AN ALTERNATIVE APPROACH TO LUMPED ELEMENT MODELING

Michael Sever

Department of Mathematics
The Hebrew University
Jerusalem, Israel

A method is described for obtaining the computational advantages of lumped element modeling, in time-dependent circuit simulations, without the need of choosing a model topology or element values by ad hoc means. The initial overhead is comparable to that of computing a stationary I-V characteristic. The method is useable whenever the quasi-stationary approximation is physically acceptable, for any particular device. In can easily be combined with "full" simulation of those devices in which the quasi-stationary approximation cannot be used, and can be used in unusual operating regions without special modifications. An example of the use of the method and a comparison with conventional simulation is included.

I. REVIEW

To first order in the time increment h, a linear, inhomogeneous relation of the form

$$(1.1)\qquad I = A(V - V_0) + G$$

exists relating the terminal voltages and currents of a semiconductor device. In (1.1), I is the vector of terminal currents, V the terminal voltages, and V_0 the terminal voltages at the previous time level; these are m-vectors for a device with m external contacts. The $m \times m$ matrix A, hereafter called the admittance matrix, and the m-vector G are computable from a knowledge of the internal device variables, the carrier densities and electrostatic potential distributions in particular. A derivation is given below, or may be found in [3]; for the moment, it suffices to note that these quantities also depend on the time step h.

Thus the quantities A,G, and V_0 completely describe the device, in this approximation, as seen by the outside world, and this suggests a simple procedure for time-dependent simulation of multiple, coupled devices, while performing the internal computations for each device separately. First for each device in the network, the quantities A,G are computed, using the latest available approximations for the carrier densities and electrostatic potential functions. Then equations (1.1) are assembled and solved simultaneously with the Kirchoff's laws relations for the network. As the matrix A is symmetric, this can effectively be done using Thevinin equivalent circuits for the devices and basic circuit analysis methods [2,3]. Solving this system yields, among other things, the updated terminal bias voltages for each device. And given the updated terminal voltages, the internal variables for each device can be updated separately. Computationally, this final step is the expensive one, but it is necessary if the quantities A,G are to be found at the subsequent time level and the process continued.

To clarify this, we show how the quantities A,G can be obtained. Differentiating the Poisson equation with respect to time and using the two carrier continuity equations, we obtain the well-known equation of total current continuity, i.e.,

$$(1.2)\quad \nabla\cdot(\varepsilon\nabla\psi_t) + \nabla\cdot((\mu_u u+\mu_v v)\nabla\psi) - \nabla\cdot(\mu_u\nabla u-\mu_v\nabla v) = 0,$$

in which ψ,u,v are the electrostatic potential, electron density, and hole density; ε is the dielectric constant and μ_u,μ_v the respective carrier mobilities. A backward difference is used in the displacement current term for stability. Thus we obtain

$$(1.3)\quad \nabla\cdot\left(\varepsilon\nabla\left(\frac{\psi-\psi_0}{h}\right)\right) + \nabla\cdot((\mu_u u+\mu_v v)\nabla\psi) - \nabla\cdot(\mu_u\nabla u-\mu_v\nabla v) = 0,$$

where ψ_0 is the electrostatic potential function at the previous time level.

Equation (1.1) is essentially a consequence of (1.3) and the boundary conditions. There are at least two ways to proceed at this point; we show one method here, appropriate for models in two space dimensions, which has been successfully used in computations for some time. Equation (1.3) is rewritten

$$(1.4)\quad \nabla\cdot(\sigma\nabla(\psi-\psi_0)-f) = 0,\quad \sigma = \mu_u u+\mu_v v + \frac{\varepsilon}{h},$$

$$f = \mu_u(\nabla u-u\nabla\psi_0) - \mu_v(\nabla v+v\nabla\psi_0).$$

We introduce normalized stream functions $\theta_1,\ldots,\theta_{m-1}$, the amplitudes of the current components $J_1,\ldots,J_{m-1}$, and an interaction term ϕ, so that an expression for the total current density is

