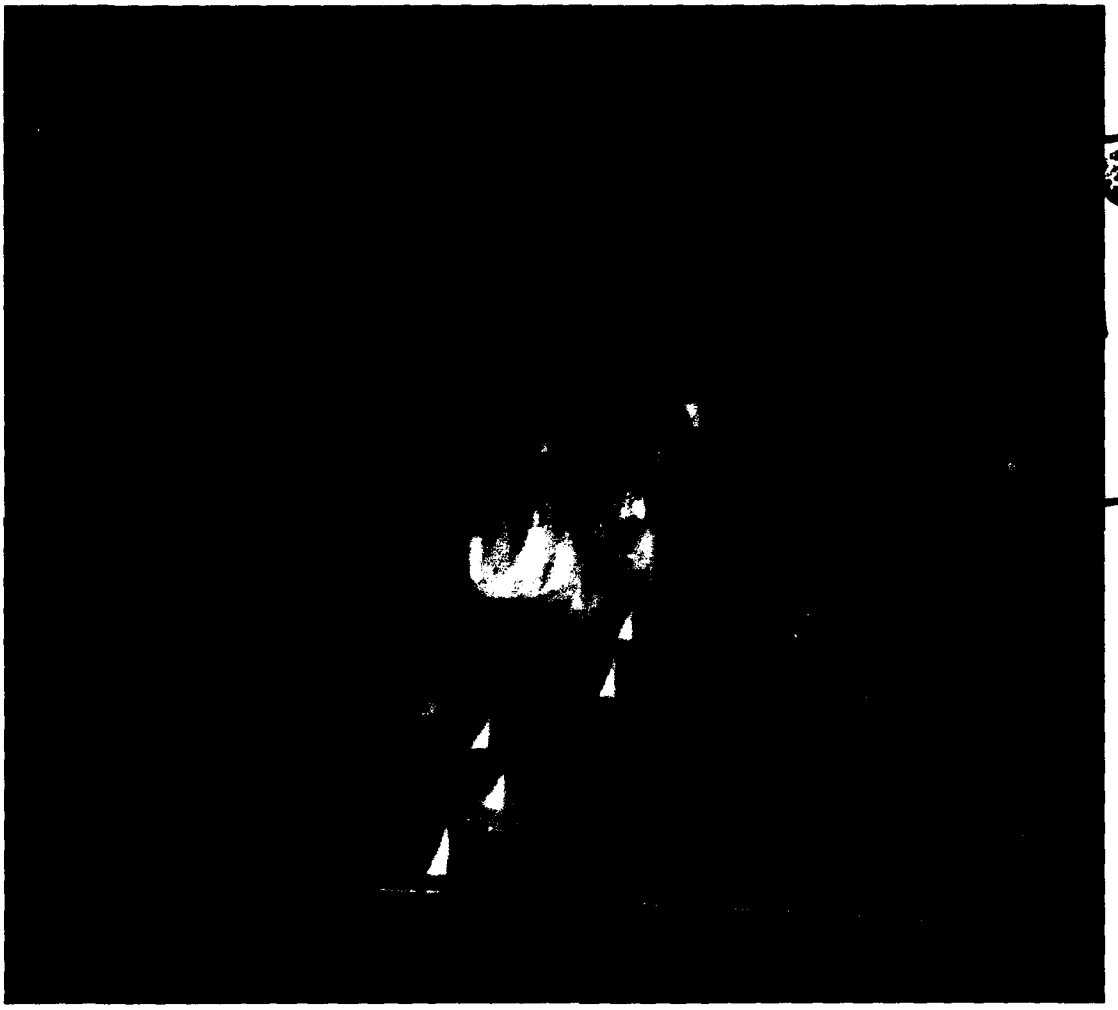
AD-A284 214

Proceedings of the Third International Workshop on Computational Electronics



DTIC ELECTE OCT 0 2 1994 U

Portland, Oregon
May 18-20, 1994

94-28454

9 4 9 01 044

This document has been approved for public release and sale; its distribution is unlimited

DTIC QUALITY INSPECTED 1

Third International Workshop On Computational Electronics

Chairman
S. Goodnick, Oregon State University

Program Committee
R. Dutton, Stanford University
D.K. Ferry, Arizona State University
M. Fischetti, IBM Yorktown Heights
C. Hamaguchi, Osaka University
K. Hess, University of Illinois
M.J. Howes, University of Leeds
M. Littlejohn, North Carolina State University
M. Lundstrom, Purdue University
M. Osman, Washington State University
W. Porod, University of Notre Dame
U. Ravaioli, University of Illinois
E. Sangiorgi University of Udine
K. Smith, AT&T Bell Laboratories
C.M. Snowden, University of Leeds

Advisory Committee
H. Bennett, NIST
F. Buot, Naval Research Laboratory
J. Cole, Rensselaer Polytechnic Institute
L. Cooper, Office of Naval Research
W.M. Coughran, AT&T Bell Laboratories
W. Fichtner, ETH, Zurich
W. Frensley, University of Texas at Dallas
M. Fukuma, NEC, Japan
H. Grubin, Scientific Research Associates
G. Iafrate, ARO
J. Jerome, Northwestern University
T. Kerkhoven, University of Illinois
S. Laux, IBM Yorktown Heights
T. McGill, California Institute of Technology
J. Meza, Sandia National Laboratories
M. Pinto, AT&T Bell Laboratories
C. Ringhofer, Arizona State University
D. Rose, Duke University
M. Schultz, Yale University
T. Sugano, University of Tokyo
A. Tasch, University of Texas, Austin
T. Thurgate, Intel
P. Vogl, Walter Schottky Institute, Munich
P. Yang, Texas Instruments
K. Yokoyama, NTT, Japan
Z. Yu, Stanford University

Sponsors
G. Lea, National Science Foundation
L. Cooper, Office of Naval Research
M. Stroscio, Army Research Officer
Oregon Center for Advanced Technology Education
Oregon Joint Graduate Schools of Engineering

Cover Illustration: Potential profile due to random impurities in a deep submicron MESFET (courtesy of Jing-Rong Zhou and David K. Ferry, Arizona State University)

