Relaxation-Based Harmonic Balance Technique for Semiconductor Device Simulation

Boris Troyanovsky, Zhiping Yu, Lydia So, and Robert W. Dutton
Center for Integrated Systems
Stanford University, Stanford, CA 94305

Abstract

Harmonic and intermodulation distortion effects play an important role in numerous analog applications, particularly in such areas as wireless communication systems. In this paper, we present a two-dimensional harmonic balance semiconductor device simulator which accurately models these nonlinear effects at the physical (drift-diffusion) level. The simulator is based on Stanford University's PISCES code, and supports the full range of physical models and features present in the time-domain version of the program. A modified block Gauss-Seidel-Newton nonlinear relaxation scheme is developed to efficiently handle the extremely large size of two-dimensional harmonic balance semiconductor device simulation problems.

1 Introduction

Harmonic and intermodulation distortion generated by nonlinear semiconductor components often plays a key role in the design and performance of analog RF/microwave circuits. While these distortion effects can in principle be simulated in the time domain through extensive transient analysis, the inherent frequency domain nature of the problem results in relatively poor efficiency and/or accuracy when time domain algorithms are brought to bear [1]-[3]. Not only must the transient simulations be run until all the transients have died out and the system has reached steady state, but the wide range of frequency components present in intermodulation distortion analyses necessitates an extremely large number of very small time steps.

Harmonic balance (HB) is a nonlinear frequency domain analysis technique which is optimally suited for obtaining the quasi-periodic steady-state solution of systems with widely varying frequency components. Although HB simulation tools are available for analog circuit designers, these programs are circuit simulators which utilize compact models or lumped equivalent circuits for the active semiconductor devices. To fill the need for a physics-based HB simulation tool, we have developed a harmonic balance version of the PISCES [4] device simulator. This two-dimensional device simulation program solves the drift-diffusion transport equations in the frequency domain, using the full complement of standard PISCES physical models. External device parasitics and packaging are handled via a mixed-mode circuit simulation capability, which accepts S- or Y-parameter descriptions of external linear elements. Simulation runs yield the spectra for terminal currents and voltages, as well as the internal distributions of potential and carrier concentrations.

The direct HB approach has memory requirements which are $(2H+1)^2$ times higher than those of time domain methods when fully-coupled Newton-Raphson is used. (H here is the number of harmonics in the HB analysis, not counting DC --- see sec. 2.2). While such rapid growth may be tolerable for small circuits, it is clearly too expensive for realistic 2D semiconductor device simulation. In Section 3, we will demonstrate a modified nonlinear block Gauss-Seidel-Newton algorithm that overcomes these memory problems, dramatically improves on the speed of Newton-Raphson, and still offers highly robust convergence. This algorithm brings numerical HB device simulation capability well within the means of ordinary workstations.

2 Harmonic Balance Applied To The Semiconductor Equations

2.1 Motivation

The advantages of solving intermodulation distortion (IM) problems in the frequency domain can be seen by examining a typical two-tone input used for an IM test:

$$V_{IM}(t) = A\cos(\omega_a t + \varphi_a) + B\cos(\omega_b t + \varphi_b) \qquad (1)$$

For very small amplitudes A and B, the system will respond only at the frequencies ω_a and ω_b. As the amplitudes are increased, however, distortion components will also begin to appear. In general, the frequencies present in the response will then be given by the set

$$\omega_{k_1 k_2} = k_1\omega_a + k_2\omega_b \geq 0\,, \qquad (2)$$

where k_1 and k_2 are integers. Because ω_a and ω_b are typ-

ically closely spaced, certain frequencies in the response (such as $\omega_a - \omega_b$) may be many orders of magnitude smaller than those in the two-tone input. For example, consider a wireless application where a power RF transistor is being checked for intermodulation distortion arising from tones at $f_a = 1GHz$ and $f_b = 1GHz - 30kHz$. An accurate time domain analysis must deal with the fact that there will be frequency components present ranging from 30kHz to several GHz. Because of this wide range, ordinary transient analysis is extremely difficult --- the time steps have to be small enough to capture the largest frequencies, and yet there have to be enough time steps to cover several periods of the lowest frequency harmonics.

2.2 Solving the Time-Dependent Semiconductor Equations with Harmonic Balance

The harmonic balance algorithm overcomes the aforementioned shortcomings by taking a frequency domain approach, and thus bypassing time discretization altogether. The standard PISCES simulator solves the time-dependent drift-diffusion equations

$$\nabla \bullet (-\varepsilon \nabla \psi) = q\left(p - n + N_D^+ - N_A^- \right) \tag{3}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \bullet J_n - U$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \bullet J_p - U$$

where in addition we have

$$J_n = qD_n \nabla n - q\mu_n n \nabla \psi \tag{4}$$
$$J_p = -qD_p \nabla p - q\mu_p p \nabla \psi$$

Linear frequency-dependent device parasitics and packaging effects may be included by introducing additional KCL equations at appropriate PISCES terminal nodes.

