Efficient Transient Device Simulation with AWE Macromodels and Domain Decomposition

Howard C. Read, Shigetaka Kumashiro*, Andrzej Strojwas
Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA 15213
*NEC Corporation, Kawasaki, Japan

Numerical simulation of multiple semiconductor devices is necessary to analyze dynamic two- and three- dimensional interactions among devices such as in MOS inverters or more complicated gates. With the advent of complete 3D process simulation, an alternative to brute force device simulation must be found for large contiguous silicon regions. In the realm of transient simulation, many of the spatially-dependent device variables on a finite-difference grid need not be temporally integrated at each time step if a simpler macromodel exists. This work extends the strategy of computing macromodels with Asymptotic Waveform Evaluation (AWE) in passive devices to active devices [1]. An event-driven simulator combines AWE-generated macromodels with traditional numerical integration and uses an error prediction/correction method based on boundary matching to assure the temporal validity of the AWE macromodels.

To lower the computational overhead of managing different temporal integration lengths, we group finite-difference nodes into distinct *subdomains*, each with a unique time-step, using the local time-step difference between nodes as a distance function. A hierarchical, single-linkage clustering algorithm yields time-step density-contour clusters using this criterion..

AWE macromodels in each subdomain are computed using a linearized state equation describing the coupled semiconductor device and circuit equations. We choose $\mathbf{x} = (\psi, n, p)$ as state variables, because they make Poisson's equation linear and the current-continuity equations nearly linear with respect to n and p. Using a control-volume approach in 2-D and expanding in a Taylor's Series at a time point, t_j, yields:

$$\boldsymbol{C}_C\boldsymbol{x}(t) = \boldsymbol{C}_A\left[\frac{\partial F}{\partial x(t)}\right](\boldsymbol{x}(t) - \boldsymbol{x}(t_j)) + \boldsymbol{C}_A\boldsymbol{F}(\boldsymbol{x}(t)) + \boldsymbol{B}\boldsymbol{u}(t) \qquad (1)$$

where $\boldsymbol{C}_C$ and $\boldsymbol{C}_A$ are control-volume matrices, $\boldsymbol{F}(\boldsymbol{x}(t))$ is the matrix of the semiconductor equations, $\boldsymbol{x}(t)$ is the vector time-varying device variables and $\boldsymbol{B}\boldsymbol{u}(t)$ is the vector of time-varying boundary conditions. It has been shown [1] that the computation of the moments corresponding to (1) in the form $\dot{x}(t) = Ax(t) + Bu(t)$ can be expressed:

$$\boldsymbol{C}_A\left[\frac{\partial F}{\partial x(t)}\right]m_r = \boldsymbol{C}_C m_{(r-1)} \qquad (2)$$

Moments are calculated recursively using the *original* Device Jacobian Matrix. We have found that a single-pole (two-moment) approximation for each device variable gives an accurate approximation to the true response. The macromodels thus represent time-varying boundary conditions compactly as the sums of exponential, step and ramp responses.

The event-driven simulator, AWETOPSY (AWE for Transient Optimized and Partitioned Simulation) schedules *events* that indicate when a temporal integration must be performed. The subdomain causing the event is *active* and others are *dormant*. Events are scheduled according to time-step prediction based on *a priori* error estimation. After each event, the difference between predicted and integrated charge quantities are checked in each subdomain and if this *a posteriori* error does not satisfy the criterion, the event is rejected and the present time is reduced.

We have analyzed two methods of error calculation: an explicit integration and a boundary-value matching method. The first is based on a truncated Taylor Series expansion; it has been shown [1] that the error vector $\mathbf{e}(\tau)$ is bounded by:

$$|\boldsymbol{e}(\tau)| \le \left|\int_0^{\tau} \boldsymbol{I}_{error}(s)ds\right| \quad \text{where } \boldsymbol{I}_{error}(t) = -\dot{\boldsymbol{x}}^{approx}(t) + \boldsymbol{F}(\boldsymbol{x}^{approx}(t)) + \boldsymbol{B}\boldsymbol{u} \qquad (3)$$

where $\boldsymbol{x}^{approx}$ are the approximated solutions (from AWE) and $\boldsymbol{I}_{error}$ represents an error current vector. This bound is pessimistic and also expensive to calculate due to the integration. The latter method uses a polynomial to approximate the (possibly non-monotonic) error function between two points, $[t_j, t_{j+1}]$. By matching the calculated state variables and their derivatives to the approximated values at the end points, we find a third order error polynomial, $e_i(\tau)$ for the i^{th} state variable. The *a posteriori* error is calculated using this polynomial. For the *a priori* case, values at t_{j+1} are not available, but it is shown in [1] that modified moments shifted by $(1/h_j)$ provide the necessary boundary conditions. Both polynomial error functions have been found to give a conservative upper limit on the calculation error.

We have applied AWETOPSY to the simulation of a single 0.8um NMOSFET and the full structure of an 0.8um CMOS inverter with capacitive loading (Figure 2.) The domain decomposition of the each is shown in Figure 1 and reflects the time-step distribution during a simulation event. In both examples, the smaller subdomains are combined into one subdomain. Only 50% of the nodes are *active* on average. The overhead caused by macromodeling takes four Newton iterations of the **F** matrix per event which takes 25% of the total simulation time as shown in Table 1. Near the switching phase of the inverter and switch, all the domains are activated, but elsewhere a large amount of latency is exploited because the larger (bulk) subdomains are mostly dormant (Figure 3.) Table 2 shows the increased number of time-steps but decreased overall simulation time due to this latency as compared with classical temporal simulation based on truncation-based error control. Figures 4 shows the close match of fully numerical and macromodeled simulations. As the size of the simulation domains grow,, this approach remains nearly as accurate as full simulation with improved efficiency; although small examples are still best handled by traditional simulation, those with coupling distributed over large spatial regions are best suited to this divide and conquer technique.

