SIMULATION OF SEMICONDUCTOR DEVICES AND PROCESSES Vol. 4
Edited by W. Fichtner, D. Aemmer - Zurich (Switzerland) September 12-14, 1991 - Hartung-Gorre

Three-Dimensional Adaptive Semiconductor Device Simulation*

J.V. Ashby, C.J. Fitzsimons, R.F. Fowler, and C. Greenough
Rutherford Appleton Laboratory, Chilton, DIDCOT, Oxon OX11 0QX, UK

P.J. Mole
BNR Europe Limited, Harlow, Essex, UK

W.H.A. Schilders
Philips Research Laboratories, Eindhoven, The Netherlands

June 18, 1991

Abstract

We present results from a three-dimensional device simulator, using adaptive meshing and solving the drift-diffusion equations. The adaption algorithm and the criteria used for adaption are discussed. Three devices of industrial interest are presented: a bipolar transistor, a $1.25\mu m$ n-MOS device and a CCD, illustrating the range of devices which may be successfully simulated.

1 Introduction

The three-dimensional analysis of semiconductor devices is invaluable for the understanding of some aspects of modern device technologies. The high cost of performing this analysis requires new algorithms which enable the economic solution of the device equations. While it is possible for the experienced designer to generate meshes manually to capture device features accurately for one-dimensional and two-dimensional problems, it is much more difficult in three dimensions. This is due to factors such as the much larger number of nodes and difficulties in visualising the mesh. Since the number of nodes used in the discretisation is an important factor in determining the solution time, it is desirable to minimise this. On the other hand too few nodes will give an inaccurate solution and a balance has to be sought. The solution of the device equations exhibits fine-scale features which depend on the geometry, physics and operating conditions of the device. In this paper we describe a method which automatically adapts a very coarse mesh to these features of the solution.

We solve the three drift-diffusion equations [1] using a fully-coupled Newton process in which the equations are discretised by a Voronoi-based finite volume technique using the Scharfetter-Gummel expression [1] to approximate the flux term in the continuity equations. The resulting linear systems are solved using preconditioned conjugate gradient methods. We use the electrostatic and quasi-fermi potentials ψ, ϕ_n and ϕ_p as the solution variables for the linear systems

*This research has been partially supported by the European Community under ESPRIT Project 962 ("EVEREST: Three-Dimensional Algorithms for a Robust and Efficient Semiconductor Simulator with Parameter Extraction") and by the UK government under the EASE+ programme.

and apply the correction transformation [2] to make the updates to ψ and the charge densities, models is available including bandgap narrowing, doping, temperature and field dependent mobilities and both Shockley-Read-Hall and Auger recombination.

The meshing strategy involves the use of both hexahedra and tetrahedra which satisfy the Delaunay criterion. Hexahedra are used where possible, the tetrahedra are used to mesh geometrically irregular portions of the domain and to provide a transition between hexahedra of different size. The mesh is locally structured, a feature we exploit in the adaption process.

2 Adaption Strategy

Our analysis starts from a geometric description of the device which is built up from points, lines, surfaces and volumes. This description is passed to an initial mesh generator which produces a semi-regular lattice of nodes (i.e. the lattice spacing is irregular in each direction but the lattice cells are right hexahedra) in the device by projecting planes from each geometric point perpendicular to the cartesian axes. Additional planes can be inserted in each direction at the user's discretion. The intersection of planes with other surfaces in the device generates additional planes and so on until a complete lattice which contains the device is formed.

Deleting lattice points which lie outside the device gives a set of bricks with between 1 and 8 nodes which is then turned into a mixed mesh of hexahedra and tetrahedra, the latter being used to mesh geometrically irregular parts of the device. The mesh produced at this stage may consist of only a few hundred nodes, even for a quite complex device. The software has been written to assemble the elements efficiently. The use of hexahedra where possible gives the immediate advantage that the Voronoi volumes associated with a node and pipe areas associated with an edge connecting two nodes are easily computed while the added flexibility of tetrahedra is important in limiting the size of the mesh.