For ordinary transient analysis, these equations are discretized in both time and space, and the basic variables at each internal device node k (ψ_k, n_k, p_k) $(1 \le k \le K)$ along with the voltages at each PISCES terminal q (v_q) $(1 \le q \le Q)$ are solved for each time step t_j. The HB version of the code retains the space discretization, but assumes that the basic variables have the form

$$x_n(t) = X_{n0} + \sum_{h=1}^{H} \left(X_{nh} \cos(\omega_h t) + X_{n,h+H} \sin(\omega_h t) \right) \tag{5}$$

where $\boldsymbol{x} = [\psi_1, n_1, p_1, \ldots, \psi_K, n_K, p_K, v_1, \ldots, v_Q]$. The frequencies used in the above expansion represent some finite subset of (2), and H, the total number of harmonics, must be chosen large enough so that the contribution from neglected frequencies is insignificant [1]. The goal of the HB analysis is to determine the appropriate set of values for basic variable quasi-Fourier coefficients X_{nh} to satisfy both (3) and the external KCL equations, subject to the appropriate boundary conditions.

Spatial finite-volume discretization of (3) yields a time-domain nonlinear system of the form

$$f_n(t) = \delta_{cont} x_n'(t) + g_n(\boldsymbol{x}(t); t), \quad 1 \le n \le 3K \tag{6}$$

where δ_{cont} is unity for the continuity equations, and zero for the Poisson equation. An additional Q auxiliary KCL equations describing the surroundng linear net are formulated directly in the frequency domain using admittance-matrix relationships, thus giving a total of $N = 3K + Q$ residual functions. Because the basic variables $x_n(t)$ are quasi-periodic (see (5)), the residual functions of (6) are as well. Consequently, they too may be expanded in a quasi-Fourier series of the form

$$f_n(t) = F_{n0} + \sum_{h=1}^{H} \left(F_{nh} \cos(\omega_h t) + F_{n,h+H} \sin(\omega_h t) \right) \tag{7}$$

For a given choice of the coefficients X_{nh}, each $f_n(t)$ may be sampled at $2H + 1$ appropriate time instants to determine its quasi-Fourier coefficients F_{nh} through a DFT (or FFT) matrix transformation. Letting $\boldsymbol{F}$ and $\boldsymbol{X}$ denote the vectors of F_{nh} and X_{nh}, respectively, we can write the harmonic balance system as a set of $(2H + 1)N$ nonlinear equations in as many unknowns:

$$\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{0} \tag{8}$$

3 Nonlinear Relaxation Methods

Solving (8) via conventional Newton-Raphson is prohibitively expensive for even small semiconductor device problems. Because the number of equations and unknowns scales as $(2H + 1)$ with the number of harmonics, the memory needed to factor the Jacobian will scale roughly as $(2H + 1)^2$. Thus, even a very modest HB intermodulation distortion analysis with, say, $H = 10$, would result in memory requirements which are 441 times as large as those used in DC/transient analysis. Since typical DC/transient analyses for semiconductor device problems require on the order of several tens of MB of RAM, a fully coupled Newton-Raphson approach for HB device simulation is clearly beyond the means of ordinary engineering workstations.

To resolve the above difficulties, we employ a modified block Gauss-Seidel-Newton relaxation method [5] to solve the harmonic balance system of equations (8). To simplify notation, we introduce complex phasors for the variables as follows: $\xi_{n0} = X_{n0}$, $\xi_{nh} = X_{nh} - jX_{n,h+H}$ for $h \ge 1$, along with $\Phi_{n0} = F_{n0}$, $\Phi_{nh} = F_{nh} - jF_{n,h+H}$ for the residuals. In order to determine the ordering of the Gauss-Seidel-Newton iterations, we will make use of the following error norms at each harmonic $0 \le h \le H$:

$$e_h = \sum_{n=1}^{N} |\Phi_{nh}| \tag{9}$$

The nonlinear relaxation algorithm is:

1. Start from an initial guess for $\boldsymbol{X}$. Typically, a PISCES DC analysis is a sufficient starting point.

2. Examine the error norm at each harmonic, and select the harmonic m ($0 \le m \le H$) with the largest error norm e_m.
3.Perform a Newton-Raphson step on the ***complex*** *system*

$$\Phi_{nm}(\xi_{1m}, \xi_{2m}, \ldots, \xi_{Nm}) = 0 \qquad (10)$$

where $1 \le n \le N$, and all omitted quasi-Fourier coefficients are held fixed.
4. Recompute the full residual vector $\boldsymbol{F}$, and then compute all e_h for $0 \le h \le H$.
5. If all error norms are below a specified tolerance, terminate. Otherwise, loop back to 2.

