Large Signal Analysis of RF/Microwave Devices with Parasitics Using Harmonic Balance Device Simulation

Boris Troyanovsky

EEsof Division,
Hewlett-Packard Co.,
1400 Fountaingrove Pkwy,
Santa Rosa, CA 95403
(formerly with Stanford Univ.)

Francis M. Rotella
Zhiping Yu
Robert W. Dutton

Center for Integrated Systems,
Stanford University, Stanford, CA 94305

Junko Sato-Iwanaga

Electronics Research Lab.,
Matsushita Electronics Corp.,
3-1-1 Yagumo-Nakamachi,
Moriguchi-shi, Osaka 570, Japan

Abstract -- A harmonic balance technique has been developed for large-signal, steady-state analysis of semiconductor devices. The algorithm has been implemented in the PISCES-2H device simulator, and has been applied to simulating a number of industrial device structures. The new tool is capable of accurately and efficiently simulating harmonic and intermodulation distortion for devices embedded in a realistic biasing and parasitic network. Krylov subspace techniques are employed to solve the extremely large system of harmonic balance equations arising from two-dimensional physics-based device simulations. A new preconditioning scheme is proposed to address the large size of multi-tone distortion problems.

I. Introduction

Large-signal distortion arising from nonlinearities of semiconductor devices plays a critical role in the performance of RF/microwave wireless communication circuits. Rigorous physics-based device-level simulation of harmonic distortion offers a number of advantages over analysis based solely on compact circuit-level transistor models. First of all is the improvement in accuracy that can be achieved from a detailed solution of the semiconductor equations. Secondly, the detailed physical insight may be gained from internal large-signal AC carrier and potential distributions. These features can be used by device designers to help identify the sources of signal distortion and thus to improve the device design. The impact of device structure and physical parameters on terminal I-V distortion characteristics can be directly examined.

In this paper, we present a device simulator capable of performing large-signal, steady-state analysis in the frequency domain using the harmonic balance (HB) technique. Harmonic balance is superior in certain applications compared to standard time-domain transient approaches. Perhaps chief among these is its ability to handle input signals of either widely or tightly spaced incommensurate frequencies. The technique is also very effective in directly capturing the steady state of systems with widely-varying time constants. In addition, the harmonic balance exhibits excellent dynamic range for resolving low-level distortion products and is robust enough to handle the levels of nonlinearity typically seen in the majority of RF/microwave applications.

A major setback of HB relative to time domain methods, however, is the considerable increase in the number of state variables involved. For a harmonic balance solution with waveform of H harmonics, the number of state variables is increased by a factor of $2H + 1$ (H here denotes the number of harmonics in the HB analysis, not counting DC component -- see Section II). When harmonic balance is applied to the semiconductor equations, the number of state variables can easily approach or exceed a million. Direct HB solution techniques suffer from the problems of memory requirement which scales as $(2H + 1)^2$ and execution time which scales as $(2H + 1)^3$. In practical terms, this implies that a device structure which normally requires 10MB of RAM for conventional device simulation would take up over 400GB of memory for a 100-harmonic HB analysis.

For some time, nonlinear relaxation algorithms have been applied to solving such large-scale semiconductor device problems [1]. These techniques enjoy the advantage of very low memory requirements that are approximately constant regardless of H, and are robust enough to solve the problems arising in semiconductor device simulation. Unfortunately, they exhibit non-quadratic convergence behavior which leads to an excessive number of device evaluations. On the other hand, harmonic balance solution algorithms based on Krylov subspace methods [2]-[3] offer faster, quadratic rate of convergence and are preferred in solving large systems of HB equations.

In the subsequent sections, we will present algorithms for solving the HB device equations. These algorithms are based on Newton-Raphson solution strategies where the linear solver is implemented using preconditioned Krylov subspace methods -- our present Krylov solver uses the restarted GMRES technique [4]. Special-purpose preconditioners will be introduced to deal with both single-tone and multi-tone distortion problems. Following the algorithm description, two examples will be presented to illustrate the

simulator's ability to handle the nonlinear device embedded in a linear biasing and parasitic network. The first example is a GaAs MESFET for microwave applications, surrounded by parasitics, which are resulted from the packaging, and the matching circuitry. The second example is an SOI BJT for high-speed and power applications. For the first example, comparisons with actual measurements are provided. The second example is a demonstration of program's capability in identifying internal distortion sources within a device.