$$(1.5)\qquad \sigma\nabla(\psi-\psi_0) - f = \sum_{i=1}^{m-1} J_i \quad \nabla\times\theta_i + \nabla\times\phi \; ;$$

dividing by σ and taking the curl of both sides, we can choose the θ_i, to satisfy

$$(1.6)\qquad \nabla\times\left(\frac{\nabla\times\theta_i}{\sigma}\right) = 0, \quad i = 1,\ldots,m-1; \quad \nabla\times\left(\frac{\nabla\times\phi}{\sigma} + \frac{f}{\sigma}\right) = 0.$$

Equations (1.2-1.6) hold in the device interior Ω; the boundary $\partial\Omega$ is assumed divided into contact and insulating segments, $\partial\Omega_D = \partial\Omega_1 \cup \partial\Omega_2 \cup \ldots \cup \partial\Omega_m$, $\partial\Omega_N$ respectively, where $\partial\Omega_j$ is the j-th contact. No tangential current flows on $\partial\Omega_D$ and no normal current at $\partial\Omega_N$. Using (1.5), we can choose boundary conditions of the form

$$(1.7)\qquad \nu\cdot\nabla\theta_i = 0, \quad i = 1,\ldots,m-1, \quad \nu\cdot\nabla\phi = 0, \quad \text{on} \quad \partial\Omega_D;$$

θ_i = constant, $i = 1,\ldots,m-1$, $\phi = 0$ on each segment of $\partial\Omega_N$, where ν is the unit normal vector at each point on $\partial\Omega$.

The current/voltage components and boundary values of the θ_i are specified in terms of a constant matrix Q, of dimension $m \times m-1$, of rank $m-1$, with vanishing column sums. Then the change in the boundary value of θ_i, as contact $\partial\Omega_j$ is crossed in counterclockwise direction, is given by Q_{ji}, $j = 1,\ldots,m$, $i = 1,\ldots,m-1$; the θ_i are thus determined up to an arbitrary additive constant.

The θ_i can be obtained from (1.6), (1.7); ϕ need not be computed. Then multiplying (1.5) by $\nabla\times\theta_j/\sigma$ and integrating over the device area, we obtain

$$(1.8)\quad \iint_\Omega \nabla\times\theta_j\cdot\nabla(\psi-\psi_0) - \iint_\Omega \frac{1}{\sigma}\, f\cdot\nabla\times\theta_j = \\ = \sum_{i=1}^{m-1} J_i \iint_\Omega \frac{\nabla\times\theta_i\cdot\nabla\times\theta_j}{\sigma} + \iint_\Omega \frac{\nabla\times\theta_j\cdot\nabla\times\phi}{\sigma} \quad .$$

In (1.8), the first integral depends only on the boundary values of ψ, ψ_0, and θ_j; we denote the second integral by H_j and the third by Z_{ij}, $i,j = 1,\ldots,m-1$; the last integral vanishes identically. We rewrite (1.8) in vector form

(1.9) $$W - W_0 = H + ZJ,$$

where W is the vector of voltage components, W_0 the voltage components at the previous time level, $H = (H_1,\ldots,H_{m-1})^T$, $Z = [Z_{ij}]$, and $J = (J_1,\ldots,J_{m-1})^T$. These voltage and current components are related to the terminal quantities by

(1.10) $$I = QJ, \quad W = Q^T V, \quad W_0 = Q^T V_0 .$$

Finally, multiplying (1.9) by QZ^{-1} and using (1.10), we obtain (1.1), with

(1.11) $$A = QZ^{-1}Q^T, \quad G = -QZ^{-1}H.$$

We note that Z is positive definite symmetric, so that A is positive semidefinite symmetric, with kernel $(1,1,\ldots,1)^T$. As G is in the range of Q, the sum of its components vanishes as required. More details of this procedure are found in [3].

