DRAM Design Utilizing 2D and 3D Device Simulation

E. M. Buturla, J. B. Johnson, T. D. Linton, S. H. Voldman
IBM General Technology Division
Essex Junction, Vermont 05452, USA

Abstract - Numerical simulation of semiconductor operation is important in the development of DRAM technologies since it provides insight into complex phenomena that cannot be obtained through experimentation or simple models. Simulation tools also provide a controlled and repeatable experiment that can yield information that cannot be measured with present-day equipment. For these tools to be useful in a practical environment, they must be physically accurate, computationally robust and usable by persons other than the software developers. These tasks are difficult in the present fast-paced DRAM business since decreasing device dimensions inevitably draw out physical effects that, in earlier technologies, were of second order and could be neglected. Recently, we have added new formulations for leakage mechanisms which we will report here. In addition, we also discuss some of the computational and visualization aspects of 3D device simulation which can be very imposing because of the large amount of data in the models.

Physical Model

Thermal- and electric field-dependent leakage mechanisms are becoming increasingly important design considerations for trench DRAM structures [1- 3]. These mechanisms affect reliability, data retention time in stacked capacitor and trench DRAM structures and low power and low temperature devices. These phenomena can be accurately modeled across a wide spectrum of device structures and operating conditions by standard drift-diffusion simulation techniques with the addition of physical models and associated numerical discretization methods. In this work, we utilize a modification of the standard generation model to account for the leakage mechanisms.

A bulk generation model which includes a generalized Shockley-Read-Hall (SRH) term and band-to-band tunneling term of the form

$$U = \frac{(np - n_i^2)}{\tilde{\tau}_n(n + n_1) + \tilde{\tau}_p(p + p_1)} - A F^2 e^{-B/F} \tag{1}$$

has been proposed and implemented [4 – 6] . In this model, the generalized time constant, $\tilde{\tau}$, is of the form

$$\tilde{\tau}_k = \frac{\tau_k(N_D, N_A, T)}{\aleph_k(F, T)} , \tag{2}$$

where

$$\tau_k(N_D, N_A, T) = \frac{\tau_{k0}(T)}{\left(1 + \frac{N_D + N_A}{N_{Ref}}\right)} , \tag{3}$$

$$\aleph_k(F, T) = \aleph_k(F, T)_{CL} + \aleph_k(F, T)_{QM} . \tag{4}$$

The doping dependence of the trap state density and zero-applied electric field capture cross-section temperature dependence are contained in $\tau_t(N_D, N_A, T)$, the electric field emission rate enhancement factor $\aleph_t(F, T)$ is the sum of a classical trap-state barrier lowering term (the Frenkel-Poole effect) and a quantum mechanical trap state-to-band term. A and B are physical parameters (see[5]) and F is the local electric field. Generation due to band-to-band tunneling is modeled using the Kane formulation [7] . The new generation model has been recently added to our device simulator, FIELDAY II.

Numerical Algorithm

FIELDAY II [8] is a general two- and three-dimensional simulation program created for use by technology developers who are not necessarily simulation experts. It utilizes a Voronoi based control cell approach.

Given an arbitrary set of points in two dimensions, the Voronoi polygons for that set of points denote the regions in the plane closest to each point in the set. The shared polygon sides between adjacent points are straight lines, of zero length in degenerate cases, perpendicular to, but not necessarily intersecting, the (imaginary) line connecting the adjacent points. If the length of the side is zero (it is never negative), then the two points are not considered to be connected and the current flow between them does not enter into the analysis.

Given an arbitrary set of points in three dimensions, the Voronoi polyhedra for that set of points denote the regions in space closest to each point. In a Voronoi polyhedra mesh, the shared polyhedra sides between adjacent points are planes, occasionally of zero area, perpendicular to the (imaginary) line connecting the adjacent points. If the area of the side is zero, then the two points are not connected and current flow between them does not enter into the analysis.