II. Applying Harmonic Balance to the Semiconductor Equations

A. Fundamentals

The PISCES-2H device simulator solves the time-dependent drift-diffusion equations

$$\nabla \bullet (-\varepsilon \nabla \psi) = q(p - n + N_D^+ - N_A^-) \tag{1}$$

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

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

where

$$= \quad \nabla n - q\mu_n n \nabla \psi \tag{2}$$

$$J_p = -qD_p \nabla p - q\mu_p p \nabla \psi$$

For carrying out distortion analysis on a device structure, we are typically interested in embedding the device in a linear network. Since we are dealing with single-device simulations, the voltage sources driving the network are usually either one- or two-tone sinusoids having the form

$$= \quad \cos(\Omega_1 t + \varphi_1) + A_2 \cos(\Omega_2 t + \varphi_2) \tag{3}$$

In a single-tone harmonic distortion analysis, is set to zero and only one frequency is present in the stimulus. In an intermodulation distortion test, both and are non-zero, and the frequencies Ω_1 and Ω_2 are usually very closely spaced. In either type of simulation, the goal is to capture the large-signal, sinusoidal, steady-state response. This is time-consuming and inefficient when standard time domain algorithms are brought to bear, since the stiff biasing network often introduces prohibitively long time constants. In addition, the presence of both widely and tightly spaced intermodulation distortion products requires the integration over many periods of sinusoid with the highest frequency. These factors make harmonic balance the method of choice in simulations of device distortion, particularly in the case of multi-tone problems.

To solve the system of equations (1) in the frequency domain, the partial differential equations are first discretized in space, and all state variables are then expanded in sines and cosines in time domain. With H harmonics, we have

$$= \quad + \sum_{h=1}^{H} [\quad \cos(\omega_h t) - X_{nh}^I \sin(\omega_h t)]$$

$$= \Re e\left\{ \sum_{h=0}^{H} X_{nh} \exp(j\omega_h t) \right\} \tag{4}$$

where . In the device simulation, the state variables at each internal node k are (ψ_k, n_k, p_k), representing electrostatic potential, electron and hole concentrations. In addition, there are state variables representing voltages at each external contact (i.e. terminal) connecting the device to its surrounding linear network. Assuming a total of K internal nodes and Q terminals attached to linear network ports, the time-domain state vector of length takes the form

$$= [\psi_1, n_1, p_1, \ldots, \psi_K, n_K, p_K, v_1, \ldots, v_Q]^T . \tag{5}$$

The harmonic balance state vector $\boldsymbol{X}$ thus has a length of . , as it requires values to represent each state variable waveform in (4).

The set of frequencies present in the waveforms resulting from excitation by the two-tone stimulus (3) is

$$\omega_{k_1 k_2} = k_1\Omega_1 + k_2\Omega_2 \tag{6}$$

for integer combinations of and . In principle, there is an infinite number of such frequencies. In practice, only a finite number of these spectral components contain significant energy. Thus, we employ the so-called diamond truncation scheme [5] to restrict the frequency components to

$$\omega_{k_1 k_2} = k_1\Omega_1 + k_2\Omega_2 \text{ s.t. } | \quad | + | \quad | \le \tag{7}$$

where P is the truncation order. In general, this set of frequencies is not harmonically related to a single fundamental frequency, and thus the standard discrete Fourier transform (DFT) cannot be applied directly. To circumvent this problem, and hence to enable use of the efficient FFT algorithm, we use the widely accepted "frequency remapping" technique of Hente and Jansen [6].

In order to determine the set of harmonic balance equations, we first note that after spatial discretization, (1) can be written in time domain as

$$\delta_{cont}\frac{dx_n}{dt} + g_n(\boldsymbol{x}(t)) - w_n(t) = 0 \ , \quad \le \ \le \tag{8}$$

In the equation above, δ_{cont} is zero for the Poisson equation, and unity for the continuity equations. The term represents the coupling of the state variable on the n-th node with others, and is a driving function that represents external sinusoidal excitation. Each of these time domain equations is sampled and then Fourier-transformed into frequency domain relations of the form

$$\omega_h \delta_{cont} X_{nh} + G_{nh}(\boldsymbol{X}) - W_{nh} = 0 , \quad \le \ \le \tag{9}$$

where is the h-th harmonic of , and is the h-th harmonic of . This is the set of harmonic balance equations which is to be solved for $\boldsymbol{X}$.

