terms of real parts and imaginary parts of Fourier coefficients as described in [2]. $\Omega \in C^{MN \times MN}$ is the frequency differentiation operator used to convert nonlinear capacitor charges to the corresponding currents:

$$\Omega = \begin{bmatrix} j\omega_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & j\omega_M \end{bmatrix}, \qquad (2)$$

where, $j\omega_k$ is a block-diagonal matrix of dimension of $N \times N$ representing the differentiation operator at frequency ω_k. The linearizations of nonlinear resistors and capacitors at sampled time points can be represented by following two block-diagonal matrices, respectively:

$$G = \begin{bmatrix} \left.\frac{\partial i}{\partial v}\right|_{t_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \left.\frac{\partial i}{\partial v}\right|_{t_M} \end{bmatrix}, \qquad (3)$$

$$C = \begin{bmatrix} \left.\frac{\partial q}{\partial v}\right|_{t_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \left.\frac{\partial q}{\partial v}\right|_{t_M} \end{bmatrix}. \qquad (4)$$

When Newton-Raphson is used to solve (1), the Jacobian is in the form [11][13]:

$$J_F = Y + \Omega\Gamma C\Gamma^{-1} + \Gamma G\Gamma^{-1}, \qquad (5)$$

where, Γ and Γ^{-1} are the discrete Fourier transform and the inverse discrete Fourier transform matrices, respectively. Notice that matrices G and C are usually sparse, but the second and third terms in (5) are dense, especially for highly nonlinear circuits. The dense Jacobian matrix of Newton-Raphson method makes the direct solution of the linear system via Gaussian Elimination at each iteration impractical for large circuits. Krylov-subspace based iterative method has been proposed to solve large the linear system at each Newton-Raphson iteration [11][13]; however, to ensure proper converge rate of linear system solutions, a preconditioning matrix is usually required. When large nonlinearities are encountered, the simple low-cost preconditioner obtained via discarding off-diagonal blocks in (5) is no longer effective[14].

4.2 SC Balance

In our linear-centric modeling approach to harmonic balance method, we employ the linearized large-signal models described in Section 3. For a circuit such as the one

shown in Fig. 5(a), that includes nonlinear devices such as MOSFETs and nonlinear capacitors, we replace each nonlinear device by its linearized large-signal model. In general, such a linearized model is composed of linear R's, L's C's and controlled sources (chords) plus a varying chord current source. By combining the chords of nonlinear elements with other linear elements in the circuit into a collapsed invariant linear circuit, we reach the overall circuit model used in our linear-centric approach, as shown in Fig. 5(b).

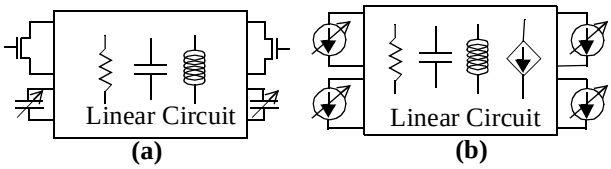


Fig. 5: (a) a circuit containing nonlinear devices, and (b) overall circuit representation used in the linear-centric approach.

As one can expect, such a circuit representation appears most intuitive for "almost linear" circuits. For these well-behaved circuits, the chords of nonlinear elements can be simply chosen according to their operating point small-signal parameters, such as small-signal conductances, transconductances and small-signal capacitances etc. In this case, a chord current source would readily represent the *nonlinear distortion* caused by the corresponding nonlinear element. Importantly, chords need not necessarily be chosen according to dc operating point small-signal values. In practice they are chosen specifically to ensure convergence for strongly nonlinear circuits -- more on this later.

