# SIMULATION OF SUBMICRON Si and GaAs DEVICES: PROBLEMS AND TECHNIQUES

Jeffrey Frey
Cornell University,
Ithaca, New York 14853

## 1. ABSTRACT

Simulation approaches for Si and GaAs semiconductor devices have generally followed somewhat different lines, but as the Si devices are shrunk toward the submicron size range, the problems being addressed and the techniques being used to address them for both materials are converging. The physical problems common to devices made of both materials in smaller size ranges are enumerated, and the techniques currently used for their solution are reviewed. These techniques are now moving beyond that of iterative solutions of the current-flow and Poisson equations, toward descriptions of charge carrier flow in terms of average carrier energy or temperature. A means of determination of heated electron temperature is also described. Finally, particular problems in the simulation of devices made of materials such as GaInAs or InP, and such devices as the HEMT and the heterojunction BJT, are also discussed.

## 2. "TRADITIONAL" SEMICONDUCTOR DEVICE SIMULATION

"Semiconductor Device Simulation" is here taken to mean the determination of measurable device terminal parameters as functions of geometry and applied boundary conditions by means of solution of the basic carrier transport and continuity equations and Poisson's equation. Additional variables, such as surface-state densities in MOS oxides, may be taken as independent or dependent. "Traditional" simulations utilize classical descriptions of carrier transport, such as the equilibrium relationship of electron velocity to electric field. In addition to determining measurable parameters, traditional simulations may also provide information on internal electric field and charge distributions and other quantities of interest. Whether analytical or computer-based, traditional simulations have been able to provide descriptions of devices with active lengths greater than about two microns which agree well with experiment.

Analytical models of MOSFETs generally proceed on the basis of Shockley's gradual-channel model, in which the longitudinal field in the channel is taken to be much smaller than the transverse field [1]. While Shockley originally used the constant-mobility relationship $v=\mu E$, later workers modified his model to incorporate the effects of velocity saturation.

The analysis of Schottky-barrier, Metal-Semiconductor FETs (MESFETs) can also proceed traditionally along lines laid down by Shockley, but in this case the material of greatest interest is GaAs which has a velocity-field relationship that is considerably more complicated than that of Si. The GaAs MESFET can be analyzed using a v-E curve similar to that of silicon, and such an analytical treatment can be made to agree with experimental results for specific devices by adjustment of low-field mobility or saturation velocity. However, a continuous analytical solution valid through saturation has not been available. Further, diffusion effects are difficult to incorporate. It is possible, however, to obtain a one-dimensional analytical MESFET solution by breaking the real GaAs v-E curve into three parts (constant mobility, constant negative mobility, saturation velocity) and incorporating some measure of diffusion currents [2].

True two-dimensional simulation of FETs cannot be performed analytically. For MOSFETs, the 2D computer programs GEMINI [3] and MINIMOS [4] have proved useful within the limitations of the approximations built into them in order to make their use economical. For MESFETs, CUPID [5] and CADDET [6] may be used. FIELDAY [7] can perform 2D or 3D simulations of MOS or bipolar devices, but is not generally available; SEDAN [8] and LUSTRE [9] are one-dimensional bipolar device simulators that are.

## 2.1 Basic Equations

Traditional two-dimensional device simulators solve the basic equations:

a) Poisson's equation

$$\nabla^2\psi = \frac{-q}{\epsilon_s \epsilon_o}(N_D - N_A - n + p) \qquad (1)$$

b. Carrier transport equations.

$$J_n = q\, n\, v_n + q\, D_n\, \nabla n, \qquad (2)$$

and

$$J_p = q\, p\, v_p - q\, D_p\, \nabla p, \qquad (3)$$

where

$$v_n = \mu_n E, \tag{4}$$

$$v_p = \mu_p E, \tag{5}$$

and

$$E = -\nabla\psi \tag{6}$$

Though Butcher [10] demonstrated that the diffusion term should be the form of $q(\partial/\partial x)\{Dn\}$ instead of that of $q D\{(\partial/\partial x)\}n$ for one-dimensional problems, $q\{(\partial/\partial x)D\}n$ may usually be neglected. The mobilities $\mu_n$ and $\mu_p$ and the diffusivities $D_n$ and $D_p$ as functions of electric field, and the diffusivities are also anisotropic, not scalars but tensors.