B. Solving the Harmonic Balance Equations

If we let the symbol Γ denote the single-sided DFT operator matrix, then

$$\begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix} = \Gamma \begin{bmatrix} g_n(\boldsymbol{x}(t_0)) \\ | \\ g_n(\boldsymbol{x}(t_{2H+1})) \end{bmatrix} \tag{10}$$

Thus, (9) can be formulated at each nonlinear Newton step by taking the current "guess" for $\boldsymbol{X}$, forming through (4), and then using the samples of to compute through (10). The remaining operations necessary to form the residual in (9) are straightforward.

The harmonic balance Jacobian matrix has dimensions $\times$ where . To each structurally non-zero entry $\partial \quad / \partial x_m$ of the time domain Jacobian corresponds a dense $\times$ block $\partial \quad / \partial \boldsymbol{X}_m$. Each such block consists of derivatives of the harmonics of the n-th residual with respect to harmonics of the m-th state variable.

As discussed earlier, factorization of the harmonic balance Jacobian by direct methods is prohibitively expensive. It has been shown in [2]-[3] that preconditioned Krylov-based linear solvers can be applied very effectively to solving such large-scale HB problems. Krylov subspace solvers (such as GMRES, QMR, and TFQMR) require the ability to multiply the HB Jacobian matrix by an arbitrary complex vector of length $N(H+1)$ (alternately viewed as a real vector of length) at each linear iterative step. Letting $\boldsymbol{Z}$ denote such a vector, we see that the necessary Jacobian matrix-vector products consist of block-multiplies

$$\frac{\partial \boldsymbol{G}_n}{\partial} \boldsymbol{Z}_m + \delta_{cont}\delta_{n-m} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & j\omega_1 & \dots & 0 \\ | & | & \backslash & | \\ 0 & 0 & \dots & j\omega_H \end{bmatrix} \begin{bmatrix} Z_{m0} \\ Z_{m1} \\ | \\ Z_{mH} \end{bmatrix} \tag{11}$$

The right-most product represents the contribution from the derivatives of carrier concentrations present in the continuity equations -- this contribution occurs only in diagonal blocks corresponding to the continuity equations. The left-most product can be efficiently calculated through use of the FFT. Applying the chain rule to (10) allows us to derive an efficient formulation of the product:

$$\frac{\partial \boldsymbol{G}_n}{\partial} \boldsymbol{Z}_m = \Gamma \begin{bmatrix} g_{nm}(t_0) & \dots & 0 \\ | & \backslash & | \\ 0 & \dots & g_{nm}(t_{2H+1}) \end{bmatrix} \Gamma^{-1} \boldsymbol{Z}_m \tag{12}$$

where we have defined $= \partial \quad / \partial x_m$. To calculate such products efficiently, the samples of $\partial \quad / \partial x_m$ are computed once and stored for the duration of the iterative linear solve. The inverse FFT operation $\Gamma^{-1}\boldsymbol{Z_m}$ needs to be carried out only once for each column m. The product for each structurally non-zero block in column m may then be computed with multiplies, followed by a single FFT operation. We note that in practice, unless happens to be a power of 2, we have to oversample $\partial \quad / \partial x_m$ to bring it up to the next higher power of 2 so that efficient FFT algorithms can be applied.

C. Preconditioning

Linear iterative solvers based on Krylov subspace methods must first be *preconditioned* to achieve fast and robust convergence. That is, instead of solving the original linear system

(13)

we solve

(14)

where the preconditioning matrix should ideally be a good approximation to the original Jacobian J, while also being easy to factor. It is well known that the harmonic balance equations become diagonally dominant as the levels of distortion are reduced. For infinitesimally small amplitudes, we see from (10) and (11) that each structurally non-zero Jacobian block can be represented by the complex matrix

$$= \begin{bmatrix} \gamma_{nm}[0] & 0 & --- & 0 \\ 0 & \gamma_{nm}[0] + \kappa_1 & --- & 0 \\ | & | & \backslash & | \\ 0 & 0 & --- & \gamma_{nm}[0] + \kappa_H \end{bmatrix} \tag{15}$$

where $\gamma_{nm}[0]$ is the DC component of the Fourier Transform of $\partial \quad / \partial x_m$, and $\kappa_h = j\omega_h \delta_{cont}\delta_{n-m}$ is the contribution from the derivative in the continuity equation. Note that although technically the Harmonic Balance Jacobian is a matrix of $\times$ quads [5], the diagonal can be represented by complex numbers for infinitesimally low distortion levels. Such a block-diagonal complex matrix is an excellent candidate for preconditioning the semiconductor device problem. By suitable permutation of the rows and columns, the block-diagonal preconditioner can be expressed as decoupled complex matrices having the sparsity pattern of the original time domain Jacobian. Thus, factorization time and storage of the complex-diagonal preconditioner is equivalent to that of performing AC analyses.