Proceedings of the Third International Workshop on Computational Electronics

•

Benson Hotel
Portland, Oregon

•

May 18-20, 1994

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

Preface

The Third International Workshop on Computational Electronics was held at the Benson Hotel in downtown Portland, Oregon, on May 18, 19, and 20, 1994. The workshop was devoted to a broad range of current issues in computational electronics related to the simulation of electronic transport in semiconductors and semiconductor devices, particularly those which require large computational resources. The present workshop evolved from earlier workshops on the same theme held at the Beckman Institute in Urbana-Champaign under the auspices of the NSF National Center for Computational Electronics (NCCE). In 1992, the scope of the NCCE workshop was expanded to become an international forum for the discussion of current trends and future directions in computational electronics. Thus, the First International Workshop on Computational Electronics was held on May 28-29, 1992, at the Beckman Institute. The following year, the Second IWCE was held at the University of Leeds in the United Kingdom on August 11-13, 1993. We are grateful for support of the 1994 workshop by the National Science Foundation, the Office of Naval Research and the Army Research Office, as well as local support from the Oregon Joint Graduate Schools of Engineering and the Oregon Center for Advanced Technology Education.

There were over 100 participants in the Portland workshop, of which more than one quarter represented research groups outside of the United States including Austria, Canada, France, Germany, Italy, Japan, Switzerland, and the United Kingdom. The emphasis of the contributions reflects the interdisciplinary nature of computational electronics with researchers from the Chemistry, Computer Science, Engineering, Mathematics, and Physics communities participating in the workshop. We are very grateful to the members of the Program Committee for the selection of invited speakers and review of contributions to the workshop. We thank S. Subramanian and April Melton for their help in the workshop organization. We are also grateful to the Advisory Committee for helpful comments and suggestions. Finally, we wish to thank all the participants for the general high quality of their presentations at the workshop and of the articles contained herein.

Stephen M. Goodnick

Oregon State University, June 1993

TABLE OF CONTENTS

	Page
SECTION 1: Particle Simulation and other Boltzmann Equation Solution Methods	
A Comparison of BTE-Based Electron Transport Simulations for Silicon (Invited) J. Higman	1
Cellular Automaton Simulations of Planar Doped Barrier Field Effect Transistor in Silicon A. Rein, G. Zandler, M. Saraniti, P. Lugli and P. Vogl	7
Full Band S-Matrix Simulation of Electron Transport in Si Carl Huster and Mark Lundstrom	11
An Alternative Solution of the Boltzmann Equation: The "Scattered Packet Method J.P Nougier, L. Hlou, P. Houlet, J.C. Vaissiere and L. Varani	15
Monte Carlo Simulation of Impact Ionization Processes in Silicon (Invited) K. Taniguchi, M. Yamaji, T. Kunikiyo, Y. Kamakura, H. Takenaka, M. Morifugi, K. Sonoda and C. Hamaguchi	19
Monte Carlo Simulation of Nonequilibrium Electron-Phonon System in Quantum Wires R. Gaška, R. Mickevičius, V. Mitin and Michael A. Stroscio	25
Microscopic Simulation of Electronic Noise in Semiconductor Unipolar Structures Luca Varani, L. Reggiani, Tilmann Kuhn, Tomás González and Daniel Pardo	29
Time-Step Stability for Self-Consistent Monte Carlo Device Simulation P.W. Rambo and J. Denavit	33
Monte Carlo Studies of Hot Electron Generation in Scaled MOSFETs A. Duncan, C.H. Lee and U. Ravaioli	37
A Rejection Method for Selection of Scattered States William S. Lawson	41
New Highly Efficient Method for the Analysis of Correlation Functions Based on a Spherical Harmonics Expansion of the Boltzmann Transport Equation's Green's Function Chr. Jungemann, P. Graf, G. Zylka, R. Thoma and W.L. Engl	45
Time-Dependent Analysis of the Coupled Hot-Carrier-Hot-Phonon Boltzmann Equations L. Hlou, J.C. Vaissiere, J.P. Nougier, L. Varani, P. Houlet, L. Reggiani, M. Fadel and P. Kocevar	49
Numerical Solution of the Perturbed Boltmann Equation in Frequency and Time Domains J.C. Vaissiere, J.P. Nougier, L. Varani, P. Houlet, L. Hlou, L. Reggiani, E. Starikov and P. Shiktorov	53
Scaled Ensemble Monte Carlo Studies of Impact Ionization A.M. Kriman and R.P. Joshi	57
Using the Random-*K* Approximation to Consistently Incorporate Impact Ionization into the Generalized Expansion Method for Solving the Multi-Band BTE in Silicon Yu-Jen Wu and Neil Goldsman	61

Investigation of the Diffusion into Si of Au via Frank-Turnbull Mechanism by Monte Carlo Simulation 65
Q.S. Zhang, J.A. Van Vechten and T.K. Monson

SECTION 2: Hydrodynamic and Other Expansion Methods for Device Modeling

Advances in Multi-Dimensional TCAD (Invited) (Abstract) 69
Wolfgang Fichtner

Three-Dimensional Hydrodynamic Simulation of Submicron MOSFET's 70
Rajiv Madabhushi, Qi Lin, Gwo-Chung Tai, Neil Goldsman and Isaak D. Mayergoyz

Three-Dimensional Simulation of the Effect of Random Impurity Distributions on Conductance for Deep Submicron Devices 74
J-R. Zhou and D.K. Ferry

Fast and Accurate HFET Modelling for Microwave CAD Applications 78
R. Drury and C.M. Snowden

Dual Energy Transport Model for Advanced Device Simulation 82
Zhiping Yu, Lydia So, Edwin C. Kan and Robert W. Dutton

Galerkin Methods for the Boltzmann Equation Using Variable Coordinate Systems 87
Christian Ringhofer

Application of Hierarchical Transport Models for the Study of Deep Submicron Silicon N-MOSFETs 91
S. Jallepalli, C.-F. Yeap, S. Krishnamurthy, X.L. Wang, C.M. Maziar and A.F. Tasch Jr.