Note that this nonlinear relaxation scheme requires the LU factorization of at most a $2N \times 2N$ sparse real matrix. Thus, the memory usage for this algorithm is independent of the total number of harmonics *H*, and is only four times as large as the space required for an ordinary DC analysis.

The algorithm is motivated by the observation that at low distortion levels, each residual phasor Φ_{nm} is to first order a function of only $\xi_{1m}, \xi_{2m}, \ldots, \xi_{Nm}$. More precisely, a first order expansion of (6) about the DC operating point $\boldsymbol{x}_0$ yields an expression of the form

$$f_n(t) = Re\left\{ \sum_{h=1}^{H} \Phi_{nh} e^{j\omega_h t} \right\}, \qquad (11)$$

where we have

$$\Phi_{nh} = j\omega_h \xi_{nh} \delta_{cont} + \sum_{i=1}^{N} \frac{\partial g_n}{\partial x_i}(\boldsymbol{x}_0)\, \xi_{ih}. \qquad (12)$$

The above analysis provides the motivation for the block selection used in the modified Gauss-Seidel-Newton relaxation scheme. Clearly, this algorithm is extremely effective at low distortion levels. However, as amplitudes rise, the residual harmonics begin to couple to additional variable harmonics through the Hessian (and higher order) terms. We can therefore no longer be sure of convergent behavior. (Note, however, that if the algorithm converges, the solution is always correct --- we make no approximations to the right hand side residuals).

Fortunately, experience has shown that the nonlinear relaxation scheme outlined above converges for virtually all semiconductor device problems of practical interest. Indeed, we have only observed non-convergent behavior under unrealistically large voltage swings, which gave difficulties even to full Newton-Raphson algorithms used in conjunction with extremely fine source stepping. These and other practical results are more fully addressed below.

4 Results and Examples

4.1 Speed and Convergence of the HB Relaxation Method on Practical Problems

The memory savings brought about by the relaxation method of Section 3 are crucial to the practical simulation of 2D device structures. In this section, we will provide empirical data to show that the method is also superior to full Newton-Raphson in speed, and that its convergence behavior is extremely robust under reasonable operating conditions encountered in semiconductor device simulation.

In order to compare the speed of the relaxation algorithm to the fully-coupled Newton-Raphson approach, we ran a series of tests on a small (100 grid point) 2D diode structure. The choice of such a relatively modest grid size was due to the extremely large memory requirements of the fully-coupled Newton solver. The diode was driven by a voltage source of the form $V = 0.7 + 0.05\sin(\omega t)$ (where voltage magnitudes are given in Volts). All simulation times in Table 1 were measured on an HP735 running HP-UX. The advantage of the relaxation algorithm in terms of computational time is considerable, and increases with the number of harmonics used. This relative advantage stays roughly the same across a wide range of input frequencies and AC voltage magnitudes. The performance difference could not be measured for large grid sizes or numbers of harmonics, as full Newton became far too memory intensive to even be accommodated in RAM.

TABLE 1. Time Comparison (in sec) of Relaxation vs. Full Newton

N	Relaxation	Newton
4	35	74
5	43	148
6	61	229
7	88	358
8	101	540

The convergence behavior of the relaxation scheme appears to be outstanding when applied to the drift-diffusion semiconductor equations. A wide variety of semiconductor devices, ranging from GaAs FETs to silicon BJTs to MOSFETs have been analyzed using this technique under a wide range of operating conditions. Examples with grid sizes of over 1500 nodes and 100 harmonics have been handled successfully.

To test the ultimate limits of the algorithm's convergence radius, we utilized a silicon BJT with minimal resistive parasitics. The device simulation converged well for input voltages as high as $V_{BE} = 0.7 + 0.15\sin(\omega t)$, so long as the BJT remained forward active. Convergence problems occured only at extremely high distortion levels, in those circuit configurations where the device swung periodically from forward-active mode into deep saturation. The advantages of harmonic balance over standard transient analysis in the aforementioned regime are questionable, as the number of harmonics necessary to capture all the distortion components grows extremely large. Even so, promising new relaxation algorithms are currently under

development to handle the ultra-high distortion case effectively.