In solving two-tone intermodulation distortion problems, the number of harmonics H rises as with the truncation order. For three or more tones, this rate of growth is even more severe. Given the extremely large number of state variables in semiconductor device simulation problems, this rapid growth makes the storage of an harmonic preconditioner prohibitive.

To come up with a less expensive preconditioner, we note that many two-tone problems have "bands" of harmonics that are very tightly spaced in frequency. For instance,

consider an input of the form (3), where $\Delta\Omega = \Omega_2 - \Omega_1$ is small compared to either Ω_2 or Ω_1. Then by simple algebraic manipulation, the set of harmonic balance frequencies (6) can be written as

$$\omega_{k_1 k_2} = (k_1 + k_2)\Omega_1 + k_2 \Delta\Omega \ . \qquad (16)$$

We proceed to define the frequencies Ω_1, $\leq \leq$, and the corresponding continuity-term derivative set $\hat{\kappa}_p = jp\Omega_1 \delta_{cont} \delta_{n-m}$. In the preconditioning matrix (15), we can use only these frequencies, and for each ω_h simply pick the closest p such that $\Omega_1 \approx \omega_h$ and $\hat{\kappa}_p \approx \kappa_h$. Thus, the cost of storing and factoring the preconditioner is reduced to that of AC analysis factorizations, which is a significant improvement over . Note, of course, that there will still be the need for back substitutions on each Krylov iteration once the factorizations are done. We refer to the preconditioner presented in this paragraph as the *sectioned block-diagonal preconditioner*.

III. Mixed-Level Device and Linear Circuit Simulation

Most device simulators allow resistors and capacitors to be placed on device terminals. However, for RF and microwave simulations, many additional linear elements are often necessary to fully characterize the device parasitics and to represent the surrounding circuitry. To satisfy these requirements, extensive linear circuit boundary conditions have been implemented in the PISCES device simulator.

Given an arbitrary linear network around a device with Q terminals, a Q-port AC analysis can be performed at each harmonic balance frequency to yield a total of complex admittance matrices (ω_h) having dimension $\times$. For a given harmonic balance state vector, the current at terminal q can be written as

$$— \qquad . \qquad (17)$$

The term represents the contribution from the electron and hole conduction currents, while the term involving accounts for the contribution from the displacement current $\partial \ / \partial t$. The circuit boundary conditions can thus be completely specified in the frequency domain by formulating the complex KCL equations

$$+ \ \omega_h G^D_{qh}(\boldsymbol{X}) + \sum_{r=1}^{Q} Y_{qr}(\omega_h) V_{rh} = 0 \qquad (18)$$

where $\leq \leq$, $\leq \leq$. Note that the HB terminal voltage harmonic vectors are stored as in the HB state vector $\boldsymbol{X}$. The auxiliary equations (18) are added directly to the harmonic balance system. The only modifications necessary to the aforementioned block-diagonal preconditioners is the inclusion of (ω_h) on the diagonals of each relevant harmonic balance block.

IV. Results and Application Examples

All simulation results reported in this section have been obtained on an HP-J210 workstation with a 120MHz CPU and 250MB of RAM. The memory usage below includes an additional 30MB of PISCES static arrays which are not used by the harmonic balance engine. In principle, the memory usage presented below could be reduced by 30MB with the same algorithms applied to a modified PISCES code.

A. GaAs MESFET with Packaging Parasitics

This example focuses on the ability to design a GaAs FET amplifier with low levels of harmonic distortion. Key aspects include design of an impurity profile which keeps transconductance constant and decreases the variation in gate-source and gate-drain capacitances. In order to analyze this device and compare the simulated results with measurements, the simulation needs to be set up to accurately represent the measurement environment (see Fig. 1). RF chokes and blocking capacitors lead to 50 Ohm terminations. Parasitics surrounding the device must also be included. A comparison between simulated and measured data is shown in Fig. 2.