This crude mesh is then transferred to the device modeller where refinement takes place based first on doping and then on the solution to the device equations. At this stage tetrahedra are also used to provide a transition between hexahedra of different sizes. Nodes are inserted at the midpoints of edges as indicated by the refinement criteria. The refinement is based on the hexahedra; to decide whether or not to divide a hexahedron in two, we look first at the refinement of each edge. No hexahedron in the final mesh may contain nodes at quarter-points of its edges, and no hexahedron may have more than ten nodes. We restrict the types of nine and ten-noded hexahedra allowed; a ten-noded hexahedron must have its two additional nodes on adjacent parallel edges. A hexahedron is further divided if the angle subtended at an additional node by an opposite side is greater than $\pi/2$ since this violates the Delaunay condition. This means that a mesh line (plane) introduced by refinement is continued until it reaches a suitable hexahedron at which it can terminate or the boundary of the device. Each hexahedron produced by a division is subsequently tested to see if it should in turn be divided.

The user specifies the required solution accuracy, the maximum number of nodes in the mesh due to the doping refinement criterion (N_D) and the overall maximum number of nodes in the mesh (N_ψ). The overall algorithm, where N_M is the number of nodes in the mesh, is:

1. Generate the initial, coarse mesh and assign doping values.
2. **while** $N_M < N_D$ **and** *doping accuracy is not achieved* **call** REFINE(Doping).
3. Solve the problem on the resulting mesh.
4. **while** $N_M < N_\psi$ **and** *accuracy is not achieved*

begin

call REFINE(Potential).

Solve the problem on the resulting mesh.

end

5. Output the final mesh and the solution.

If N_D or N_ψ are exceeded before the relevant criteria are satisfied the user is warned of this fact and the mesh from the last completed refinement phase is used.

The data-structures used in the refinement are based on the lattice of hexahedra used to generate the initial and refined meshes. These are of three types: those with fewer than eight vertices inside the device, those with eight, and those with nine or ten, which we call transitional hexahedra. A hexahedron is described by its twelve defining edges in a systematic way. An edge is defined by its endpoints. A list of hexahedra in the current lattice which meet at a node is maintained to enable rapid checking of neighbours. The parent-child relations which arise when edges or hexahedra are divided in two are also stored, e.g when an edge is divided the two new edges are stored as children of the old edge and the old edge is stored as the parent of the new. This enables efficient access to the mesh data needed for refinement. This approach yields a hierarchical mesh structure which allows the recovery of the mesh at a previous level of refinement. This could be used either for multigrid solution or to restart at an interim mesh for subsequent refinement at other bias points or time-steps.

Adaption using the doping criterion enriches the mesh in regions of geometric and intrinsic physical interest and it is very economical since it does not require the solution of any linear systems. All that is necessary is a means of computing the doping profile at the new nodes. Adaption using the potential criterion enriches the mesh in regions where this is required by the operating conditions. It is important for efficiency to choose an initial guess for the Newton process which will give good convergence. We do this by first calculating the doping at the new nodes and then interpolating the solution (ψ, ϕ_n, ϕ_p) on each edge to which a new node is added. Linear interpolation is used for the quasi-fermi potentials. The potential ψ is set to the lower of its two neighbouring values for n-type nodes and to the higher for p-type. Then the Poisson equation is solved *only* at the nodes generated by this refinement sweep, while holding the values at existing nodes fixed. This is followed by a similar solution of the coupled system which provides a good initial guess for the subsequent solution on the full mesh.