c) Continuity equations

$$\frac{\partial n}{\partial t} = \nabla \bullet (J_n/q) + G, \tag{7}$$

and

$$\frac{\partial p}{\partial t} = \nabla \bullet (J_p/q) - G, \tag{8}$$

where G represents the carrier generation-recombination term.

d) Total current equation

$$J = J_n + J_p - \epsilon_s\epsilon_o \frac{\partial}{\partial t} \nabla\psi \tag{9}$$

## 2.2 Materials Parameters

Equations (4) and (5) may be replaced by analytical curve fits. For electrons in silicon, for example, at the doping level of $10^{17}/cm^3$, a suitable form is:

$$v(E) = \mu_o E/\{1+(E/E_c)^2\}^{1/2} \tag{10}$$

where $\mu_o$ is the low field mobility and $E_c$ is the "critical field".

Diffusivity tensors are hard to find for any material. Since transport properties in semiconductors with impurity densities of less than $10^{17}/cm^3$, at an electric field of greater than 10 kV/cm, are almost the same as in pure material, the diffusivity parallel to the electric field for high purity Si can be extrapolated to low fields and used for doped material, using a modified Einstein relation [11]. The extrapolation requires matching the relation

$$D(E) = \frac{kT}{q}\frac{v(E)}{E} + \frac{2}{3}\tau \{v(E)\}^2 \qquad (11)$$

where $\tau$ is the energy relaxation time, to the pure-Si diffusivity at 10 kV/cm. The anisotropy of the Si diffusivity is generally negligible [12].

In GaAs, up to electric fields of 25 kV/cm the steady state drift velocities of electrons at doping levels of zero and $10^{17}/cm^3$ have been calculated using Monte Carlo techniques [13]. Velocity-field data up to electric fields greater than 100 kV/cm has been obtained experimentally [14]. This data can be combined with the low-field curve for numerical analysis of GaAs MESFET's. An analytical form of these curves, such as [15]

$$v(E) = 4.5x10^6 \bullet E \frac{1+1.5(E/7)^4}{1+4.725(E/7)^5} \qquad (12)$$

where E is in the unit of kV/cm, may also be used for GaAs. Values of diffusivity both parallel and perpendicular to the electric field are available [16,17]. A modified Einstein relation can again be used to extrapolate high-field diffusivities to the low-field region.

For InP, experimental [18,19] and theoretical [13] velocity-field curves of electrons are available. A useful analytical form is:

$$v(E) = 2.8x10^6 \bullet E \frac{1+0.327(E/12)^{2.7}}{1+(E/12)^{3.7}} \qquad (13)$$

where E is in units of kV/cm. Some diffusivity values have also been presented [19].

## 3. SPECIAL PROBLEMS IN BIPOLAR SIMULATION

In a bipolar transistor the flow of both electrons and holes must be considered. While the nature of bipolar device fabrication and design has permitted solutions in only one dimension with reasonable accuracy in the past, VLSI bipolar transistor behavior and that of parasitic bipolar devices as in, e.g., CMOS latchup paths, will not permit this.

SEDAN is a popular simulator that solves the 1-D equations numerically: LUSTRE is designed for use with microcomputers, so does not use a completely numerical approach: the basic equations were first manipulated analytically, subject to some simplifying assumptions, with the resulting expressions forming the basis of a simplified numerical solution.