The mesh used for this structure has 952 grid nodes and 2 auxiliary KCL equations, for a total of . For single-tone HB analyses at , the total system size is 88,598 unknowns. Solution time varied from under 11 min. per HB analysis at RF source levels under 100mV, to about an hour near 1V. Total memory usage was 121MB.

A larger GaAs FET was analyzed under two-tone excitation to illustrate the code's effectiveness on large-scale problems. A two-tone RF input was applied (2GHz and 1.9GHz), with both tone magnitudes at 100mV. The device had 1406 grid points and two auxiliary equations, for a time domain system size of . For the two-tone analysis at , the total number of unknowns was 932,620. Total execution time was 5 hr. 8 min., with memory usage at 360MB. Because this exceeded the 250MB physically present on our machine, extensive swapping occurred. The actual CPU time was 2hr. 8 min., and thus we anticipate that a performance increase of over a factor of two would be possible on machines having at least 360MB of RAM.

B. Distortion Analysis of an SOI BJT

Fig. 3 shows a 3D schematic of a silicon-on-insulator (SOI) bipolar device, where the active region is isolated by a 0.5μm oxide layer. While the original structure [7] has an n^+ floating collector layer underneath the vertical *npn* transistor at the Si-SiO_2 interface, this structure relies on the back gate (i.e. the substrate) bias to form a high electron concentration layer at the interface. The actual size and orientation (for the subsequent contour and perspective plots) of the device being simulated are shown in Fig. 4. When the

substrate bias is increased positively with respect to the emitter, an electron accumulation layer is formed as shown in Fig. 5. This high concentration layer helps reduce the transit time for electrons to cross the collector region [7] as is evident in Fig. 6, where the cutoff frequency (f_T) vs. the collector current density is drawn. The improvement in f_T is about 10% for V_{sub}=10V. Furthermore, as our simulation shows, the overall distortion level in the output (i.e. collector) current is reduced by as much as 20% when compared to zero substrate bias. The physical explanation for this reduction is as follows.

One key component of distortion is the nonlinearity of the base-collector junction capacitance, which is modulated by the large swing of the output signal. Due to the existence of this high electron concentration layer at the Si-SiO_2 interface, the potential at this layer is essentially clamped (or "locked"), limiting the *ac* voltage swing across the base-collector junction. This phenomenon has the benefit of reducing the distortion level for the entire transistor.

The grid size for simulating this device is 1190 nodes. An $H = 15$ one-tone simulation at a 50mV drive level with a base-emitter bias of 0.7V took 52 min. and 150 MB of RAM. The total number of unknowns for this simulation was 110,670. For a two-tone intermodulation distortion analysis with 20mV tones and $H = 90$, the total number of unknowns was 646,170, and the run-time was 4 hr. 34 min. The CPU time accounted for 3hr. 32 min., since the 305MB necessary for this simulation was more than was physically present in the machine. Fig. 7 shows the 2nd harmonic distortion as a function of substrate bias, while Fig. 8 and Fig. 9 show the actual collector current spectrum for the one- and two-tone cases, respectively. We note that in addition to being able to determine current spectra, the simulator has the ability to determine the harmonic distribution inside the device for fundamental variables like electron concentration (Fig. 10) and potential (Fig. 11). The latter capability can help in identifying the origin of the distortion and improving the device design.

V. Conclusions

This paper presented an approach to simulating harmonic distortion in semiconductor devices. Krylov subspace methods are used to apply the quadratically convergent Newton-Raphson algorithm without bearing the cost of directly factoring the harmonic balance Jacobian. Examples are presented to demonstrate the code's capabilities and performance.

Work is currently in progress to further improve the fundamental simulation algorithms. In the near future, we will investigate more robust preconditioners, compare the performance of GMRES with that of QMR/TFQMR, and carry out extensive numerical experiments to determine the limitations of the current approach. In addition, preliminary algorithms have been implemented that dramatically lower simulator memory usage. All of these results will be more fully reported in the future.