II PROPOSED METHOD

In practice, we compute successively the stream functions θ_i, $i = 1,\ldots,m-1$, the "impedance matrix" Z, the vector H, and from them the quantities A,G from (1.11). These quantities depend on the internal arrays ψ, ψ_0, u, v, and the time step h. As previously noted, it is computationally expensive to update these internal device arrays at each step; any method giving reasonable approximations to Z,H without doing this will thus result in a dramatic improvement in computation speed.

Our approach is to tabulate values of Z,H, as functions of the bias voltages and of the time step h. To build this table, the computations described in section 1 are performed using the stationary arrays $\psi = \psi_0, u, v$ corresponding to each bias voltage. The computation to build this "impedance matrix table" is slightly more complex than that of computing a stationary I-V table, since at each bias point, after solving the stationary problem for the arrays ψ,u,v, the computation of section 1 is performed for several values of time step. (In practice, however, this extra computation is not expensive relative to that of solving the stationary problem, as the elliptic systems for the stream functions tend to be well-conditioned and easily solved by iterative methods.)

Then at each time step of a transient computation, we simply interpolate Z,H from the stored table, as functions of the present device bias voltages and the time step h. The internal variables for devices treated in this manner never need to be obtained as functions of time.

Obviously there are important limitations to the use of such a method. An intrinsic approximation in this method is that the device is quasi-stationary, i.e. that the internal speed of the device is high compared with that of the circuit, so that the internal device arrays are always close to their stationary values, even though the bias voltages are changing as functions of time. This approximation appears to be used in all present "lumped element" models. Clearly there are cases in which this approximation is not justified, for example a device "struck" by a pulse of ionizing radiation, or a device subjected to an instantaneous change in its bias voltages. However, this method can be combined with ordinary time-dependent simulation in studying the response of full circuits. For example in studies of radiation induced upset of memory circuits, we can use the method of section 1 for the struck device and the above method for the other devices in the circuit, still achieving a substantial savings of computation. This type of computation was, in fact, the principal motivation for the development of this method [1].

Furthermore, the effects of the quasi-stationary approximation can be isolated and studied by this technique, as we show in an example below.

A second source of error in this method is that associated with the interpolation for Z,H, as functions of the bias voltages and h, between the tabulated values. This error can, of course, be reduced by tabulating Z,H at more points; in practice, a compromise must be chosen between the initial overhead (which can be amortized over many time-dependent computations) and the interpolation error.

Obviously, how these interpolations are done will affect the size of these errors. The results reported below were obtained using simple, piecewise linear interpolation in the bias voltages and in $\log h$. We have thus far obtained no improvement with higher order methods, presumably because Z,H were not tabulated at sufficiently many points. An experiment was performed in which Z^{-1} (equivalent to A) was interpolated instead of Z; the results were substantially worse. There are, however, some things which can be done to modestly reduce the interpolation errors. If the values of h which will be used for time-dependent runs are known in advance, then Z,H can be tabulated at precisely these values of h and this part of the interpolation avoided. Alternatively, it has been noted that Z tends to vary less rapidly, as a function of h for fixed bias voltages, than does H. The interpolation of H as a function of h can be avoided. Setting $W = W_0$, $J = J_s$, the stationary current component amplitudes, in (1.19), we see that $J_s = - Z^{-1}H$ is independent of h, even though Z,H are not. Thus if we tabulate J_s as a function of bias voltage, and interpolate Z as a function of bias voltage and time step, obtaining Z(h), we can find

$$(2.1) \qquad H(h) = - Z(h)\, J_s .$$

This procedure also assures that the computed asymptotic, i.e. stationary, values of the device currents will be independent of h.