SEDAN solves the basic set of equations straightforwardly for both carrier types. The Einstein relationship is used to express diffusion constant in terms of mobility, thus precluding the proper description of electron/hole transport in high-

field regions, such as depletion layers. The boundary conditions at the emitter, base, and collector contacts for this bipolar simulator differ, of course, from those for MOSFET simulators: charge neutrality is taken at the e,c contacts, and Boltzmann statistics are used so that the potential and carrier concentrations at these contacts can be calculated from the collector-emitter voltage and the net impurity concentrations. The majority carrier quasi-Fermi level at the base contact is required to be equal to the base-emitter voltage, so the minority carrier density at this contact can be expressed in terms of applied potential. In addition to special mobility models, SEDAN also incorporates expressions for carrier recombination and bandgap narrowing.

## 4. SOLUTION METHODS

### 4.1 Finite Elements vs. Finite Differences

The set of continuum equations that describe the physics of a problem may be transformed into a set of discrete algebraic equations for computer solution using either finite element (FE) or finite difference (FD) methods. Each method may be advantageous in certain applications.

Finite difference methods result in solutions for the physically continuous variables as specific values at the discrete points which make up a mesh. Continuum derivatives are replaced with standard finite difference approximations, with the result that the equations are discretized. The discrete variables are the values of the solutions at the mesh points. This method is conceptually simple since the finite difference derivative approximations are ready available and easy to understand, being based on the definition of the derivative itself in the limit of small mesh spacings. A price is paid for this simplicity, however: the resulting FD program can be very difficult to generalize since the difference relations become very unwieldy if they are applied to non-rectangular meshes. Local adjustment of the mesh spacing results in complicating the finite difference approximations, and implementation of higher-order (e.g., cubic or exponential) differences, necessary to reduce the number of mesh points needed, can be cumbersome. FD methods are thus characterized by simplicity obtained at the expense of generality.

The finite element method is based on an entirely different concept. First, the geometric region representing the device is divided into small subregions known as finite elements. Next, the continuum functions representing the solution are taken to have some simple form within each element, typically resulting in a global solution that is piecewise linear, cubic, etc. Finally, the original continuous physical equations are used to define an equivalent integral formulation. As a result, the problem is reduced to either minimizing a

functional over a trial space of functions, or forcing a residual function to be orthogonal to a trial space. The former problem is obtained if the original equations resulted from a variational principle; the latter problem is typically obtained where the original equations are differential equations and no equivalent variational form is known. The trial space is the space of functions which possess the assumed piecewise characteristics of the solution, which has a finite number of degrees of freedom. Associating discrete variables with these degrees of freedom yields a discrete system of algebraic equations suitable for computer solution.

While the basis of the FE method is more complicated than that of the FD method, once the necessary software is developed the generality is much greater. Since the shapes of the finite elements are arbitrary, FE algorithms work well for various arbitrary mesh structures, permitting great flexibility in the definition of device geometries, mesh refinement for improved computational efficiency, etc. Implementation of general boundary conditions is also straightforward; boundary condition restraints are often simply built into the definition of the trial function space, and the stability of the solution algorithm is often improved by the integral approach. Higher-order methods require only a modest increase in complexity, since only the piecewise characteristics of the solution are changed. However, the FE method achieves its generality at the expense of an increase in complexity.

The use of finite differences is appropriate for simple (1-D) problems and 2D programs intended to simulate specific geometries, while finite elements should be useful for programs that must be flexible in dealing with complex, particularly non-rectangular, systems.

Hybrid FE/FD approaches have been proposed [7], but seem to have inherited the complexity of the FE method without yielding any advantages.

### 4.2 Packaged Numerical Analysis Software

Powerful software packages now available for the numerical solution of systems of algebraic equations can permit device physicists to spend more time improving the physics of models than developing numerical solution algorithms. LINPACK, a package for the solution of (small) linear systems, and MINPACK, a corresponding package for nonlinear systems, are useful for 1-D simulations. In addition, EISPACK, a package for eigenvalue problems, is potentially applicable to special 1-D quantum problems. SPARSPAK, a package for large sparse linear systems, is especially useful for simulations arising from either FE or FD methods. Finally, MATLAB, an interactive matrix package, is useful for rapid testing of proposed solution algorithms. LINPACK, EISPACK, MATLAB, and MINPACK are

available from the Argonne National Laboratory (USA); SPARSPAK is available from the University of Waterloo (CANADA) Dept. of Computer Science.