Acknowledgments

The authors would like to acknowledge the following organizations. The GaAs example was done in cooperation with Matsushita Electronics Corporation, Osaka, Japan. The SOI example was provided by Dr. Torkel Arnborg of Ericsson Components, Stockholm, Sweden. Funding from SRC (Semiconductor Research Corp.) and NSF through NCCE (National Center for Computational Electronics) program is critical to the success of this project. One of the authors, Boris Troyanovsky, would like to especially thank Jeffrey W. Meyer and Niranjan Kanaglekar, his managers at HP-EEsof, for their support. He would also like to thank David D. Sharrit of HP-EEsof for many valuable technical discussions.

References

[1] B. Troyanovsky, L. So, Z. Yu, and R. W. Dutton, "Relaxation-based harmonic balance technique for semiconductor device simulation," *Proc. IEEE/ACM International Conf. on Computer-Aided Design*, pp. 700-703, Nov. 1995.

[2] R. Melville, P. Feldmann, and J. Roychowdhury, "Efficient multi-tone distortion analysis of analog integrated circuits," *Proc. of the IEEE Custom Integrated Circuits Conference*, pp. 241-244, May 1995.

[3] P. Feldmann, R. Melville, and D. Long, "Efficient frequency domain analysis of large nonlinear analog circuits," *Proc. of the IEEE Custom Integrated Circuits Conference*, pp. 461-464, May 1996.

[4] Y. Saad and M.H. Schultz, "GMRES: a generalized minimal residual method for solving nonsymmetric linear systems," SIAM J. Sci. Stat. Comput., Vol. 7, pp. 856-869, July 1986.

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

[6] D. Hente and R.H. Jansen, "Frequency domain continuation method for the analysis and stability investigation of nonlinear microwave circuits," *IEE Proceedings*, part H, vol. 133, no.5, pp. 351-362, Oct. 1986.

[7] T. Arnborg, "Modeling and simulation of high speed, high voltage bipolar SOI transistor with fully depleted collector," *IEDM Digest*, pp. 743-746, Dec. 1994.

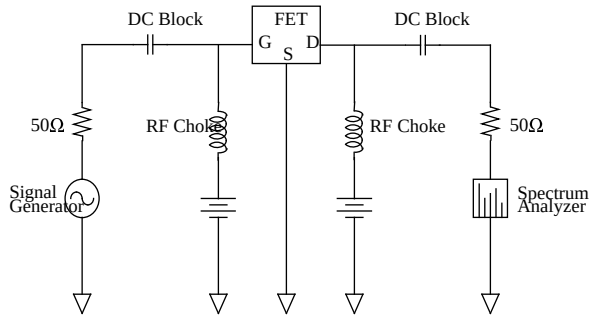
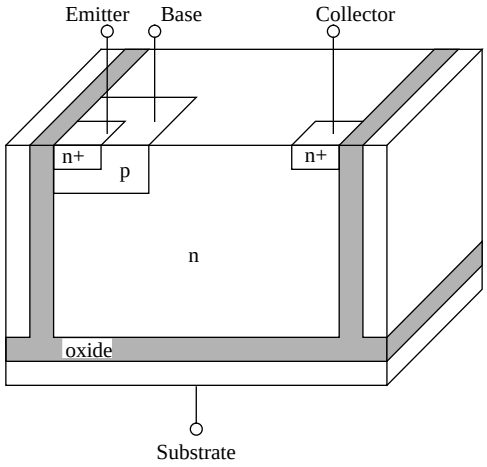
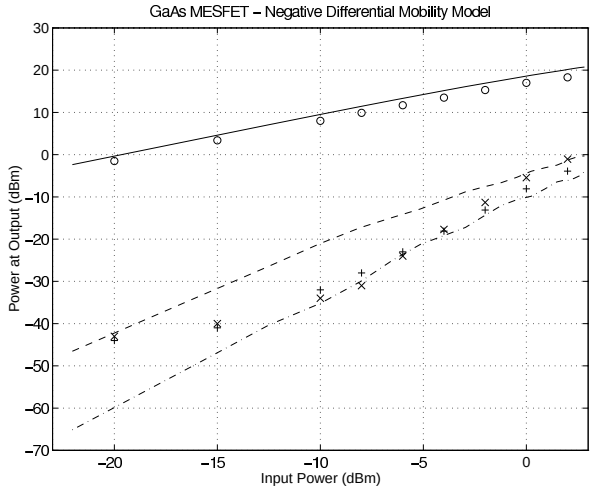


Figure 1. RF tw o-port network used for GaAs MESFET simulation and measurement.