3 Refinement Criteria

Several authors [3], [4], [5] have developed *a posteriori* adaption strategies for the device problem. In [6] the equivalence of the box method and the finite element method using piecewise linear elements is shown. Polak *et al* [2] showed that the box method may be viewed as a Petrov-Galerkin method. These results have led us to employ *a posteriori* error estimation techniques in the case of box discretisations. In [7] an error estimator is developed based on the weak formulation of an elliptic problem. Due to the problems associated with interpolation in device modelling, this must be modified slightly. This was first done in [5] in two dimensions and extended to three dimensions in [8]. We consider refinement based on the electric field explicitly; similar analysis holds for the current density. Let ψ^h denote the approximate solution to the Poisson equation and let e be the error in the solution. Defining $(a, b)_{\omega_i} = \int_{\omega_i} ab \, d\omega_i$, where ω_i is a mesh element with boundary $\partial\omega_i$. Replacing ψ by $\psi^h + e$, and applying the divergence

theorem to the weak form of the Poisson Equation, we obtain the following expression for E, an approximation to the error, e.

$$(\varepsilon\nabla E, \nabla v)_{\omega_i} + (q(n+p)E, v)_{\omega_i} = -(\varepsilon\nabla\psi^h, \nabla v)_{\omega_i} + (q(n-p-D), v)_{\omega_i} + (\varepsilon\psi_n^h, v)_{\partial\omega_i} \quad (1)$$

where the test function $v \in H_0^1(\Omega)$ and ψ_n^h is the normal derivative of ψ^h on the boundary of ω_i. n and p are the electron and hole densities respectively, D is the doping and ε is the electric permittivity. The last term dominates for odd-order elements [7] [5]. We approximate it by taking the average value of ψ_n^h on both hexahedra sharing a common face and using these values to evaluate the jump term $F^+ - F^-$ where

$$F^{\pm} = \frac{\int_{\partial\omega_i^{\pm}} \varepsilon\psi_n^h ds}{\int_{\partial\omega_i^{\pm}} ds} \quad (2)$$

Because the jump term dominates the error it suffices to monitor this quantity. Therefore we choose to refine an element if

$$\mid F^+ - F^- \mid \frac{h}{2\varepsilon} > \mathrm{TOL} \quad (3)$$

where ε is the permittivity of the hexahedron and h its length in the direction of refinement. This criterion is effectively measuring the difference between ψ on the boundary between the two hexahedra and the value given by linear interpolation between the two outer faces. TOL is measured in thermal volts, typically being given a value of 1. Other more sophisticated criteria have been suggested but experience has shown that they are insufficiently robust. In [9]

$$\frac{\mid F^+ - F^- \mid}{F_0 + \delta_r \max(\mid F^+ \mid, \mid F^- \mid)} > \delta_a \quad (4)$$

was suggested where F_0 was a cut-off field below which refinement was not considered necessary. δ_r and δ_a are user supplied parameters which control the relative effect of F_0 and the overall degree of refinement respectively. The difficulty with this approach is that the optimal choice of F_0 and δ_r depend on the bias conditions which can readily change the fields in a device by two orders of magnitude. In an effort to overcome this the same authors suggested [10] replacing F_0 by $F_{\max}$, the maximum of the displacement fields calculated over all the hexahedra in the device. Again, this can lead to problems, for instance in a device with two junctions, one of which has a high built-in field and the other a low one. It is possible that the second junction will be inadequately resolved with this criterion. It is also possible that a high field in one direction will mask weaker fields in another which are more important for device operation. For these reasons we feel that the criterion in (3) is the most practical way of monitoring the dominant term in the error estimator.

Having checked one of its neighbours and decided that a hexahedron needs refinement, each of its edges is divided perpendicular to the common face. Then the neighbours in the remaining directions are checked.

The doping criterion ensures the subdivision of each edge in the mesh for which the following is true

$$\frac{\mid D_i - D_j \mid l_{ij}}{\max(\mid D_i \mid, \mid D_j \mid, \mid D_{\mathrm{ref}} \mid)} > l_{\min} \quad (5)$$

where D_i, D_j denote the doping at the endpoints of the edge, l_{ij} is its length, D_{ref} is a user-specified reference doping, and $l_{\min}$ is a user-specified minimum edge length for the mesh.