These packages, written in standard FORTRAN, solve linear or nonlinear systems of algebraic equations. The user can start with the system of coupled partial differential equations above. Using either FD or FE methods, these equations are converted into discrete algebraic equations which can then be solved by the matrix packages. The discrete variables used might be potential, carrier concentration, etc. at each mesh point, corresponding to continuum potential, etc. as a function of space and/or time. Space and, for transient simulations, time must be discretized. Nonlinear discrete algebraic equations can be linearized with a Newton-like method, and solved iteratively using the linear equation solvers listed above. For small problems (e.g., 1-D simulations) MINPACK represents an alternative to this strategy. The packages mentioned here are only a representative sample of those now available.

## 5. SPECIAL PROBLEMS IN SUBMICRON DEVICE SIMULATION

### 5.2. Transport on a Small ($<1\mu$) Scale

The concept of equilibrium velocity-field curves is meaningless if the device is small enough that a carrier can pass through it in a time sufficient for only a few energy-exchange processes ("collisions") to occur. The carrier velocity may then exceed, on a transient basis, its steady-state value. This transient carrier transport (TCT), may lead to extremely fast devices.

The major TCT phenomenon is velocity overshoot, i.e., the transient rise of a carrier's velocity to a value above that predicted by the velocity field curve. Overshoot is a complex phenomenon related to the quality of the collisions as much as to their quantity; in GaAs and InP, overshoot arises from the different effects that collisions have on energy and on momentum, and the time required for both of these quantities to simultaneously reach values proper for intervalley transfer to occur within the conduction band. In Si, overshoot arises because scattering rates are inversely proportional to energy [13]. Transient effects in Si which are collision-dominated damp out very fast due to the strength of its acoustic phonon and equivalent intervalley scattering processes; in GaAs, however, peak velocity can reach almost $7\times10^7$ cm/sec and TCT can last over distances up to a half-micron or so.

TCT calculations for electrons in Si and GaAs at 300 degrees K are summarized in Figure 1, which may be used to determine what types of devices should be designed on the basis of TCT. If a device includes a region in which a particular combination of field and distance over which that field occurs

lies underneath the relevant curve, then the device should be affected by TCT. For example, the average electric field in a GaAs MESFET with a 6V drain-gate potential difference and 1.0 micron drain-gate space is about 60kV/cm, which lies above the GaAs curve; the distance is, in the device, so large compared to the duration of the transient at this field that TCT is of little importance. If $V_{DG}$ is reduced to 1V, however, TCT can affect electron velocity over the whole drain-gate space.

## 5.2 Device Simulation Incorporating TCT

Devices in which transient carrier transport occurs may be simulated using two-dimensional Monte Carlo methods. Even when relatively large approximations are made, however, these techniques are extremely costly. Further, many of the important materials parameters which must be used in this calculation are only approximately known, and hardly seem to justify the use of such a costly technique. Methods which reduce the description of the carrier dynamics to the "hydrodynamic" equations, which can be derived from the Boltzmann equation, have been used for the TCT problem. However, most who have used this approach have been forced to make rather severe approximations.

Alternatively, a semiclassical model based on the Boltzmann transport equation has been shown to be useful for two-dimensional simulations in which electron energy, rather than, say, velocity, is determined over the volume of the device. Local current flow can then be derived from that energy [20]. This method was originally used to analyze GaAs and Si MESFETs, in which carrier flow could be considered one-dimensional over the region where the electric field was greatest. Figure 2 is an example of the differences in behavior of a short-channel device as revealed by a traditional simulation and by a simulation that can incorporate transient carrier transport, A new finite-element MOSFET simulation program utilizing this technique, and in which the one-dimensional approximation must necessarily be removed, is now being developed as a logical extension of the original MESFET studies.