Impedance Field in Submicron n^+nn^+ InP Diodes 95
V. Mitin, V. Gružinskis, E. Starikov, P. Shiktorov, L. Reggiani and L. Varani

An Alternative Method for Compact Model Construction and Parameter Extraction 99
Edwin C. Kan and Robert W. Dutton

Transport in Two-Dimensional Quantum Well HEMTs 103
John P. Kreskovsky and H.L. Grubin

Breakdown Simulation of Semiconductor Devices Including Energy Balance and Lattice Heating 107
Y. Apanovich, R. Cottle, E. Lyumkus, B. Polsky, A. Shur, A. Tcherniaev and P. Blakey

A General Hydrodynamic Solver for Deep Submicron Silicon Devices 111
Mei-Kei Ieong and Ting-wei Tang

Inclusion of Viscous Effects in the Hydrodynamic Modeling of Ultrasmall Silicon Devices 115
Ting-wei Tang and Joonwoo Nam

Formulation of a Full Dynamic Transport Model for Heterostructure Devices 119
R. Khoie

Tunneling and Its Inclusion in Analytical Models for Abrupt HBTs 123
Shawn Searles and David L. Pulfrey

Device Modeling using Hydrodynamic Balance Equations 127
J. Cai, H.L. Cui, N.J.M. Horing, X.L. Lei, E. Lenzing and B.S. Perlman

Effect of Heavy Doping on Conduction Band Density of States: Implications for Device Modelling 131
C.G. Morton, J. Wood, W. Batty and C.M. Snowden

Hydrodynamic Model and Jump Conditions for Discontinuous Mass and Electric Potential in Heterostructure Devices 135
Emad Fatemi

SECTION 3. Quantum Transport and Quantum Devices

Quantum Device Simulation: Open Boundary Conditions for the Schrödinger Equation (Invited) 139
Craig S. Lent, Manhua Leng and P. Douglas Tougaw

Applications of 3D Quantum Transport Simulations 145
D. Z.-Y. Ting, S.K. Kirby and T.C. McGill

Self-Consistent, Quasi-3D Simulation of Quantum Waveguide Couplers 149
A. Galick, M. Macucci, U. Ravaioli and T. Kerkhoven

Open Boundary Conditions for Multidimensional Electronic Scattering States 153
Henry K. Harbury, Wolfgang Porod and R. Kent Smith

Quantum Device Simulation Including Interactions (Invited) 157
S. Datta, G. Klimeck, R. Lake, Y. Lee and M.J. McLennan

Electron Mobility in Quantum Wires Limited by Optical-Phonon Scattering 163
T. Ezaki, N. Mori, K. Taniguchi, C. Hamaguchi and U. Ravaioli

Loss of Phase Coherence in Semiconductor Heterostructures Due to the Coulomb Interaction 167
M. Žiger and W. Pötz

Incorporating Spatially Varying Effective-Mass in the Wigner-Poisson Model for AlAs/GaAs Resonant-Tunneling Diodes 171
K.K. Gullapalli and D.P. Neikirk

Evaluation of the Electron Density of States in a Si-SiO_2 Interface Using the Zero-Temperature Green's Function Formalism 175
Dragica Vasileska-Kafedziska, Paolo Bordone and David K. Ferry

Resonant Tunneling Calculations via the Density Matrix in the Coordinate Representation 179
H.L. Grubin and T.R. Govindan

Quantum Transport Using Liouvillean Quantum-Field Dynamics and Functional Approach to Self-Consistent Many-Body and Scattering Effects 183
F.A. Buot and A.K. Rajagopal

Quantum Transport and Coherence in Boundary Limited Electronic Devices: Recursive Dyson Hamiltonian Green's Function and Finite Element Techniques in Heterostructures 187
Clifford M. Krowne

Confined Acoustic Phonon Controlled Relaxation Times in Free Standing Quantum Wells 191
N. Bannov, V. Aristov and V. Mitin

Monte Carlo Simulation of Electron Streaming Caused by Inelastic Acoustic-Phonon Scattering in Quantum Wires 195
R. Mickevičius and V. Mitin

Acoustic Phonon Spectrum and Density of States in Free Standing Quantum Wells 199
V. Aristov, N. Bannov, V. Mitin and M.A. Stroscio

Dynamical Transport Model of a Coupled Electron-Phonon System in a Quantum Wire 203
Julie A. Kenrow and T. Kenneth Gustafson

The Inclusion of a Finite Capture Time in the Numerical Simulation of Quantum Effect Devices 207
M. Gault, H. Matsuura, K. Furuya, P. Mawby and M.S. Towers

A Self-Consistent Solution of One-Dimensional Schrodinger-Poisson Equations by Newton-Raphson Iteration Technique 211
S. Subramanian

Numerical Simulation of the Effect of Surface Charges on Electron Confinement in Quantum Dot Structures 215
Minhan Chen and Wolfgang Porod

A Linear Eigenvalue Method for Calculating the Positions of Transmission Poles and Zeros in Resonator Structures 219
Zhi-an Shao, Wolfgang Porod and Craig S. Lent

Magnetic-Quasi-Bound-State Induced Resonant Coupling of Edge States 223
Manhua Leng and Craig S. Lent

Dynamic Behavior of Coupled Quantum-Dot Cells 227
P. Douglas Tougaw and Craig S. Lent

Monte Carlo Study of Ionized Impurity Scattering in Quantum Wires 231
Lucio Rota and Stephen M. Goodnick

Intrinsic High Frequency Characteristics of Tunneling Heterostructure Devices 235
Chenjing L. Fernando and William R. Frensley

Quantum Device Modeling with Non-Equilibrium Green Functions 239
R. Lake

An Accurate, Efficient Algorithm for Calculation of Quantum Transport in Extended Structures 243
T.J. Godin and Roger Haydock

SECTION 4: High Performance Computing and Algorithms

Domain Decomposition Applied to the Drift-Diffusion Equations (Invited) (Abstract) 247
W.M. Coughran, Jr. and Eric Grosse

Adaptive Grid and Iterative Techniques for Submicron Device Simulation 248
Anand L. Pardhanani and Graham F. Carey

Transport Effect, Hyperbolicity, and Shock Capturing Algorithms for Device Simulations 252
Joseph W. Jerome and Chi-Wang Shu