The main advantage of this method, as compared with traditional lumped element models, is that one does not need to specify the topology or the element values of the model. These are automatically obtained. Results obtained by this method can be directly compared with those obtained by conventional transient simulation, with only the two sources of error described above between the results. Finally, the method can be used in problems where a device enters an unfamiliar region with respect to its bias voltages. For example, in the sample computation reported below we use this method with IGFET models with substantial forward bias on the source-substrate junction.

III A SAMPLE PROBLEM

Here we present the results of a sample problem, selected to demonstrate the capabilities and limitations of the proposed method. The circuit topology for this problem is shown in fig. 1. Two silicon n-channel IGFET devices of conventional geometry (metallurgical channel length 0.8μ, substrate doping 5×10^{16} cm^{-3}, oxide thickness 250 Å, and width 10μ) are coupled by their substrate contacts, together with two dependent current sources. It is perhaps not obvious whether this structure could act as a memory element, i.e. whether it has multiple stable stationary states, but if it does one of the source-substrate junctions is likely to become strongly forward biased, making the use of a conventional lumped-element model difficult.

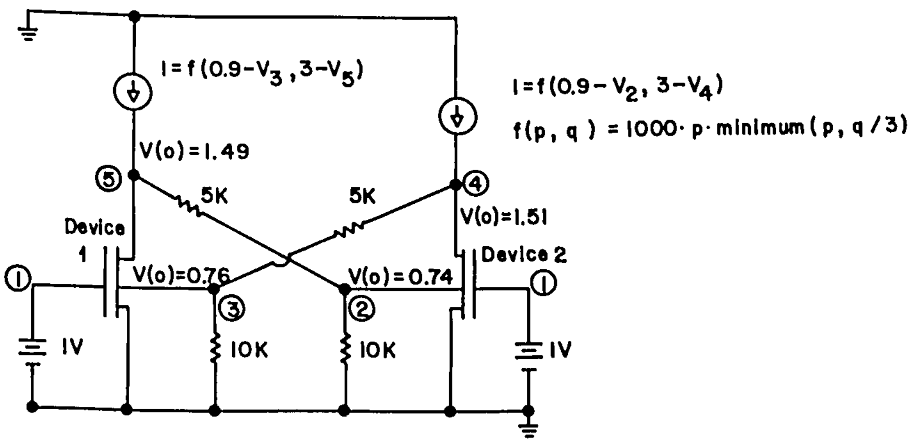

Fig. 1: Circuit topology

Also shown in fig. 1 are the device and node numbers, for reference in the figures below, and the initially applied voltages. Voltages are in volts, currents in microamperes resistances in ohms and time in nanoseconds throughout.

The evolution of this circuit, given these initial voltages, was calculated twice, by the two methods described. In the first calculation, the two devices were included as two-dimensional models and their internal arrays updated at each step, as described in section 1. The initial values of the internal arrays corresponded to stationary conditions for the given bias voltages. The results from this computation are indicated by "R" or "Real" in the figures below.

In preparation for the second run, the impedance matrix Z,H was tabulated, assuming stationary conditions, as a function of drain (to source) voltage, substrate (to source) voltage, and time step. The gate voltage was held constant at one volt throughout, consistent with the circuit diagram. The drain voltage was varied uniformly in 0.5 volt steps from zero to 3V; the values of substrate voltage at which Z,H were tabulated were 0, 0.5, 0.7, 0.8, 0.85, and 0.9V. The values of time step began at 0.001 nsec and were successively increased uniformly by factors of $\sqrt{2}$. The values of these parameters occurring in the computation lie within these limits, so that no extrapolations were needed.

The second computation of the transiet response of the circuit was performed by the method described in section 2, i.e. interpolating the values of Z,H for each device at each time step. The results of this computation are designated by a "Z" in the figures below.

The improvement in computation speed by this second method is at least a factor of several thousand, ignoring the time required for building the impedance matrix table. A precise figure was not obtained, because the computation time for the second method is dominated by overhead. The same values of time step were used for both computations, so the number of time steps is the same in the two cases.