The direct energy-calculation method greatly facilitates the calculation of electron temperature, a quantity which can be used to gain insight into device noise problems, and also into mechanisms of failure related to electron heating. Useful results can be obtained even if electron temperature calculation is decoupled from the electric field calculation: electric field profiles can be obtained using a simple two-dimensional device simulator such as MINIMOS or GEMINI, and an approximate solution of the BTE in terms of electron energy can then be obtained using this field. The BTE for the energy of electrons in silicon can be written in the following one-dimensional form provided the heating is primarily caused by a one-dimensional field in the direction of transport [20]:

$$\frac{dw}{dx} = \frac{3}{5}\left(eE_x(x) - \frac{w-w_o}{v_x(x)\tau_w}\right) \tag{14}$$

where $\tau_w$ is the averaged electron energy relaxation time. For more exact solutions this one-dimensional approximation should be removed. Electron temperature may be obtained from Equation (14) using the approximation

$$W \triangleq \frac{3}{2} kT_e \tag{15}$$

To determine the electron energy distribution along the length of a device, the longitudinal electric field must be determined, and $T_e(x)$ can then be evaluated using Simpson's rule.

We have utilized the method outlined above to determine electron temperature profiles in normal and lightly doped drain (LDD) MOSFET structures. These results can be used to gain insight into the design of devices immune to provide an indication of the hot-electron gate current injection and other hot electron effects. Electric field profiles were obtained for these calculations from GEMINI. Figure 3 shows the longitudinal electric field and corresponding electron temperature profiles for both a standard MOSFET and for an LDD device [21]. The reduction in the electric field in the LDD device and associated reduction in $T_e$ are apparent.

## 6. NEW MATERIALS and NEW DEVICE STRUCTURES

Our understanding of the physics of such materials as InP or GaInAs, and our capabilities for simulating such device structures as heterojunction FETs and bipolar devices, trench memory cells or even just very-short-channel MOSFETs, are being pressed to keep up with the technologists' abilities to make the devices themselves. Device simulation must, then, proceed in new directions in order to remain useful in the prevention (and analysis) of expensive fabrication mistakes.

### 6.1 Materials Other than Si or GaAs

The accuracy of simulations of devices made of such materials as InP, GaInAs, AlGaAs, etc. will be limited by that of the materials parameters used. Some materials data even for silicon, particularly that which describes inversion-layer transport, is not yet fully understood. However, enough has been measured or calculated for InP and some variations of GaInAs [22] to permit either traditional or more modern types of simulations to be performed for devices made of these materials.

## 6.2 New Device Structures

Some programs--such as IBM's finite-element FIELDAY--have no problems in dealing with interesting geometries in inhomogeneous media. However, the finite-difference programs that are available are constructed to deal with particular geometries, such as planar MOSFETs, and are not so easily modified. If, however, the geometry of interest has a planar free surface and represents only a modification of symmetry, some finite-difference programs may be adapted by manipulating doping profiles.

An example of the usefulness of such an approach is shown in Figure 4, which shows a set of equipotentials obtained for a trench memory cell from a GEMINI which has been modified to utilize quite flexible junction profiles [23]. Such simulations were used to determine maximum permissible bias levels, subthreshold performance, amount of charge stored, and a measure of hot electron gate currents as functions of dimensions and substrate doping, as part of the design of this cell.

Heterostructure devices (HEMT, TEGFET, MODFET, etc.) should be treatable as logical extensions of currently familiar devices. The correspondence between MOS and modulation-doped structures has already been pointed out [24] and the correspondence between governing equations of these devices is clear [25]. However, the dimensions of the regions in which carriers are confined in these devices, as in modern Si MOSFETs, are small enough that quantum effects should be properly considered [26]. Confinement of carriers in very small regions leads to quantization of allowed energy levels, creating sub-bands within the valence and conduction bands. The attendant problems of band population distribution and inter-band scattering have been a subject of discussion for some time, but little is yet understood of the extend to their importance in submicron MODFET or MOSFET simulation.