Numerical Simulation of High-Field Transport Using a Flux-Corrected Transport Algorithm 256
M. G. Ancona and C.R. DeVore

Parallel Solution of Elliptic Boundary Value Problems (Invited) 260
Babak Bagheri, Andrew Ilin, L. Ridgway Scott and Dexuan Xie

Parallel Implementation of a GaAs MESFET Electro-Thermal Simulation on a Transputer-Based System 266
C.S. Tsang-Ping, D.M. Barry and C.M. Snowden

A Parallel Multigrid Solver for Semiconductor Device Equations 270
X. Han, D.M. Barry and M.J. Howes

An Efficient Full-Zone k-p Method for 'on the fly' Calculation of Valence Band Energies in Hole Transport Studies 274
John P. Stanley and Neil Goldsman

SECTION 5: Modeling of Optical Processes and Optoelectronic Devices

A Generalized Monte Carlo Approach for the Simulation of the Coherent Ultrafast Dynamics in Photoexcited Semiconductors (Invited) 278
Fausto Rossi, Stefan Haas and Tilmann Kuhn

Monte Carlo Modeling of the Dynamic Screening Effects on Ultrafast Relaxation of Photo-Excited Carriers in GaAs 284
N. Nintunze and M.A. Osman

Fullband Ensemble Monte Carlo Modeling of High-Field Transport in the ZnS Phosphor of AC Thin Film Electroluminescent Devices 288
Shankar S. Pennathur, Keya Bhattacharyya, John F. Wager and Stephen M. Goodnick

Advances and Opportunities in the Design and Modeling of Vertical-Cavity Surface-Emitting Lasers (Invited) 292
J.W. Scott, B.J. Thibeault, S.W. Corzine and L.A. Coldren

A Transition Matrix Study of Laser Dynamics 298
Muhammad A. Alam and Mark S. Lundstrom

Coupling Classical Carrier Transport, Capture, and Size Quantization in a Quantum Well Laser Simulator 302
Matt Grupen and Karl Hess

Monte Carlo Simulation of High Field Electron Transport in ZnS with Modified Density of States 306
John Fogarty, Weiran Kong and Raj Solanki

Simulation of Periodically Segmented Waveguides as Concurrent Bragg Reflectors and Quasi-Phase-Matched Second Harmonic Generators 310
Cangsang Zhao and Reinhart Engelmann

Numerical Modeling of Injection Induced Carrier Confinement in Quantum-Well (QW) Structures 314
Yijun Cai and Reinhart Engelmann

A Self-Consistent Simulation of the Modulation Response of Quantum-Well Lasers Including the Effect of Strain (Abstract) 318
R. Baca, M. Grupen and K. Hess

Electron-Hole Imbalance in the Active Region of QW Lasers, and its Effect on the Threshold Current 319
G.A. Kosinovsky, M. Grupen and K. Hess

A Numerical Model for Computing the Emission Spectrum in Time-Resolved Photoluminescence Experiments 323
Steven M. Durbin, Dean H. Levi and Jeffery L. Gray

Time-Dependent Simulations of Filament Propagation in Photoconducting Switches 327
P.W. Rambo, W.S. Lawson, C.D. Capps and R.A. Falk

A COMPARISON OF BTE-BASED ELECTRON TRANSPORT SIMULATIONS FOR SILICON

J.M. Higman *
Motorola
Advanced Products Research and Development Laboratory, MD K-10
3501 Ed Bluestein Blvd., Austin, Texas 78721

ABSTRACT

A unique study of several previously reported computer simulations which solve the Boltzmann Transport Equation (BTE) for electrons in silicon has been completed. A total of 47 individuals representing 19 laboratories in 6 countries participated in this comparison, with a total of 21 data sets contributed. Most of the simulations are based on the Monte Carlo particle technique, and have been used here to calculate a set of transport characteristics for intrinsic silicon at room temperature with a homogeneous electric field. From a global perspective the results vary widely, but they provide for the first time a quantitative comparison of many silicon transport models. If we group the data sets according to their bandstructure models and compare groups to one another the separate effects of bandstructure and phonon scattering rates can be seen. In the group of full-band models we observe a striking and unexpected agreement between 4 of the data sets, and the possible significance of this observation will be discussed.

I. INTRODUCTION

In this study, initiated within the National Center for Computational Electronics, we present a comparison [1] of many of the computer simulation codes which have been developed throughout the world for simulating electron transport in silicon. A number of researchers solved the BTE (all but one using the Monte Carlo particle method) for a specific set of conditions, and the results of all of these simulations have been compiled allowing the first direct comparison of a broad spectrum of transport models. The focus of this study is on the physical models and how they might affect the calculated distributions, and not programmer-dependent qualities such as computational speed or the algorithms.

Each simulation code was used to calculate the energy distribution of electrons in homogeneous, intrinsic silicon at room temperature with time-invariant applied electric fields of 30kV/cm, 150kV/cm and 300kV/cm. For each field, the percentage of electrons above 1.1eV, 1.8eV and 3.1eV were calculated.

II. RESULTS

Each of the 21 data sets are listed after the references, with a brief description of the bandstructure model, a list references to the literature which describes the model, and the names of the contributors. It is useful to separate the data sets into three groups according to their bandstructure models, and all of the figures are organized according to this scheme. The first group (sets 1 through 7) consists of models which use effective mass bands; these are either spherical or ellipsoidal constant-energy surfaces, and can be either parabolic or nonparabolic. The second group (sets 8 through 16) contains data sets which use "fit" bandstructure models. These models differ from one another in detail but in principle they are all constructed of analytic functions which contain a number of adjustable parameters such that some properties of the full bandstructure of silicon at higher energies can be emulated, while retaining the simplicity of analytical expressions. For example, the bandstructure may be fit to the density-of-states extracted from a full bandstructure calculation. Set number 8 is included in the fit-band group although it uses only effective-mass bands, but since it uses both X- and L-valleys it is similar to the other models in this group. The

third group (sets 17 through 21) consists of the full-band Monte Carlo simulations where the $E(\mathbf{k})$ relation is calculated using the pseudopotential method, tabulated on a three-dimensional grid in k-space, and interpolated as needed during the monte carlo simulation.