The results of the two computations are compared in figs. 2-6. The two substrate voltages as functions of time are shown in fig. 2; the two drain voltages in fig. 3. The gate and drain currents of device 1 (the device ending up in the "on" state) are shown in fig. 4, and those of device 2 (ending up in the "off" state) in fig. 5. The substrate currents for both devices are shown in fig. 6.

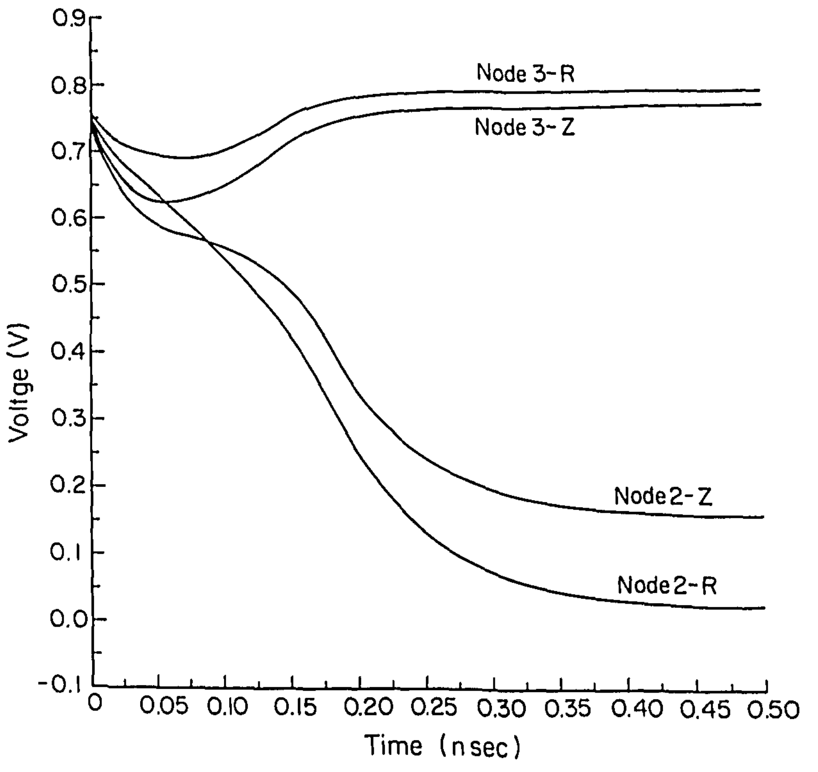

Fig. 2: Substrate voltages as functions of time

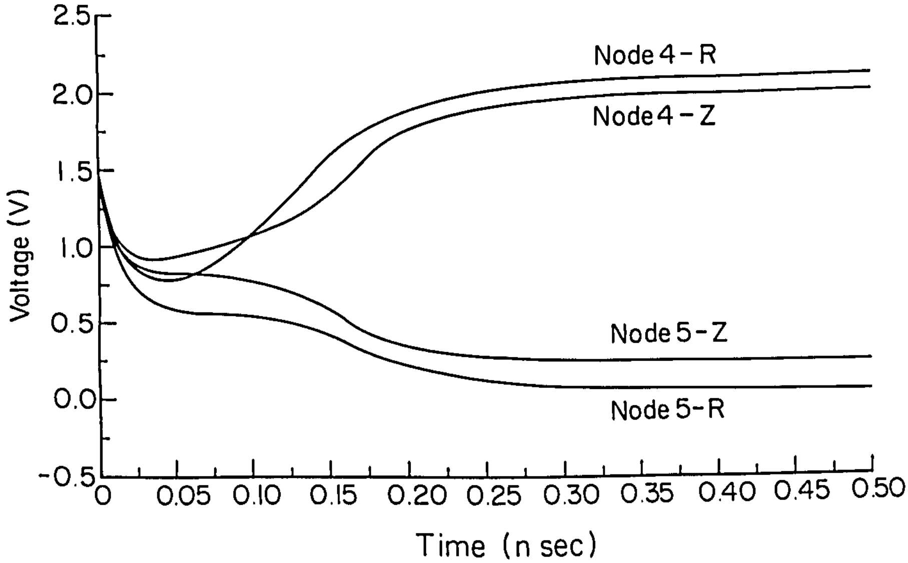

Fig. 3: Drain voltages on functions of time

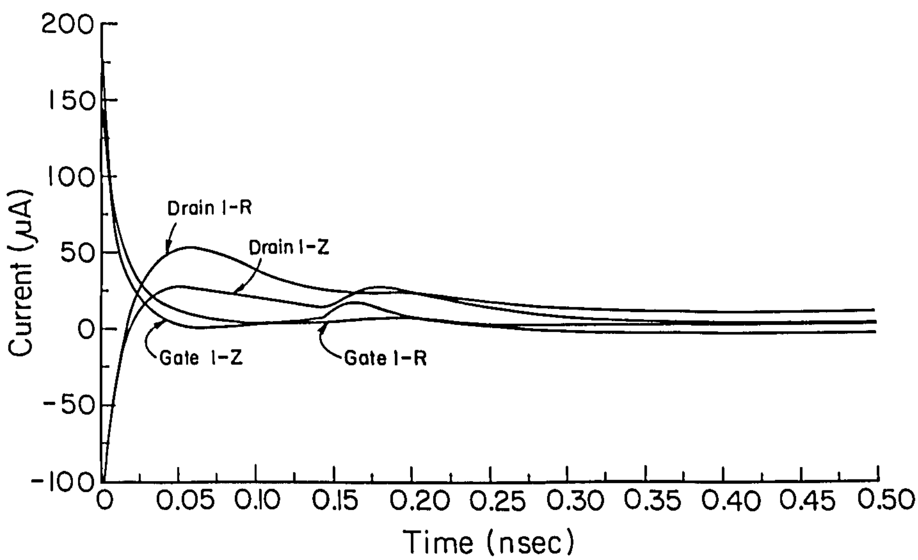

Fig. 4: Drain and gate current for device 1

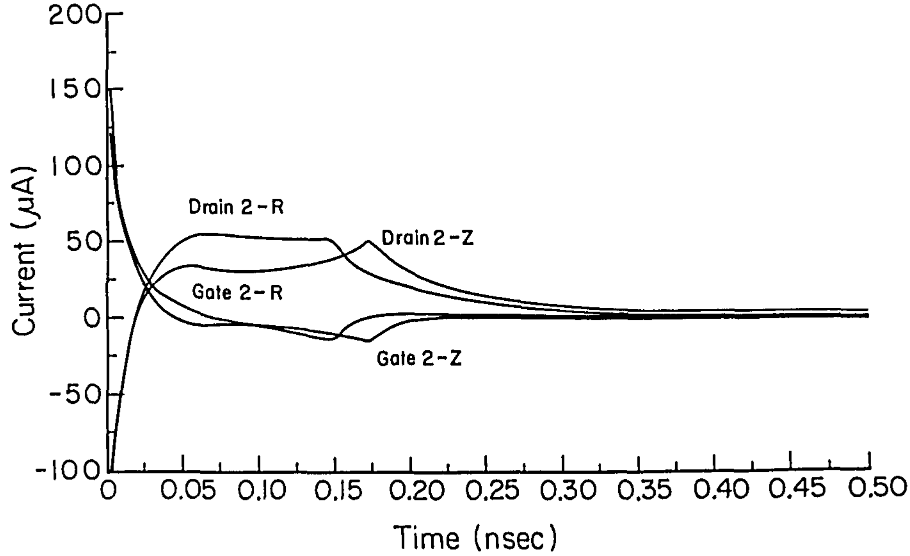

Fig. 5: Drain and gate current for device 2

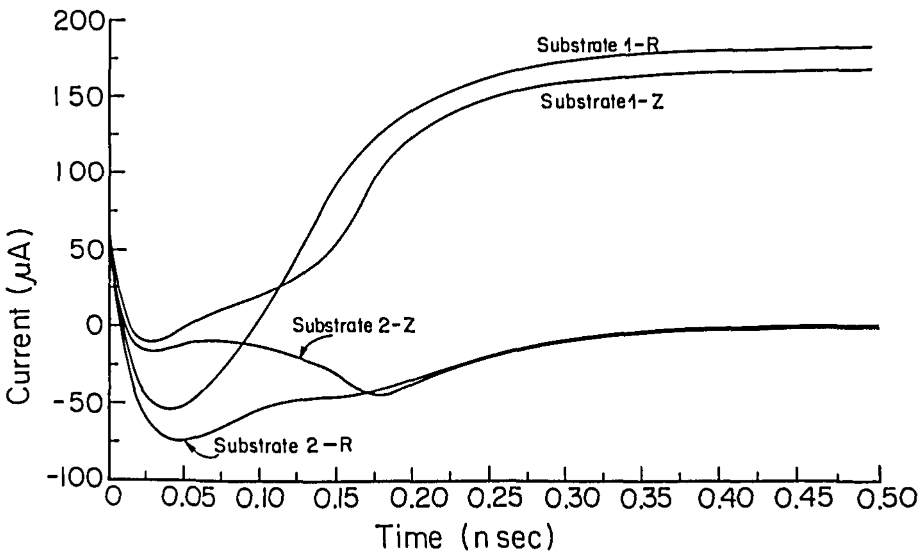

Fig 6: Substrate currents as functions of time

The discrepencies in the results for relatively large time, say $t \geq 0.3$ nsec, arise from the error of interpolation in the impedance matrix table. For such times, the circuit is nearly in a stationary state and the quasi-stationary approximation used in tabulating the Z,H arrays is clearly well justified and not a significant source of error. These errors could be made smaller by tabulating Z,H at more bias voltage points.