## 6.3 3D Structures: Circuit Simulation Approach

Simulation of three-dimensional structures using standard techniques can consume considerable CPU time. For some problems, such as those of CMOS latchup simulations and some bipolar logic circuits, a cheaper alternative is desirable. 3D simulation may be performed with an acceptable level of accuracy by breaking the three-dimensional region of interest into a three-dimensional network of one-dimensional devices. The active elements may be interconnected with current-modulated resistances. Circuit simulation programs may then be used to determine the overall performance of the three-dimensional circuit. [29,30]. This method may well prove very useful for the approximate determination of properties of large three-dimensional systems.

## 7. SUMMARY

Semiconductor device simulation techniques using materials properties that are taken as decoupled from the action of charge carriers within the device itself can only with difficulty, and the use of nonphysical approximations, be used to simulate the behavior of submicron-scale devices. At this scale, transient transport and quantum effects must be taken into consideration. Energy-transport techniques may be more efficient for such device simulation than any other method at this time. Large three-dimensional problems may be efficiently simulated as three-dimensional networks of one-dimensional devices, but the limits to the accuracy of this approach in specific problems have not yet been established.

Acknowledgements

Robert Byrnes and Carol J. Shedlock made significant contributions to Sections 3 and 4 above.

## REFERENCES

1. W. Shockley, Proc. IRE 40, 1365 (1952)

2. R. H. Dawson and J. Frey, Solid State Electron. 22, 343 (1979)

3. J. A. Greenfield and R.W. Dutton, IEEE Trans. ED-27, 1520 (1980)

4. S. Selberherr, et al., IEEE Trans. ED-27, 1540 (1980)

5. T. Wada and J. Frey, IEEE Trans. ED-26, 476 (1979)

6. T. Toyabe and S. Asai, IEEE Trans. ED-26, 453 (1979)

7. E.M. Buturla, et al., IBM J. Res. Develop. 25, 218 (1981).

8. D.C.D'Avanzo, et al., "One-Dimensional Semiconductor Device Analysis (SEDAN); Stanford University, Integrated Circuits Laboratory Report G-201-5, (1979)

9. C. Shedlock, "LUSTRE: A Fast Bipolar Process and Device Simulator using Analytic Methods", M.Sc. Thesis, Cornell University, 1984