4 Examples

We illustrate the usefulness of this approach with results from our three-dimensional device modeller, EVEREST [11], on three typical devices. The first is a polysilicon emitter transistor with a fully three-dimensional structure around the emitter. Figure 1 shows the basic geometry of the device with the base and raised emitter contacts. The collector contact is on the back face simulating a buried collector layer. This is a device for which experimental results and two-dimensional simulations are available for comparison. SIMS profiles were used to define the doping for emitter, base and collector regions. Two stages of refinement are carried out, first on doping to ensure that features above $10^{14}cm^{-3}$ are resolved to within $0.033\mu m$ followed by a further three levels of refinement on potential with a collector bias of $2V$. The final mesh has 12163 nodes. This mesh is then used in all subsequent simulations. The predicted currents and Early voltage have been compared with measured data and with the two-dimensional simulator HFIELDS [13]. The same physical models were used in both simulations so that any differences are attributable to the methods used to solve the device equations. The two simulations agree well but the base currents are over-estimated compared with measurements. However, accuracy is consistent with two-dimensional simulators. Figure 1 shows the logarithm of the net doping across the base region. The doping gradient has been well captured by the mesh used. Note that the isometric projection software is based on triangular elements so that diagonals have been inserted on quadrilaterals.

In Figure 2 we show the electrostatic potential, ψ, on the plane $z = 0\mu m$. This shows how well the depletion regions around the base-emitter and base-collector junctions have been resolved.

The second device is based on a $1.25\mu m$ gate length n-MOS structure. The doping profiles were obtained from SUPREM [12] simulations with lateral diffusion of the source and drain implants approximated by a lateral error function with $\sigma_\perp = 0.04\mu m$. In this device we are looking for refinement of three features: the source and drain junctions, the channel under the gate and the spreading of the drain depletion region at high source-drain bias. The field oxide is approximated by a step as the isolation edge effects are not investigated. A schematic of the geometry in which the y-coordinate has been shrunk by a factor of 0.1 is shown in Figure 3. In this case the refinement is performed in three stages. Following refinement on doping to locate the source and drain junctions, $1.45V$ is placed on the gate so that refinement on potential captures the inversion layer. Then the gate is turned off and the drain biased at $2V$ to perform extra refinement around the drain depletion region. The final mesh contains 22731 nodes. Again, reasonable agreement with experiment is found, particularly given the uncertainties in the exact channel length and mobilities for the measured device. In particular the threshold voltage agrees well, although the drain current is greater than measured. This discrepancy could be due to an error in the channel length or in the mobility value used, or it could be due to not refining on current. However, the correct prediction of threshold indicates that the adaptive refinement has adequately described the channel doping and the inversion layer location.

In Figure 3 we show the customary MOS cross section on the end plane of the device where nearly two-dimensional behaviour is to be expected. For these simulations the mesh generated by the above technique was used but the bias conditions were $V_g = 1.95V$ and $V_d = 0.05V$. It can be seen that the source and drain regions and the channel are well resolved and a careful examination shows that there has been some extra refinement around the drain end depletion region. In Figure 4 we show a section through the centre of the gate contact. The use of a step field oxide introduces rapid curvature into the potential since it poses a Laplace problem with

a Dirichlet boundary condition on an internal corner. The refinement has coped well with this difficult problem, although it could be argued that it has gone too far in trying to resolve the corner.

Thirdly we have simulated a charge-coupled device used in imaging applications. The complicated doping profile for this device is incorporated into EVEREST by means of a Fortran subroutine. It consists of a background dope and several implantations leading to a three layer structure on one side of the device and a four-layer structure on the other. There are five gate contacts as well as contacts on an n-channel, a p-well and the substrate. In operation the two outer gates are held at $0V$ and the remaining three gates at $10V$. The substrate is biased at $15V$, the p-well at $4.5V$ and the n-channel at $10V$ reducing to $7V$.