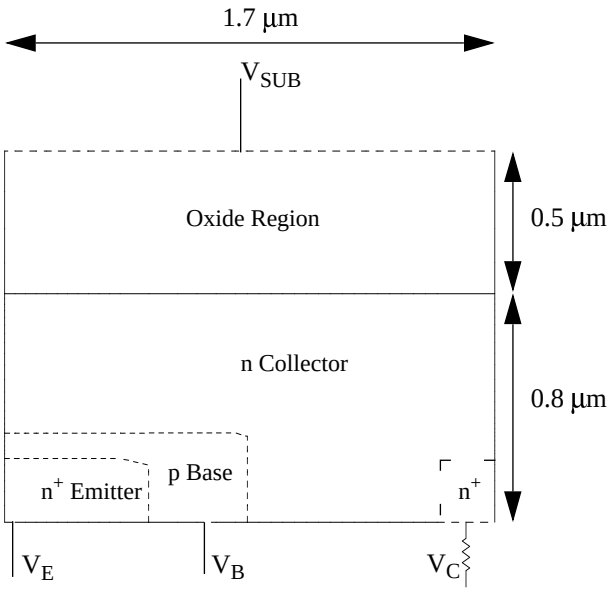


Figure 4. 2D cross-section of the above device, drawn to scale.

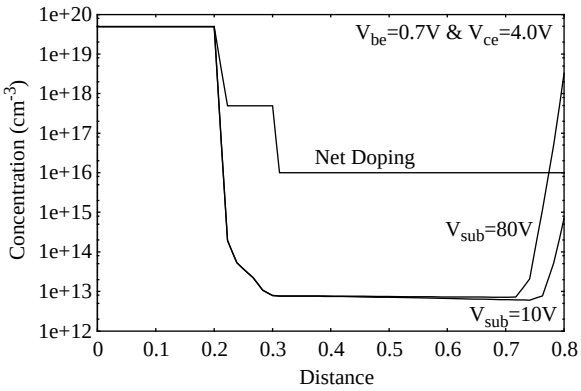
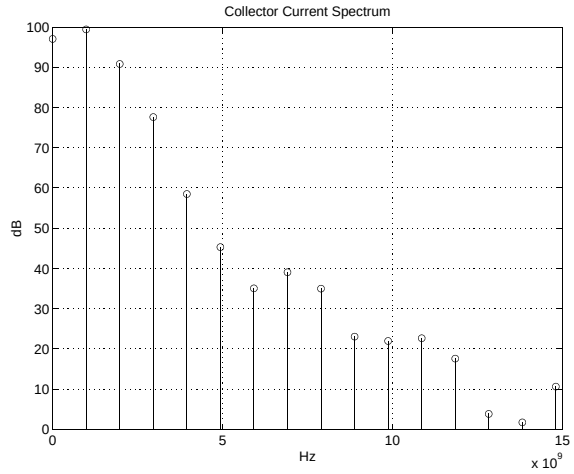
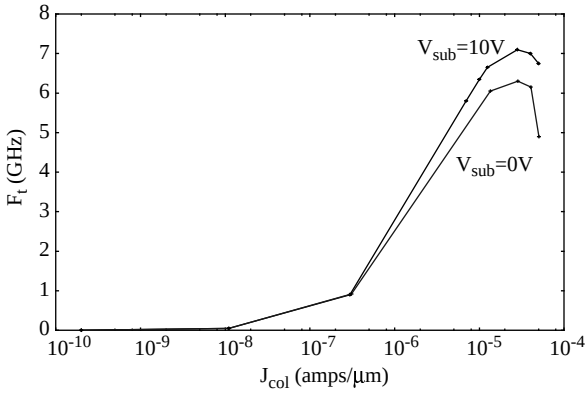
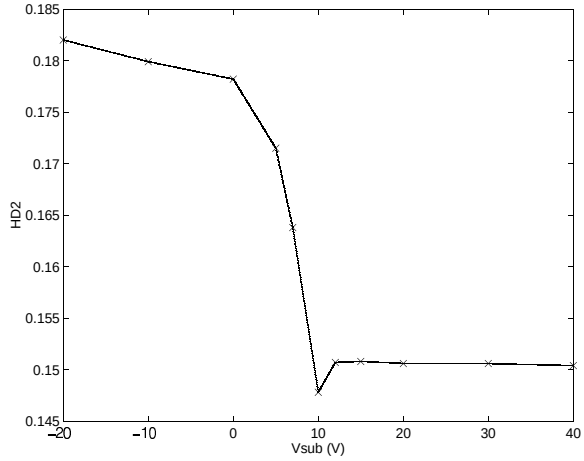
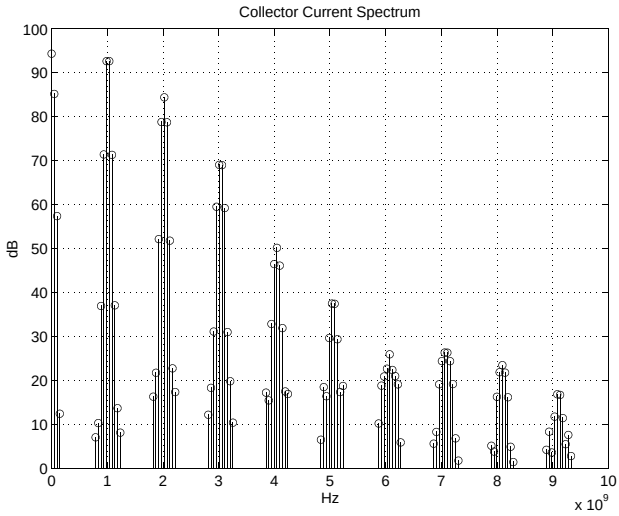