10. P. Butcher, Rep. Prog. Phys. 30, Part I, 97 (1967)

11. K. Yamaguchi, et al., Japan J. Appl. Phys. 14, 1069 (1975)

12. G. Persky and D. Bartelink, J. Appl. Phys. 42, 4414 (1971)

13. T.J.Maloney, "Nonequilibrium Electron Transport in Compound Semiconductors", Ph.D. Thesis, Cornell University, 1976

14. P. Smith, et al., Appl. Phys. Lett. 37, 797 (1980)

15. T. Wada, "Effects of Hot-Carrier Phenomena on Performance of MESFETs", M.Sc. Thesis, Cornell University, 1979

16. W. Fawcett and H.D.Rees, Phys. Lett. 29A, 578 (1969)

17. P.E. Bauhahn, et al., El. Lett. 9, 460 (1973)

18. P.M. Boers, El. Lett. 7, 625 (1971)

19. T.H. Hill, et al., Appl. Phys. Lett. 42, 725 (1983)

20. R.K. Cook and J. Frey, COMPEL 1, 65 (1982)

21. N. Goldsman and J. Frey, to be published

22. T.P. Pearsall, ed., "GaInAsP Alloy Semiconductors", J. Wiley and Sons, New York (1982)

23. T. Furuyama and J. Frey, to be published

24. P.M. Solomon, Proc. IEEE 70, 489 (1982)

25. R.F. Pierret and M.S. Lundstrom, IEEE Trans. ED-31, 383 (1984)

26. T. Ando, et al., Rev. Modern Phys. 54, 437 (1982

27. K. Hanihara, et al., IEEE Trans. ED-29, 418 (1982)

28. C. Werner, et al., Compel 2, 195 (1983)

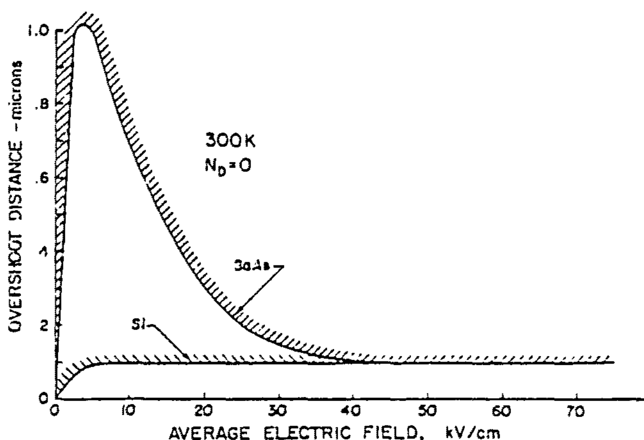


FIGURE 1

Regions of operation of Si and GaAs devices in which velocity overshoot may be an important factor in determining device performance. In each case the regions below the curves are those in which overshoot should be considered. For example, overshoot should be incorporated into the design analysis of a GaAs "active length" (channel, perhaps) is e.g. 0.6 micron, and the average electric field in the channel is 5 kV/cm; overshoot would probably have negligible effect in such a device if the average field were greater than about twelve kV/cm.

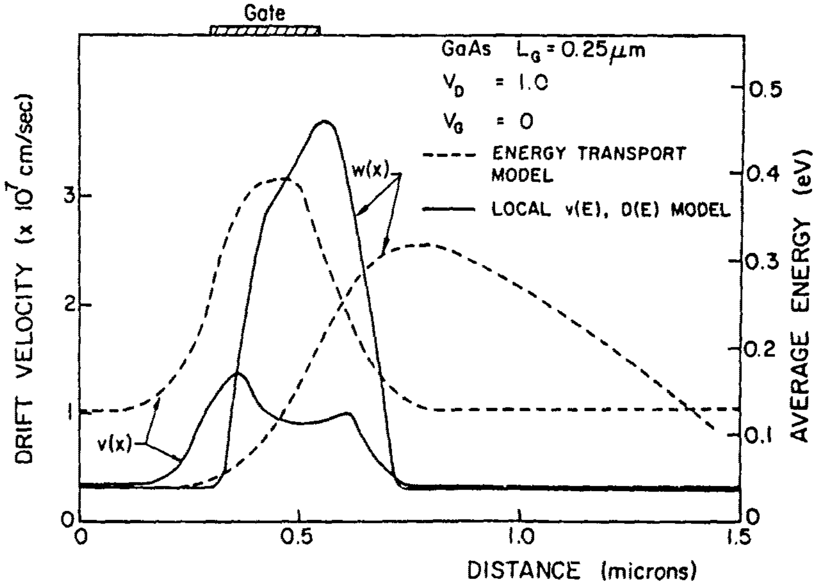


FIGURE 2

Average electron velocity and energy as functions of distance along the channel of a 0.25 micron gate GaAs MESFET, using both an energy-transport and a field-dependent (localized) transport model. Channel thickness = 0.12 microns, doping = 1.5 X $10^{17}$, drain voltage = 2.0V, applied gate voltage = 0V, built-in gate voltage = -0.6V.