Figure 5 shows three levels of mesh on the centre plane of this device. Top left is the initial coarse mesh of only 784 nodes where only minimal user input has been used to place nodes close to the gate contacts. Top right is the mesh following refinement on doping, which now contains 6027 nodes and is beginning to display features of the device. Then $2V$ is applied to the inner gates and the mesh refined to 9459 nodes as shown at bottom left. Now we can begin to see the gate structure emerge, particularly the high field region between the outer and inner gates. Finally the full $10V$ potential is applied to the inner gates and the mesh refined to 16609 nodes. An isometric plot of potential is shown bottom right. Some of the further refinement visible on this plot arises from the need to resolve potential features in other regions of the device and so appears not to be necessary on this plane. It is a feature of the restrictions which we place on the mesh that refinement spreads through the device until it can heal sensibly by a transition to a larger hexahedron or until it reaches the device boundary.

It is expected that the currents will increase exponentially as the p-well voltage is reduced and this is seen. In fact, the height of the potential barrier experienced by the carriers in the device can be calculated and the current shows an exponential dependence on this barrier height. Finally it is worth noting that the simulations have shown discrepancies with the one-dimensional based rules of thumb commonly used to explain the operation of such CCD-cells. Clearly the three-dimensional nature of the device is manifesting itself at a fundamental level and the use of simulators such as EVEREST will be valuable in understanding how such devices work and can be optimised.

5 Summary

We have presented an algorithm for performing mesh adaption in three-dimensional semiconductor device simulation. We have discussed the choice of refinement criterion in detail and surveyed some of the alternatives. Finally we have presented results which demonstrate the usefulness of the method when applied to a range of devices of industrial interest.

References

[1] **Selberherr S.** *Analysis and Simulation of Semiconductor Devices* Springer Verlag (1984)

[2] **Polak S.J., den Heijer C., Schilders W.H.A., Markowich P.A.** "Semiconductor Device Modelling from the Numerical Point of View", Int. J. Num. Meth. Eng. **24** 763-838 (1987)

[3] **Edwards S.P., De Meyer K., De Wilde Ph.** "Adaptive Meshing Applied To A Two-Dimensional Device Simulator" from *Fundamental Research On The Numerical Modelling Of Semiconductor Devices and Processes* (Ed. J.J.H. Miller) Boole Press, Dublin (1987)

[4] **Ciampolini P., Forghieri A., Pierantoni A., Gnudi A., Rudan M.V., Baccarani G.** "Adaptive Mesh Generation Preserving the Quality of the Initial Grid" *IEEE Trans. Comp. Aid. Dev.* **CAD–8**, No. 5, 490–500 (1989)

[5] **Philips** *Report P3.2.3* Project EVEREST, ESPRIT 962, April (1988)

[6] **Bank R.E., Rose D.J.** "Some Error Estimates for the Box Method" SIAM J. Num. Anal., **24**, 777-787 (1987)

[7] **Babuška I., Miller A.** "*A Posteriori* Error Estimates and Adaptive Techniques for the Finite Element Method" *Techn. Note BN–968* Inst. for Phys. Sc. and Techn., Lab. for Numer. Anal., University of Maryland (1981)

[8] **Fitzsimons C.J., Miller J.J.H.** "Algorithmic Advances in Three-Dimensional Semiconductor Device Modelling" presented at the Fourth Conference on Circuit Design, Dresden, 12–14 February (1990)

[9] **Philips** *Report P2.3.3* Project EVEREST, ESPRIT 962, April (1989)

[10] **Philips** *Report P3.2.4* Project EVEREST, ESPRIT 962, April (1990)

[11] **Greenough C., Fitzsimons C.J., Fowler R.F.** "Software for Modelling Semiconductor Devices in Three Dimensions" Rutherford Appleton Laboratory Report RAL-91-042 (1991)

[12] **Ho C.P., Plummer I.D., Hansen S.E. and Dutton R.W** "VLSI Process Modelling - SUPREM III" IEEE ED **30** 1438 (1983)

[13] **Baccarani G., Guerrieri R., Ciampolini P., Rudan M.V.,** "HFIELDS, a highly flexible 2-D Semiconductor-Device Analysis Program" *Proceedings of the NASECODE IV conference* (Ed. J.J.H. Miller) Boole Press, Dublin (1985)