The Voronoi polygon mesh eliminates the obtuse angle problem of the control cell method: a positive current cross-section is provided for every pair of connected nodes, for any node distribution in two or three dimensions, not merely those forming regular point meshes, thus satisfying the primary requirement of Mock's analysis of consistency and convergence of the current continuity equations. The physical consistency of the discretization of both Gauss' Law and the current continuity equations is assured by the orthogonality of the current cross-sections and the calculated flux components.

Connecting each pair of points that share a Voronoi polyhedron side by a straight line yields a triangulation of the original point set in two dimensions, and a tetrahedra mesh in three dimensions, known as the Delaunay tessellation. Since these two meshes are complementary, in practice one finds the Delaunay tessellation of the set of mesh points, a process for which efficient algorithms exist, and then extracts from this triangle mesh the Voronoi polyhedra.

The Voronoi polyhedra mesh is the logical mesh to use for the control cell method. By using it one is freed from restricting the mesh points to those from regular arrays, as in the finite difference method, or that from prisms, as in the original FIELDAY method: points can be added anywhere in the mesh without worrying about overlapping control areas. This leads naturally to efficient mesh refinement procedures.

A typical finite element program calculates on an element by element basis to compute the global coefficient matrices. In our approach, the connection between two nodes, actually an edge of the mesh, is the basic computational element. All data for these edges is kept in vector form. Solution of the resulting matrices is by use of bi-conjugate gradient iterative solvers.

FIELDAY II has written which utilizes the Voronoi control cell method. The program runs on the IBM 3090-600 machine which is has a vector facility and runs under the MVS/EA (Multiple Virtual System/Enterprise Architecture). This architecture allows addressing up to 2 gigabytes. Our software utilizes this address space by trading memory for speed. We calculate and save in vectors where at all possible rather than recalculate to save space. We also utilize FIELDAY II on the Risc System/6000 both for interactive

tasks such as pre- and post-processing as well as for 2D and 3D device simulations. We prefer to use the mainframe environment for large 3D analyses since the amount of DASD is unlimited to us from a user point of view.

Analyses of problems to 350,000 unknowns have been made to date. These problems require memory sizes of about 700 megabytes. Much of the memory is virtual and of course, the actual distribution and manipulation of the memory is virtual and under control of the operating system. For problems of this size, the solution of a single matrix equation requires about 600 seconds of CPU time. This time fluctuates with degree of non-linearity of the problem.

To summarize, FIELDAY II implements the semiconductor device equations in control-volume form. Each mesh node is associated with a unique area of the device in a two-dimensional problem and unique volume in three dimensions. These nodal regions are the Voronoi cells of the device structure. The generation rate computed via equation (1) at each mesh node must be distributed within the area or volume associated with that node, which requires a corresponding nodal value of F.

To integrate the generation models easily into the well established Shockley-Read-Hall terms, which are typically calculated at each mesh node and simply weighted by the control-volume area, we use a novel least-squares technique described elsewhere[6] to compute the local field F. On Cartesian grids, this least-squares technique has the desirable property of being identical to the first-order finite-difference technique under the proper weighting scheme of the equations, and is independent of the simulation dimension or of the element types used in the discretization.

Additional complications exist when calculating leakage currents since finite precision arithmetic limits the ability of a simulation code to resolve small currents. In FIELDAY II, we solve this problem by solving the device equations in 16-byte arithmetic. If requested, FIELDAY II passes the double-precison (8-byte) solution to extended-precison (16-byte) routines. While the extended-precision code is approximately 3X slower than the double-precision code on an IBM 3090 system, the double-precision code does most of the work; typically, only a few extended-precision Newton iterations are needed for each bias point. The benefit is that extended-precision solutions have approximately twice the decimal digit-precision of the double-precision code. For certain problem, this permits FIELDAY II to resolve currents that could not otherwise be extracted.