All data sets except 19, 20, and 21 use electron-phonon models based on phenomenological coupling constants. Both the acoustic deformation potential and the intervalley coupling constants are adjusted to reproduce (some) measured data. Data sets 19 through 21 represent the newest attempts at physical models for the transport without any fitting parameters *per se*. In these transport models the pseudopotential description of the crystal is used to calculate both the bandstructure and the electron-phonon interaction, thus treating the free propagation of the electron on the same footing with scattering.

Figures 1(a) through (c) show the total electron-phonon scattering rates (emission plus absorption) for the effective mass, fit-, and full-band groups of data, respectively. Although this is an incomplete representation of any particular transport model - the relative magnitudes of rates for different mechanisms and the impact ionization rate are not shown - it does give some indication about structure in the electronic density-of-states and the strength of the electron-phonon coupling used in the model. At high electric fields the impact ionization scattering rate may have a strong effect on the electron energy distribution and the details of the ionization scattering rates can be found in the references given for each data set.

Figures 2(a) through (c) show the energy distribution of electrons for 30kV/cm, for the three groups of data. Each of the three groups has one model with a much more highly populated tail than the others (sets 2, 10, and 19). The model for set number 2 uses a parabolic, spherical band resulting in a low scattering rate and large population at high energies. In order to understand Figure 2 further we use the effective deformation potential $(DK)_{ij} = [\sum_{\eta} \Delta_{\eta}^2]^{\frac{1}{2}}$ and effective phonon energy $(\hbar\omega)_{ij}^{-1} = \frac{1}{(DK)_{ij}^2} \sum_{\eta} \frac{\Delta_{\eta}^2}{\hbar\omega_{\eta}}$ for the conduction band edge, which were introduced in [14]. The subscripts ij indicate a pair of conduction band minima and Δ_{η} is the coupling constant for phonon branch η. Averaging $(DK)_{ij}$ and $(\hbar\omega)_{ij}$ over all possible minima ij give average effective scattering parameters shown for each model in Fig. 3. The arrows on Fig. 3 indicate the three apparently outlying data sets of Fig. 2, showing that these data sets have the smallest average coupling $< (DK) >$ in their respective groups. Judging from the fit- and full-band data, at lower fields (30kV/cm in this case) the electron-phonon coupling has the strongest influence on the tail of the distribution.

Figure 4 shows the three groups of data as in Figure 2, for an applied electric field of 150kV/cm. In Fig. 4(a) we see again that due to the bandstructure model used in set 2 it falls far from the nonparabolic band models. The fit-band model distributions shown in Fig. 4(b) exhibit considerable scatter in the tail populations at this intermediate field value. The consistency among the full-band models is improving as the field is increased, with set 19 still showing a much larger population at high energies.

The electron distributions for an applied electric field of 300kV/cm are shown in Figure 5. For the effective mass bands, Fig. 5(a), the non-parabolic models predict more consistent distributions than do the fit-band models of Fig. 5(b), and the agreement among the full-band models has improved over the 150kV/cm case. This observation concerning the full-band models is unexpected since the electron-phonon interaction models differ dramatically, from phenomenological coupling constants of models 17 and 18, to the pseudopotential-based, anisotropic coupling of models 19, 20, and 21, and quite different impact ionization models. This suggests that at higher electric fields the bandstructure has more influence on the electron distribution; the fit-band models, which use a variety of analytic expressions and fitting schemes to determine the bandstructure give widely varying results at higher fields. The non-parabolic, effective mass band models (excluding for the moment the spherical-parabolic model of set 2) and full-band models, which have well-defined, consistent bandstructures from one model to another, become more consistent with one another at higher fields. It is important to note that consistent results among any group of models does not necessarily indicate that they are close to the *correct* result.

Figures 6 and 7 show information about the integrated distribution, fraction of electrons above 1.1eV (near the impact ionization threshold) for an applied field of 150kV/cm, and the average energy for each data set at 150 and 300kV/cm. These figures emphasize the disagreement across all data sets, as we observe that the average energy varies by a factor of 5 at these fields (a factor which becomes *worse* at lower fields).

III. CONCLUSION

In summary, we can say that the disagreement among the results is unacceptably large, and from a global perspective the data is discouraging, but by dividing the data into groups according to the band-structure models we see that some encouraging trends exhibit themselves. In addition, the data hints at some potentially fundamental observations about the relative role of bandstructure and electron-phonon scattering in different field ranges: At higher fields the distribution is much less dependent on the details of the scattering rates and is determined largely by the bandstructure. These conclusions are tentative at best, but they indicate the kind of information that is available, and the importance of such broad-based studies.

ACKNOWLEDGMENT

The author would like to thank all of the participants for their cooperation, many of whom made valuable comments and suggestions in addition to contributing data. Thanks also to Chiang-Sheng Yao (Stanford U.) who collected and compiled the data, and to Karl Hess and Bob Dutton who supported this effort from its beginning.

* This work was done at the University of Illinois at Urbana-Champaign while the author was employed there, supported by the NSF through the National Center for Computational Electronics.

REFERENCES

1. These results were initially shown and discussed by several of the participants and others during the Workshop on Computational Electronics, May 1992, University of Illinois at Urbana-Champaign. A detailed manuscript including all contributors as authors has been submitted to *IEEE Trans. Electron Devices*

2. Th. Vogelsang, W. Haensch, J. Appl. Phys., vol. 70, pp. 1493-1499 1991.

3. J.Y. Tang and K. Hess, J. Appl. Phys., vol. 54, pp. 5193-5144, 1983.

4. R. Brunetti, C. Jacoboni, F. Venturi, E. Sangiorgi, and B. Ricco, Solid-State Elect., vol. 32, pp. 1663-1667, 1989.

5. S. Ramaswamy, MS thesis, University of Massachusetts, Amherst, 1992.

6. P.D. Yoder, J.M. Higman, J. Bude, and K. Hess, Semicond. Sci. and Technol., vol. 7, B357-B359, 1992.

7. T. Iizuka, M. Fukuma, Solid-State Electronics, vol. 33, pp. 27-34, 1990.

8. C. Jacoboni, R. Minder, and G Majni, J. Phys. Chem. Solids, vol. 36, pp. 1129-1133, 1975.

9. R. Thoma, H.J. Peifer, W.L. Engl, W. Quade, R. Brunetti, and C. Jacoboni, J. Appl. Phys., vol. 69, pp. 2300-2311, 1991.