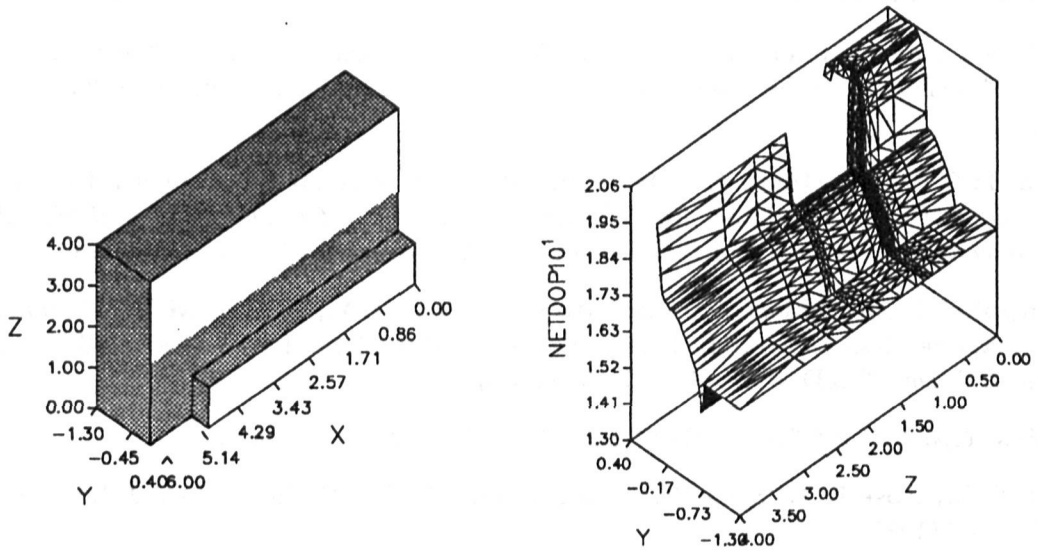

Figure 1: Geometry of the 3D *npn* bipolar (left). The white areas are the base and emitter contacts. An isometric view of the net doping ($\log|N_A - N_D|$) on the plane $x = 0\mu m$ is also shown (right), showing good resolution of the base doping.

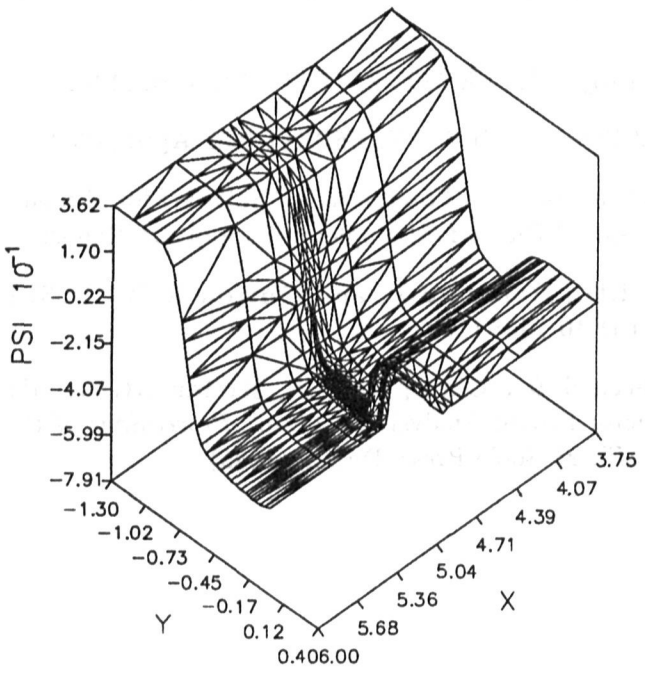

Figure 2: Potential on the plane $z = 0\mu m$ for the 12163 node mesh with $V_e = -0.8V$ and $V_c = V_b = 0V$.