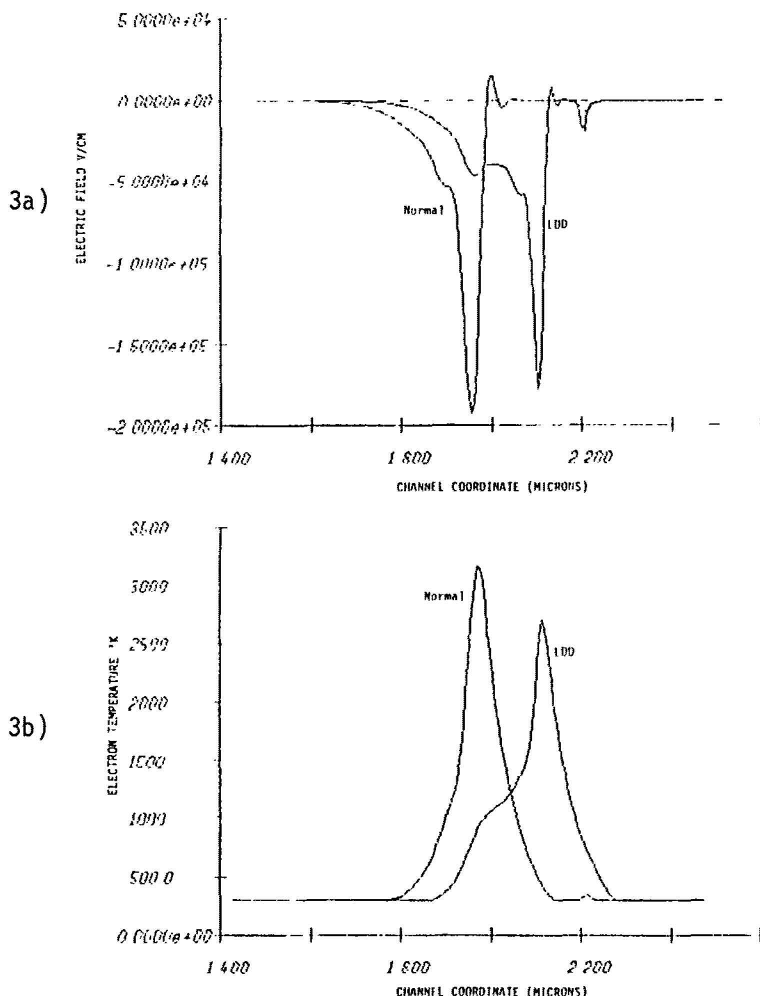


FIGURE 3

Electric field(a) and electron temperature(b) profiles for a standard MOSFET and an LDD device. The parameters of these devices are:

Standard: Substrate doping $3x10^{15}/cm^3$
n+ source and drain doping $3x10^{18}/cm^3$
Drawn channel length = 1.4 microns
Effective channel length = 0.9 microns, extending to distance coordinate 1.95 microns
S/D junction depth = 0.25 microns
Oxide thickness = 15 nm.

LDD: n- low-field region doping $3x10^{16}/cm^3$, extending from distance coordinate 1.95 microns to 2.1 microns.

In both devices only the region near the drain junction is shown, and the drawn gate metal ends at coordinate 2.2 microns.

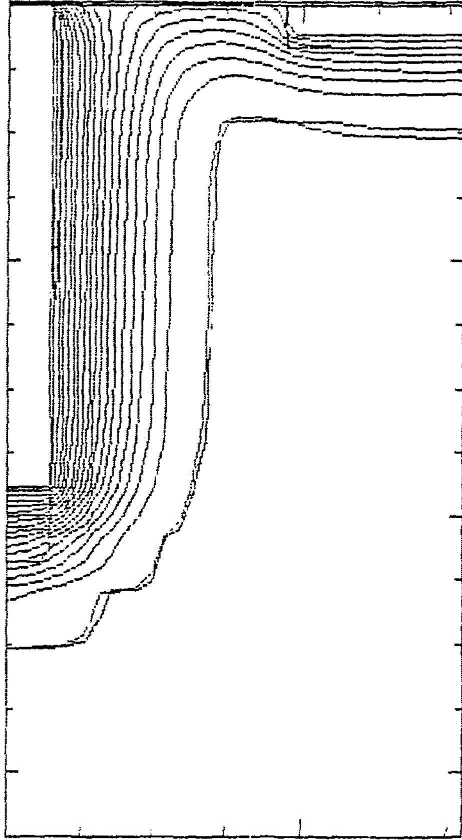

FIGURE 4

Typical equipotentials in a sample trench-capacitor memory cell under development for memories at the 16mB/chip scale. An etched trench at the left contains one capacitor plate, made of polysilicon; the other plate is the vertical n+ diffusion. The drain of the access transistor is the horizontal n+ diffusion at the upper right. Each small division = 0.5 micron.