Application Example

The physical models and their implementation in FIELDAY II discussed above have been used to analyze leakage mechanisms in a substrate-plate-trench (SPT) DRAM cell [9]. This model is valuable in analyzing the p+ channel gate-induced drain leakage (GIDL) mechanism and the trench gate-induced diode leakage (TGIDL) mechanism. It has been shown experimentally that both these mechanisms consist of an electric-field dependent thermal generation and band-to-band tunneling mechanism [2,3]. Our simulations show a strong correlation between the TGIDL mechanism in a SPT cell DRAM structure and the field-dependent leakage models discussed in this paper. Note that extended-precision arithmetic was necessary to resolve the node leakage currents. Our earlier work was to simulate these effect in 2D and recently, we have studied these effects in 3D. The 3D studies are much more complex to interpret since so much more simulation data results and we have found that advanced visualization techniques and software are required for 3D.

Scientific visualization software for semiconductor technology simulation must meet several requirements. First, it must provide a simple human interface with a seamless connection to the simulation data base, it must be highly interactive and the user must therefore be able to quickly manipulate the visualization package in order to study the phenomena of interest. This requirement also impacts the computer hardware being used for visualization. The graphics processing must be fast enough to allow the user to try many different visualization methods in a reasonable amount of time. In our work, we utilize the IBM Risc System/6000 as the hardware of choice since it has a powerful local computing capability. We run our simu-

lations on the host IBM 3090 and then do the visualization locally. This combination has proven to be very flexible and powerful.

A number of scientific visualization packages are being used at IBM for semiconductor technology simulation. These include both internally developed software as well as vendor software packages such as the Data Visualizer from Wavefront Technologies [10] With this software, flexible 2D and 3D visualizations can easily be made including scalar contour maps, vector displays, point and volume rendering as well as particle trace and animation capabilities. These are combined with the ability to utilize both regular and unstructured grids in both 2 and 3 dimensions.

Concluding Remarks

We have shown how topical physics effects have been implemented into a state of the art device simulation program and how simulation aids in the design and analysis of the semiconductor devices used in state-of-the-art integrated circuits. If used properly, simulation can substantially help in cutting development costs. Indeed, some researchers are of the opinion that cost savings can be as large as 40% of the development cost.

REFERENCES

[1] S. Voldman, J. Bracchitta, and D. Fitzgerald, "Band-to-Band Tunnelling and Thermal Gate-Induced Drain Leakage," 46th Annual Device Research Conference, June 1988.

[2] S.Voldman, A. Bryant and W. Noble, "Vertical Trench Gated Diode Leakage," 45th Annual Device Research Conference, June 1987.

[3] W. Noble, S. Voldman and A. Bryant, "The Effects of Gate Field on Leakage Characteristics of Heavily Doped Junctions," IEEE Transactions on Electron Devices, Vol. 36, No.4, p. 720, 1989.

[4] S. Voldman, "Unified Generation Model for the Analysis of Electric Field Dependent Leakage Mechanisms in Silicon Devices," PhD. Thesis, University of Vermont, 1991.

[5] S. Voldman, J. Johnson, T. Linton and S. Titcomb, "Unified Generation Model with Donor and Acceptor-Type Trap States for Heavily Doped Silicon," IEDM Technical Digest, p. 349, December 1990.

[6] J. Johnson, T. Linton and S. Voldman, "Discretization Methods and Physical Models for Simulating Leakage Currents in Semiconductor Devices with Applications to Modeling Trench DRAM Cells," NASECODE VII, April 9-12, 1991, Coper Mountain, Colorado.

[7] E. O. Kane, "Theory of Tunneling," Journal of Applied Physics, Vol. 32, No.1, p. 83, 1961.

[8] E. Buturla, J. Johnson, S. Furkay and P. Cottrell, "A New Three-Dimensional Device Simulation Formulation," NASECODE VI: Proceedings of the Sixth International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits, Boole Press, Dublin, Ireland, p. 291, 1989.

[9] D. Kenny, E. Adler, B. Davari, J. DeBrosse, W. Frey, T. Furukawa, P. Geiss, D. Harmon, D. Horak, M. Kerbaugh, C. Koburger, J. Lasky, J. Rembetski, W. Schwittek and E. Sprogis, "16-MBIT Merged Isolation and Node Trench SPT Cell (MINT)," 1988 Symposium on VLSI Technology, May 1988.

[10] Wavefront Technologies, Inc., 530 East Montecito Street, Santa Barbara, California.