4.2 Examples

In this section, we present simulation results from some harmonic balance PISCES runs. Figure 1 shows the cross-section of a silicon npn BJT [6]. Both single-tone harmonic distortion and two-tone intermodulation tests are performed on this device. Figure 2 shows the collector current spectrum when the base of a forward-active BJT is driven by a 100MHz, 120mV sinusoid with a dc offset of 0.7V. Fourteen harmonics are utilized for this HB analysis, with an hour of simulation time necessary for this case on an HP-735 workstation. Figure 3 shows the results of a two-tone intermodulation distortion analysis of the same device structure. In this case, the base-emitter voltage input consists of two 25mV sinusoids, biased around a dc offset of 0.65V. One sinusoid has a frequency of 1.0GHz, with the other at 1.05GHz. Seventy harmonics were used in the analysis.

5 Conclusion

We have presented a physically-based two-dimensional harmonic balance semiconductor device simulator. This code is a fully operational frequency domain version of the PISCES time-domain simulation tool. Nonlinear relaxation algorithms were developed to combat the problem of extremely high memory requirements necessary for fully-coupled Newton-Raphson approaches. The relaxation algorithms seem to be particularly well-suited for the HB device simulation problem. Compared with conventional Newton methods, they offer drastic reductions in memory usage and execution time, while retaining outstanding convergence behavior. The result is a device analysis tool which brings harmonic balance device simulation capability within reach of ordinary engineering workstations.

References

[1] K.S. Kundert, J.K. White, and A. Sangiovanni-Vincentelli, *Steady-State Methods for Simulating Analog and Microwave Circuits*, Kluwer Academic Publishers, 1990

[2] K.S. Kundert and A. Sangiovanni-Vincentelli, "Simulation of nonlinear circuits in the frequency domain," IEEE Trans. on Computer-Aided Design, pp.521-535, Oct. 1986

[3] V. Rizzoli and A. Neri, "State of the art and present trends in nonlinear microwave CAD techniques," IEEE Trans. on Microwave Theory and Techniques, pp.343-365, Feb. 1988

[4] Z. Yu, D. Chen, L. So, and R.W. Dutton, *PISCES-2ET --- Two-Dimensional Device Simulation for Silicon And Heterostructures*, Stanford University, 1994

[5] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970

[6] P. Vande Voorde, D. Pettengill, and S.Y. Oh, "Hybrid simulation and sensitivity analysis for advanced bipolar device design and process development," IEEE 1990 Bipolar Circuits and Technology Meeting, Minneapolis, Minnesota

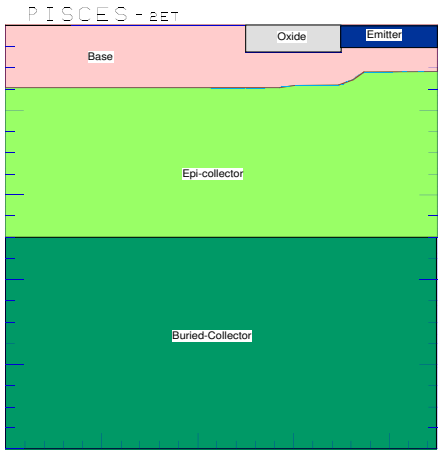

Figure 1. Cross-section of silicon npn BJT used in harmonic balance analysis.

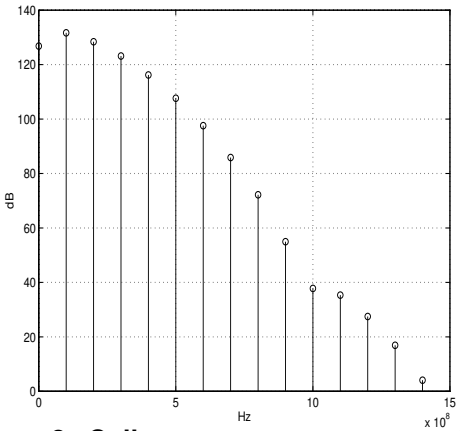

Figure 2. Collector current spectrum in response to a 100MHz, 120mV sinusoidal input.

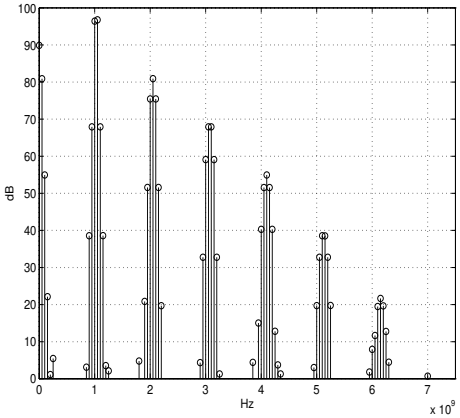

Figure 3. Collector current spectrum in response to a two-tone intermodulation distortion test, with 25mV tones at 1.0GHz and 1.05GHz.