At earlier times, both sources of error contribute to the differences in the results. For example, in fig. 4, the abrupt change in slope of the two "Z" curves at approximately 0.15 nsec is also interpolation error, caused by the voltage at node 2 (the substrate of device 2) crossing the table value 0.5V at this time. In fig. 6, the large differences between the "R" and "Z" curves around t = 0.05 nsec presumably includes a substantial contribution from the quasi-stationary approximation - this will be discussed further below. In fig. 5, both pairs of curves show sharp changes in slope between t = 0.15 and t = 0.18; the qualitative behavior is presumed correct, as it also appears in the "R" curves. The significant difference in the drain current curves in fig. 5, between t = 0.05 and t = 0.15, is again primarily interpolation error, as we shall now describe.

IV ISOLATION OF THE QUASI-STATIONARY APPROXIMATION

As is well known, there are two basic sources of error in lumped element models. One is in choosing a model topology, or in interpolating in tables. In principle, this error can be made very small by additional computation and perhaps cleverness in choosing a model. The second approximation comes from assuming that the device is in a stationary condition for purposes of evaluating the model parameters (or impedance matrix in the present terminology). This approximation appears to be intrinsic, in the sense that no realistic alternative is known. Since no practical method is known for reducing this second source of error, short of abandoning lumped element modeling entirely, it is desirable at least to have a method for appraising its magnitude. The method described in section 2 permits this to be done by an additional computation, which we now describe as applied to the sample problem discussed in section 3.

From figures 5 and 6, we note that for $t \leq 0.15$ substantial differences are observed in the values of the drain and substrate currents of device 2, as computed by the two methods. These two quantities were selected for further study. We want to know how much of these errors are due to interpolation and how much are intrinsic, i.e. the result of the basic quasi-equilibrium approximation.