References

1. S. Kumashiro, *Transient Simulation of Passive and Active VLSI Devices Using Asymptotic Waveform Evaluation*, Ph. D Thesis, Carnegie Mellon University, Pittsburgh, PA, Research Report No. CMUCAD-92-40

Table 1. Simulation Times with Macromodeling

Operation	% of Total Time	
	NMOS	CMOS
Hierarchical Decomposition	0.2	0.06
AWE *A Posteriori* Error Eval.	14.5	15.5
AWE Macromodeling/Prediction	20.6	26.6
Solution of Device Equations	64.7	58.8

Table 2. Comparison of Simulation Methods (DecStation 5000)

Simulation	Total Sim. Time (sec)	# Time-Steps	Average Tim (sec) / Time-Step	Total # Mesh Points	Average # Active Points
AWETOPSY (NMOS)	935	21	44.5	840	481 (57%)
Classical (NMOS)	910	14	60.6	840	840 (100%)
AWETOPSY (CMOS)	7666	27	283	2167	1284 (59%)
Classical (CMOS)	8350	23	363	2167	2167 (100%)

FIGURE 1. Time-step distribution.Darkest are smallest (t=4e-12). Solid lines are clustered subdomains, dashes are junction boundaries

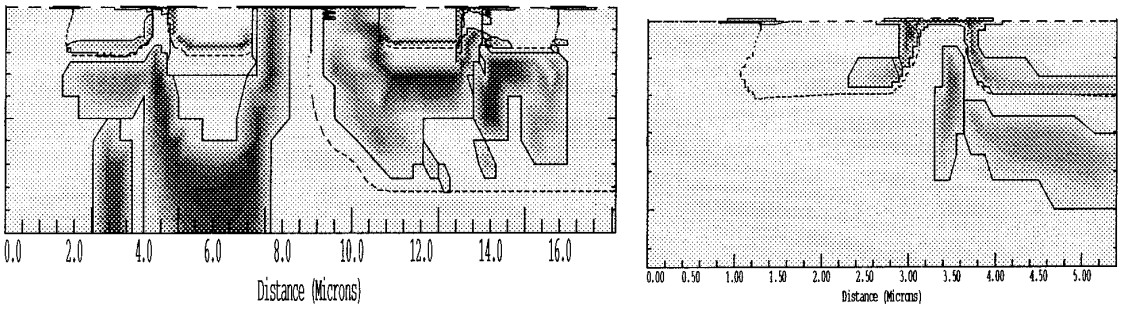


FIGURE 2. Application of macromodeling to NMOS switch and CMOS Inverter with capacitive circuit element.

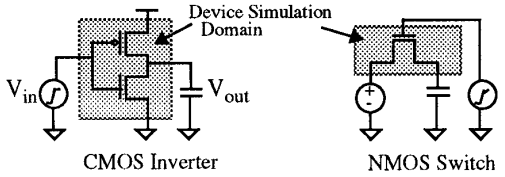


FIGURE 3. Event Maps showing events in largest (bulk) subdomain versus those in the smaller (channel region) subdomain group (all forced to same time-step.)

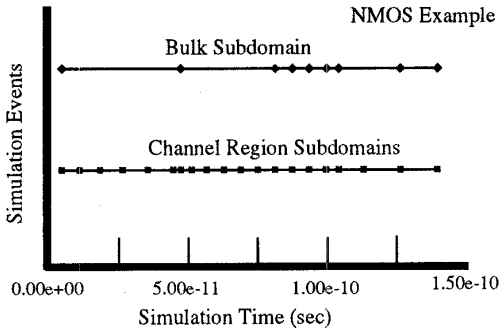
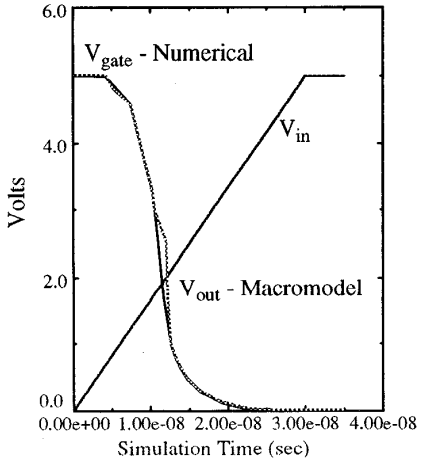
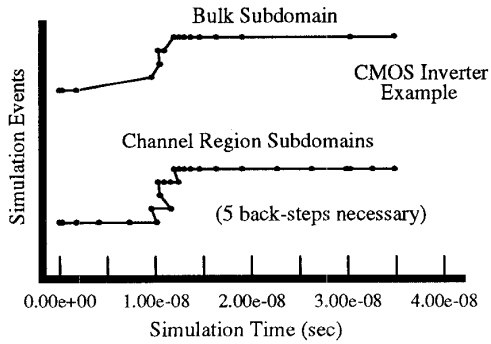


FIGURE 4. Comparison of CMOS Inverter output using macromodeling to that using full numerical integration.





Each dot represents a numerical integration. Each level represents a back-step.