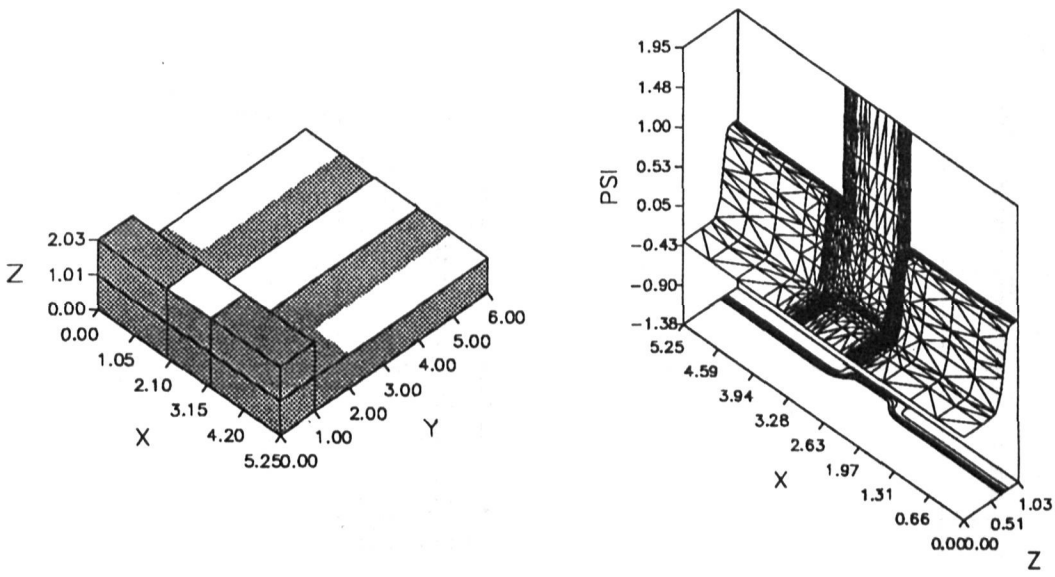

Figure 3: Geometry of the MOSFET (left, in μm, except the y-axis which is compressed by a factor of 10). The central white strip is the gate contact with source and drain to either side. The block of field oxide extends from 0 to $10\mu m$ in y. To the right is shown the potential on the face $y = 60\mu m$.

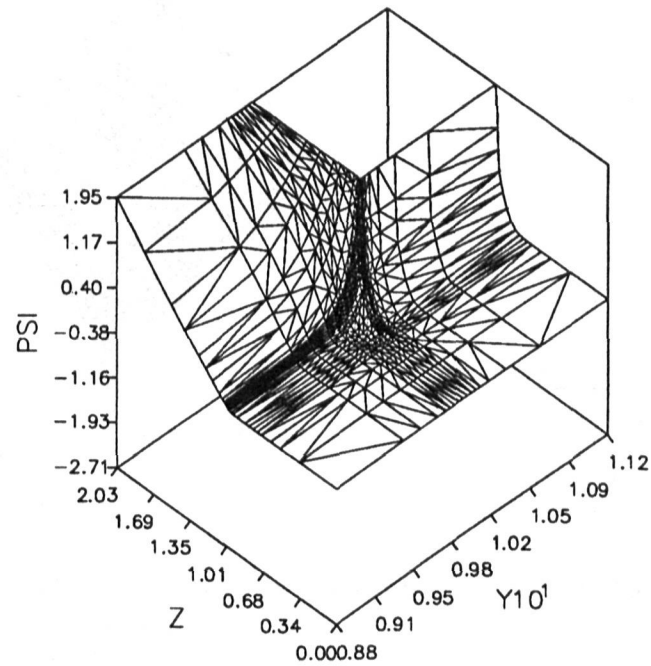

Figure 4: Potential on the plane $x = 2.625\mu m$, along the centre of the gate. Only the region around the field oxide corner is shown. Solution with $V_g = 1.95V$ and $V_d = 0.05V$.