From a circuit simulation perspective, the adoption of the circuit model in Fig. 5(b) is equivalent to simulating the nonlinear circuit using successive chord iterations in harmonic balance. To see this connection more clearly, we apply the constant linearized large-signal model for each nonlinear element at every sampled time points, then equations (3) and (4) become

$$G = \begin{bmatrix} G_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & G_0 \end{bmatrix} \tag{6}$$

and

$$C = \begin{bmatrix} C_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_0 \end{bmatrix}, \tag{7}$$

respectively. In (6) and (7), G_0 and C_0 are the time-invariant linearizations of nonlinear resistors and capacitors comprised of chords, respectively. Substituting (6) and (7) into (5), we obtain

$$J_{SC} = \begin{bmatrix} Y(j\omega_1) + G_0 + j\omega_1 C_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Y(j\omega_M) + G_0 + j\omega_M C_0 \end{bmatrix}. \tag{8}$$

Recognizing that J_{SC} in (8) is a constant matrix, we apply the following successive chord iterations to solve for system unknowns

$$J_{SC}(\hat{V}_{i+1} - \hat{V}_i) = -F(\hat{V}_i), \tag{9}$$

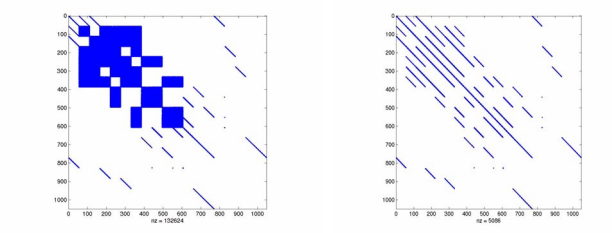
or

$$J_{SC}\hat{V}_{i+1} + \left(I(\hat{V}_i) + \Omega Q(\hat{V}_i) - \begin{bmatrix} G_0 + j\omega_1 C_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & G_0 + j\omega_M C_0 \end{bmatrix} \hat{V}_i \right) + U = 0 \tag{10}$$

In (10), J_{SC} represents the constant linearized circuit, and the second term is simply the DFT of the sum of chord currents of nonlinear resistors and capacitors. Based upon (5), (8) and (10), our linear-centric approach presents the following important advantages:

- To construct the Jacobian matrix in (5), one needs to evaluate not only the currents (charges) of nonlinear resistors (capacitors), but also the derivatives of currents (charges) w.r.t. terminal voltages on which they depend at each sampled time point. This also requires the continuity of the derivatives to ensure the stability of N-R iterations. This limitation demands the use of smoothing functions wherever there is a discontinuity of these derivatives. However, a constant Jacobian matrix as in (8) is used for all iterations in our linear-centric approach such that there is no need to evaluate any derivative. Only the currents (charges) themselves need to be evaluated. Hence, our approach allows the use of lookup table based device models, in which sensitivities are not directly required [3][4][5][6]. In the detailed device model implemented in our prototyped simulator, N-R method needs to evaluate 17 derivatives for a MOSFET and its nonlinear capacitive parasitics at each sampled time point. Then, 17 DFTs are required to transform these 17 derivatives into frequency domain. Saving in the model evaluation using our linear-centric approach would be more significant as more complicated device models are used.
- While (5) is usually dense for circuits with large nonlinearities, (8) is still sparse. The Jacobian matrices of two methods for a double-balanced mixer are shown in Fig. 6. Fig. 6(a) has 132,624 non-zero entries compared to 5,086 non-zeros in Fig. 6(b). The sparsity of constant Jacobian matrix in (8) can be exploited to solve large nonlinear circuits efficiently. Prior to the start of the iteration loop, (8) can be prefactorized via LU factorization and stored sparsely. Thereby a new solution can be solved very efficiently though simple forward and backward substitutions using the constant sparse LU factorization at each iteration. The overall procedure is

equivalent to pre-solve a linear circuit, then update the circuit response by changing the excitements to the linear circuit at each iteration step. Finding responses to new excitements only requires the re-evaluation of the chord currents for the nonlinear elements. Moreover, due to the fact there is no frequency translation in (8), the linear system solution at each SC balance iteration can be also solved individually for each frequency. In this case, a total of M smaller linear circuits are solved, each of which is of dimension of N x N.



(a) **(b)**

Fig. 6: (a) Jacobian of N-R, and (b) Jacobian of SC.