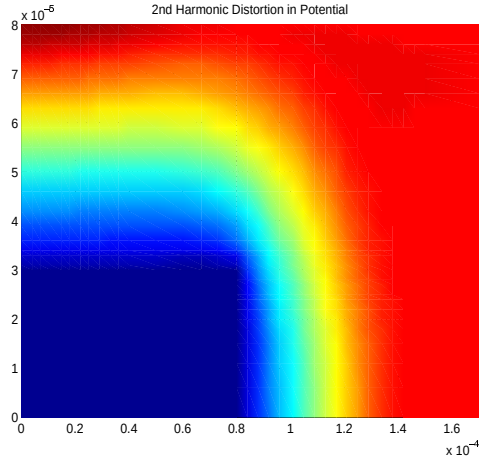
Figure 5. Formation of electron accumulation layer at the Si-$SiOR_2$ interface induced by the positive substrate bias.

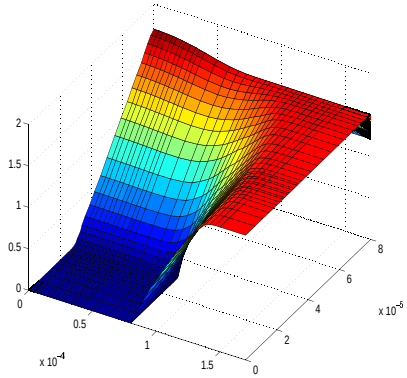


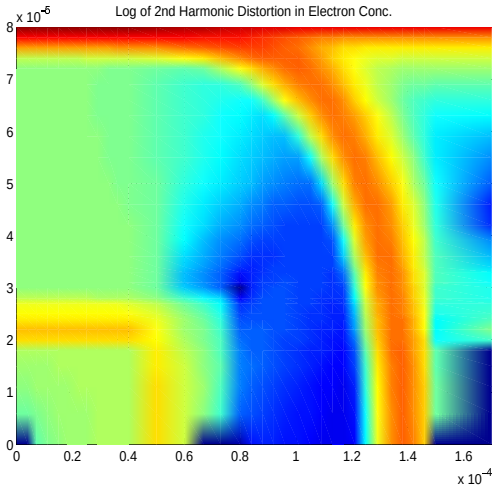


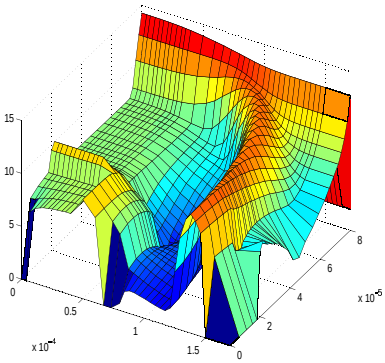



Collector Current Spectrum
dB
Hz
x 10^9


2nd Harmonic Distortion in Potential
x 10^-5
x 10^-4


x 10^-4
x 10^-5


Log of 2nd Harmonic Distortion in Electron Conc.
x 10^-6
x 10^-4


x 10^-4
x 10^-5