10. H.-J. Peifer, B. Meinerzhagen, R. Thoma, and W.L. Engl, IEDM Tech. Digest, pp. 131-134, 1991.

11. C. Fiegna and E. Sangiorgi, IEEE Trans. Electron Devices, vol. ED-40, pp.619-627, 1993.

12. C. Fiegna, et al., IEDM Tech. Digest, pp. 451, 1990.

13. M.V. Fischetti and S. Laux, Phys. Rev. B, vol. 38, pp. 9721-9745 1988.

14. M.V. Fischetti, IEEE Trans. Electron Devices, vol. 38, pp. 634-649, 1991.

15. C. Jacoboni and L. Reggiani, Rev. Mod. Phys., vol. 55, pp. 645-705 1983.

16. A. Phillips, Jr., and P. J. Price, Appl. Phys. Lett., vol. 30, pp. 528-530, 1977.

17. H. Kosina and S. Selberherr, Jpn. J. Appl. Phys., vol. 29, pp. 2283-2285, 1990.
18. P. Hesto, These Doctorat es Sciences, Orsay, France 1984.
19. M. Mouis, These Doctorat es Sciences, Orsay, France, 1988.
20. H. Mizuno, K. Taniguchi, C. Hamaguchi, Semicond. Sci. Technol., vol. 7, pp. B379-B381, 1992.
21. N. Sano, M. Tomizawa, A. Yoshii, Jpn. J. Appl. Phys, vol. 30, pp. 3662-3665, 1991.
22. N. Sano and A. Yoshii, Phys. Rev. B, vol. 45, pp. 4171-4180, 1992.
23. T. Kunikiyo, T. Kamakura, M. Yamaji, H. Mizuno, M. Takenaka, K. Taniguchi, and C. Hamaguchi, Procedings of 1993 VPAD (International Workshop on VLSI Process and Device Modeling), p.40, 1993.
24. X. Wang, V. Chandramouli, C.M. Maziar, and A. F. Tasch, J. Appl. Phys., vol. 73, pp.3339-3347, 1993.
25. S.-L. Wang, N. Goldsman, K. Hennacy, J. Appl. Phys., vol. 71, pp. 1815-1822, 1992.
26. P. D. Yoder, Ph.D. Thesis, Univ. of Illinois, 1993.
27. C. Hao, J. Zimmermann, M. Charef, R. Fauquembergue, and E. Constant, Solid-State Electron., vol. 28, pp.773-740, 1985.
28. M. Charef, Thesis, Universite des Sciences et Technologies de Lille, France (1983).
29. A. Abramo, F. Venturi, E. Sangiorgi, J. Higman, and B. Ricco, IEEE Trans. Computer-Aided Design, vol. CAD-12, no.9, pp.1327-1335, September 1993.
30. C.-S. Yao, D. Chen, R. Dutton, F. Venturi, E. Sangiorgi, and A. Abramo, Procedings of 1993 VPAD (International Workshop on VLSI Process and Device Modeling), p.42, 1993.

Data Sets

1. The first three Legendre polynomials are used [15,25]. (Goldsman, Hennacy, Lin, S.-L. Wang)
2. Parbolic, spherical band model, six equivalent valleys [18,19]. (Hesto, Galdin, Dollfus, Castagne)
3. Non-parabolic, ellipsoidal band model [15,16,17]. (Kosina, Hackel, Selberherr)
4. Non-parabolic, ellipsoidal band model [40]. (K. Tomizawa)
5. Non-parabolic, ellipsoidal band model [27,28]. (Charef, DEssenne, Thobel, Baudry, Fauquembergue)
6. Non-parabolic, ellipsoidal band model [5]. (Ramaswamy, Tang)
7. Non-parabolic, ellipsoidal band model [7,8]. (Iizuka)
8. Non-parabolic, ellipsoidal band model for X-valleys, parabolic, spherical band model for L-valleys [21,22]. (Sano, M. Tomizawa, Yoshii)
9. Anaylytic fit band model [4,24]. (Wang, Maziar)
10. Analytic fit band model [2]. (Vogelsang)
11. Analytic fit band model [4]. (Scrobohaci, Tang)
12. Analytic fit band model [4,9,10]. (Peifer, Thoma, Jungemann, Engl)
13. Analytic fit band model [4,11]. (Fiegna, Brunetti)
14. Analytic fit band model [4,11,12]. (Fiegna)
15. Analytic fit band model [20]. (Mizuno, Taniguchi, Hamaguchi)
16. Analytic fit band model [29,30]. (Abramo, Yao)
17. Full band model [3,36,41]. (Higman, Hess)
18. Full band model [13,14]. (Fischetti, Laux)
19. Full band model [6]; no adjustable electron-phonon parameters in this model. (Yoder, Higman, Hess)
20. Full band model [23]; no adjustable electron-phonon parameters in this model. (Kunikiyo, Mizuno, Kamakura, Takenaka, Taniguchi, Hamaguchi)
21. Full band model [26]; no adjustable electron-phonon parameters in this model. (Yoder, Hess)