- As shown in Fig. 5(b), we model a nonlinear circuit by a constant linear circuit plus chord current sources. The linear circuit consists of linear circuit elements and the chords of nonlinear devices. Interconnect model order reduction techniques [15][16] can be extended to handle controlled sources such that a reduced linear circuit can be used in the successive chord iterations. We expect this model order reduction will further increase the simulation efficiency of our linear-centric approach.

When a circuit operates almost linearly, the chords of all nonlinear elements can be simply chosen as their small-signal operating point values. The convergence and good speed can be guaranteed easily. In [5][6], to ensure the convergence of successive chord iterations for possible large ranges of operating points, large sensitivities of the drain source current of a MOSFET w.r.t. terminal voltages are chosen as the chord values for the MOSFET. For example, g_{ds} is smaller in the saturation region compared to its value in the linear region. Therefore, the chord for g_{ds} should be determined according to the value in the linear region. For a more detailed discussion on the convergency property of successive chord method, please refer to [5][7].

In general, choosing large sensitivities leads to a slower update in the solution per step while it helps to prevent divergence of the iterations. In SC balance, we employ essentially the same approach. In addition to considering the impact of the nonlinear I_{DS} characteristics of MOSFETs, we also include the contributions of nonlinear parasitic capacitances to the Jacobian matrix. Our experiments have demonstrated that neglecting nonlinear capacitances in (8) can easily lead to divergence, especially for high-speed circuits. The chord for a junction capacitance is chosen as the small-signal capacitance value under zero-biased condition because it is the largest sensitivity of the junction charge. Similarly, the chord values for C_{gd} and C_{gs} (the fixed overlap capacitances are considered separately) are approximated by capacitance values in the linear or saturation region considering certain channel charge partitioning parameters used in the gate capacitance model. The largest sensitivity regarding to the nonlinear component of C_{gb} can be estimated by the gate capacitance in the accumulation region.

5 Experimental Results

In this section we compare our implementations of three iterative methods, N-R, block Gauss-Jacobi-Newton(GJN) relaxation and SC on several benchmark circuits. Block Gauss-Jacobi-Newton is a relaxation method of N-R. It is equivalent to neglecting frequency translations in the Jacobian of N-R method by discarding all the off diagonal blocks [2]. It can speed up the factorization of Jacobian matrix, however, it might lead to divergence when large nonlinearities are encountered. We compare these three methods on three double-balanced mixers. To make block GJN converge, circuit nonlinearities are limited by using small signal amplitudes for local oscillators. A level three MOSFET model is used for all cases.

The results are summarized in Table 1. The second column shows the problem size, and the columns that follow are number of iterations for convergence, and CPU time in seconds for the three methods. As shown in the table, SC method runtime is superior in all cases, and even outperforms block Guass-Jacobi-Newton relaxation considerably.

Circuit	Eqns	It-NR	Ctime-NR	It-GJN	Ctime-GJN	It-SC	Ctime-SC
dbmixer-A	668	8	252.4	8	19.8	10	6.1
dbmixer-B	1200	8	845.9	11	49.7	13	14.1
dbmixer-C	1542	9	1,584	12	71.4	10	14.0

Table 1: Comparison on weakly nonlinear circuits

In Table 2, more strongly nonlinear circuits are considered: switching single-balanced mixers, switching double-balanced mixers and switching double-balanced mixers driven by an LNA. For the circuits in Table 2, block Gauss-Jacobi-Newton fails to converge. Therefore, we only compare N-R method and SC method. As seen from the results in this table, our approach has significant speedups over N-R method, and we expect the improvement of simulation efficiency to be even greater for larger problems.