To determine this, an additional computation is performed. At each discrete time level (up to 0.2 nsec), a stationary solution for the device is obtained, using the bias voltages as obtained in the computation with the "real" devices. (This is not as expensive a computation as it might appear, as these bias voltages change relatively little between successive time levels, so that 2 or 3 Gummel iterations for each time level generally suffices.) Once these stationary internal arrays are obtained, the computations (1.6-1.9) are repeated, using these "quasi-stationary" internal arrays and the values of W, W_0, h as obtained in the "real" computation. Thus the only source of differences in the computed values of the terminal currents is that in the "real" case, the actual, transient arrays ψ, ψ_0, u, v are used, whereas in the "quasi-stationary" case these arrays are the stationary ones for the same bias voltages. The difference between these results is thus precisely the effect of the quasi-stationary approximation .

The results are shown in fig. 7. For comparison, we also show the stationary terminal currents at each point. It will be noted that these have no relation to the computed values by either method, which is an indication of the relative importance of the reactive part of the model if an accurate description of such transients is to be obtained.

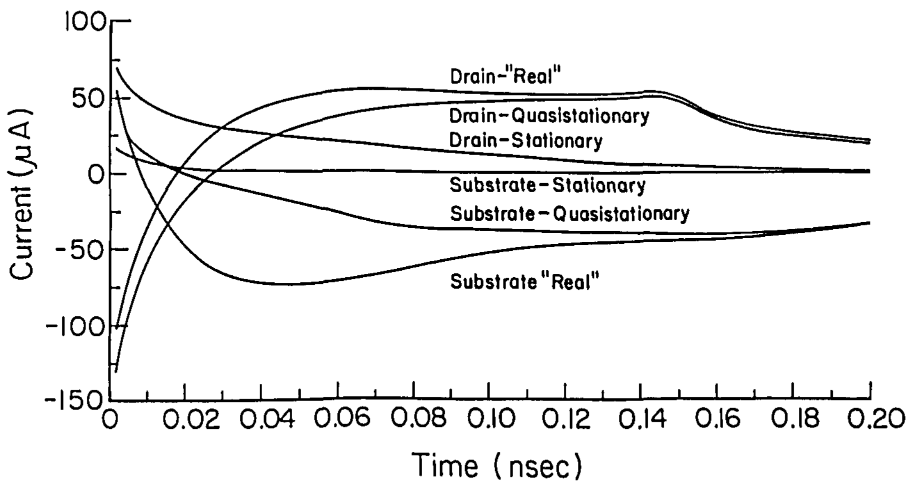

Fig.7: The quasi-stationary approximation

The comparison is significantly different for the drain and substrate currents. For the drain current, the difference between the "real" and "quasi-stationary" models is relatively modest and decreases rapidly after about t = 0.07. Thus in fig 5, most of the difference between the "R" and "Z" curves for the drain current is attributable to interpolation in the impedance matrix table. In contrast, a large, qualitative difference is observed in the substrate current results, up to about t = 0.09. This is the effect of the quasi-stationary approximation, and one is forced to conclude that such lumped element models could not be used if this portion of the transient response is to be obtained accurately. For larger times, we see that the quasi-stationary approximation becomes quite good, as evidenced by the approach of the two pairs of curves. This is of course expected, or the circuit begins to approach a stationary condition after the initial transient period.

V. REFERENCES

1. C. L. AXNESS, J. S. FU and H. T. WEAVER, "Two-dimensional simulation of single event upsets in SEU-hardened CMOS RAM cells," Proc. Fourth Int. Conf. on Num. Anal. of Semiconductor Devices and Integrated Circuits, J.J.H. Miller, Ed., Boole Press, Dublin (1985).

2. V. BELEVITCH, "Classical Network Theory, "Holden-Day, San Francisco, 1968.

3. M. S. MOCK, "Analysis of Mathematical Models of Semiconductor Devices," Boole Press, Dublin, 1983.