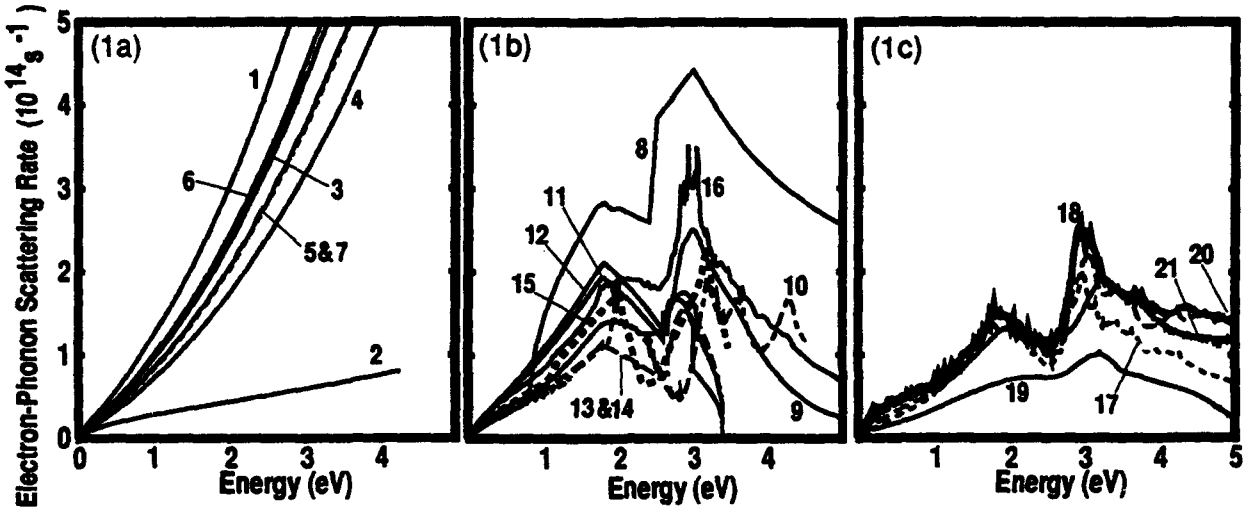


Figure 1. Electron-phonon scattering rates for (a) effective mass band models, (b) fit-band models, and (c) full-band models.

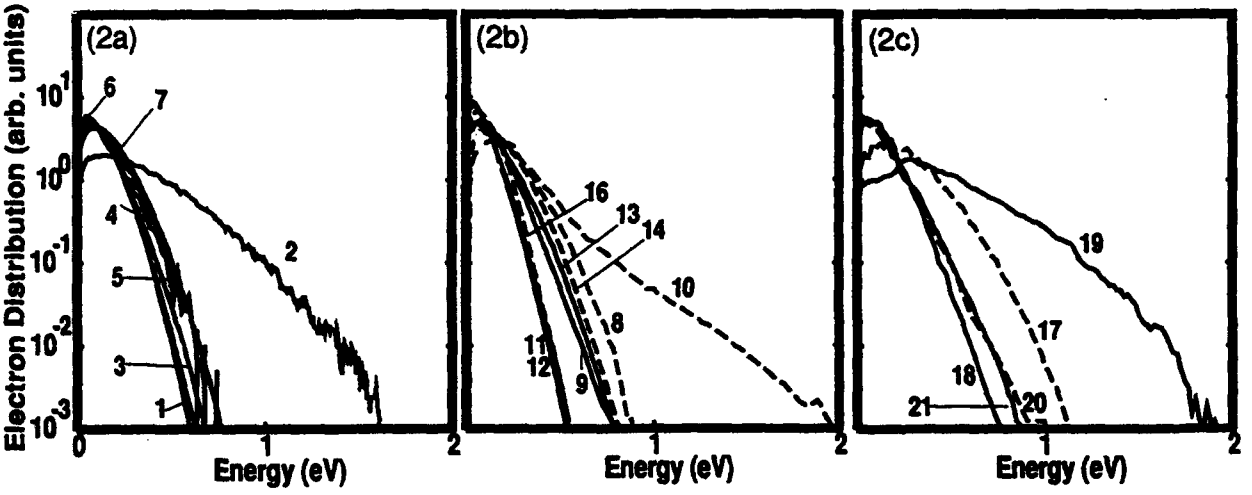


Figure 2. Electron distribution for an applied electric field = 30 kV/cm, for (a) effective mass band models, (b) fit-band models, and (c) full-band models.