Circuit	Eqns	It-NR	It-SC	Ctime-NR	Ctime-SC	Speedup
sbmixer-A	1041	12	247	675	136.7	4.9
sbmixer-B	1905	10	386	2,066	379.8	5.4
sbmixer-C	2449	10	386	3,545	489.9	7.2
dbmixer-A	1238	13	102	1,423	113.2	12.6
dbmixer-B	1618	169	356	30,802	503.3	61.2
LNA+dbmixer-A	1891	8	79	3,227	129.2	25.0
LNA+dbmixer-B	3051	16	221	30,464	621.8	49.0

Table 2: Comparison on strongly nonlinear circuits

6 Conclusions and Future Directions

In this paper we have presented a novel and efficient harmonic balance approach for the steady-state solutions of analog circuits. The proposed linear-centric modeling approach has several advantages. First, it allows much simpler device model evaluations and time-to-frequency domain transforms than traditional N-R based approaches. More importantly, it represents a nonlinear circuit by a constant linearized circuit driven by circuit inputs and varying chord current sources of nonlinear elements. Thereby, successive chord method is adopted to efficiently solve the steady-state response of the nonlinear circuit via iterative solutions of the constant linear circuit. The sparsity of the linear circuit is exploited to handle large problem size effectively.

Our future research directions include more theoretical study on the convergence of SC balance. Moreover, due to the linear-centric device models, the use of model order reduction techniques for the overall linearized circuit can potentially improve the simulation efficiency further.

7 Acknowledgments

The authors would like to thank Emrah Acar for the many helpful discussions regarding successive chords.

8 References

[1] K. Kundert and A. Sangiovanni-Vincentelli, "Finding the steady-state response of analog and microwave circuits," *Proc. of IEEE Custom Integrated Circuits Conf.*, 1988.

[2] K. Kundert, J. White and A. Sangiovanni-Vincentelli, "Steady-state methods for simulating analog and microwave circuits," *Kluwer Academic Publishers*, Boston 1990.

[3] F. Dartu and L. Pileggi, "TETA: transistor-level engine for timing analysis," *Proc. of 35thDAC*, 1998.

[4] F. Dartu, "Gate and transistor level waveform calculation for timing analysis," *Ph.D. dissertation*, Dept. of Electrical and Computer Engineering, Carnegie Mellon University, 1997.

[5] E. Acar, "Linear-centric approach for timing analysis," *Ph.D. dissertation*, Dept. of Electrical and Computer Engineering, 2001.

[6] E. Acar, F. Dartu and L. Pileggi, "TETA: transistor-level waveform evaluation for timing analysis," to appear on *IEEE. Trans. on CAD*.

[7] J. Ortega and W. Rheinboldt, "Iterative solution of nonlinear equations in several variables," *Academic Press*, 1970.

[8] W. McCalla, "Fundamentals of computer-aided circuit simulation," *Kluwer Academic Publisher*, 1988.

[9] L. Pillage, R. Rohrer and C. Visweswariah, "Electronic circuit & system simulation methods," *McGraw-Hill*, 1995.

[10] D. Ward and R. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE Journal of Solid-state Circuits*, vol. sc-13, no. 5, October 1978.

[11] P. Feldman, B. Melville, D. Long, "Efficient frequency doman analysis of large nonlinear analog integrated circuits," *Proc. of IEEE Custom Integrated Circuits Conf.*, pp. 461-464, May 1996.

[12] R. Telichevesky, K. Kundert and J. White, "Efficient steady-state analysis based on matrix-free Krylov-subspace methods," *Proc. of 32ndDAC*, 1995.

[13] R. Telichevesky, K. Kundert, I. Elfadel and J. White, "Fast simulation algorithms for RF circuits," *Proc. of IEEE Custom Integrated Circuits Conf.*, pp. 437-444, 1996.

[14] M.Gourary et. al., "The enhancing of efficiency of the harmonic balance analysis by adaptation of preconditioner to circuit nonlinearity," *Proc. of ASP-DAC*, pp. 537-540, 2000.

[15] K. Kerns and A. Yang, "Stable and efficient reduction of large, multiport RC networks by pole analysis via congruence transformations," *IEEE Trans. CAD*, vol. 16, 1997.

[16] A. Odabasioglu, M. Celik and L. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. CAD*, August 1998.