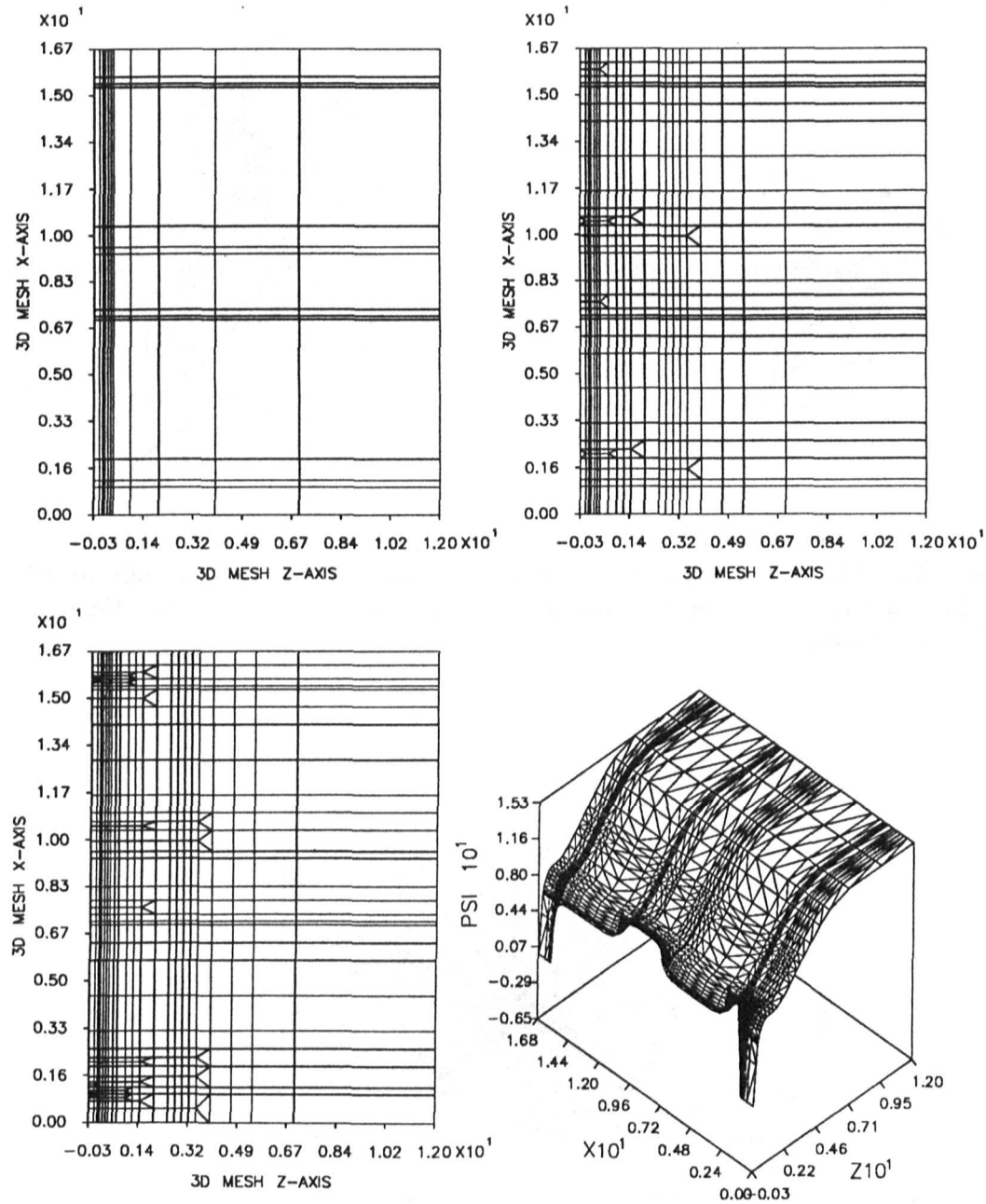

Figure 5: The mesh of the CCD device during refinement on the centre plane ($y = 1.5\mu m$) of the device. Top left: initial mesh, 784 nodes. Top right: mesh after doping refinement, 6027 nodes. Bottom left: mesh after potential refinement at low bias, 9459 nodes. Bottom right: isometric plot of potential after further refinement at high bias, 16609 nodes.