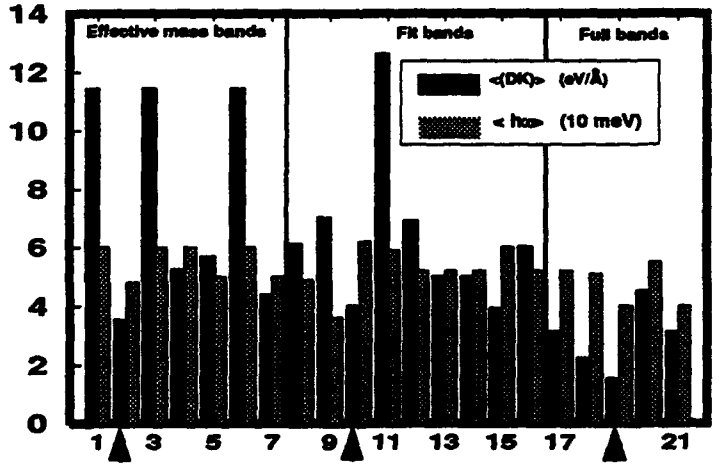
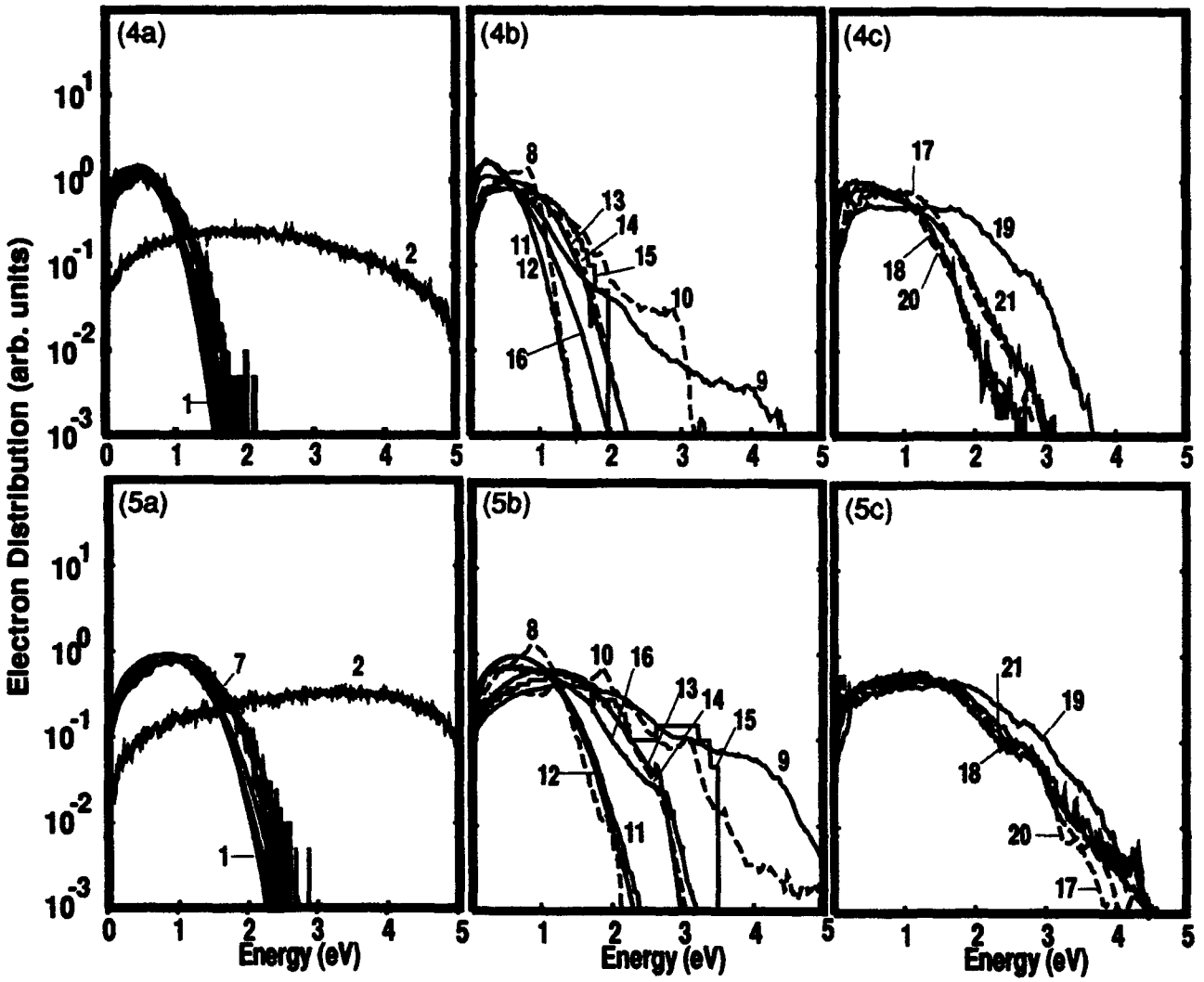


Figure 3. Effective intervalley deformation potential and effective intervalley phonon energy for each data set.



Figures 4 and 5. Electron distribution for 150kV/cm (Fig.4) and 300kV/cm (Fig.5), for (a) effective mass band models, (b) fit-band models, and (c) full-band models.

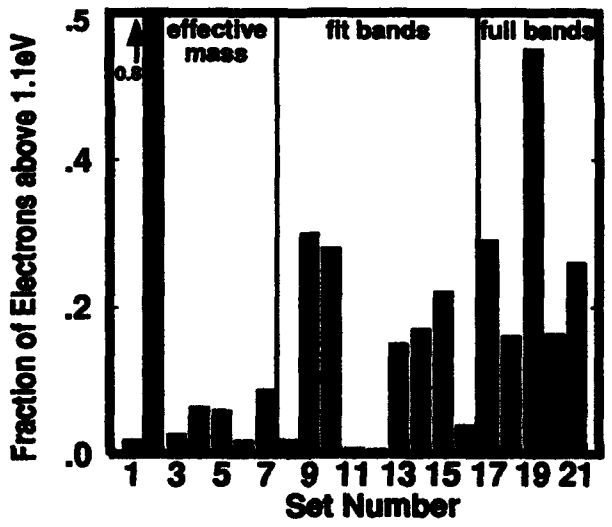


Figure 6. Fraction of electrons above 1.1eV; applied electric field = 150kV/cm.

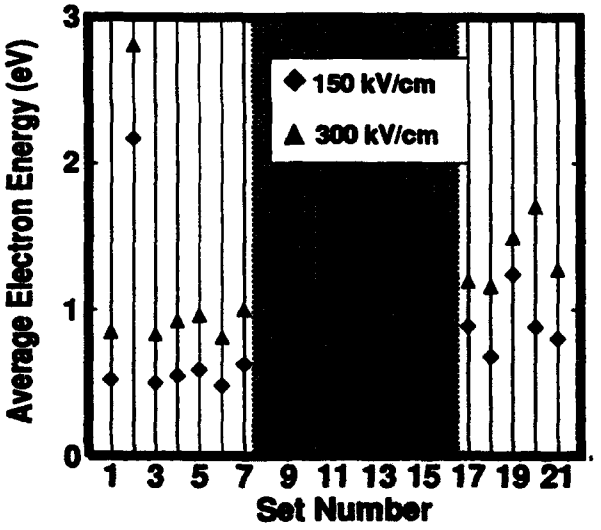


Figure 7. Average electron energy for each set at 150 and 300kV/cm.

CELLULAR AUTOMATON SIMULATIONS OF PLANAR DOPED BARRIER FIELD EFFECT TRANSISTOR IN SILICON

A. Rein, G. Zandler, M. Saraniti, P. Lugli*, and P. Vogl
Physik Department and Walter Schottky Institut, TU München, D-85747 Garching, FRG

Abstract

We report new developments in cellular automata transport simulations and present a study of vertically grown Si ultra-short channel FET's with these new methods. The probabilistic scattering rates for the electric field have been replaced in the cellular automaton by a new deterministic scattering rule in a fully three-dimensional momentum-discretization, leading to a significant suppression of statistical errors. We have also developed a fast multigrid-solver for the Poisson equation that offers the possibility to solve the Boltzmann and Poisson equations asynchronously in a multi-processor environment.

I. Introduction

Simulation of nanostructured devices has become a crucial and strategic part of today's microelectronics [1–3]. Conventional drift–diffusion approaches [4] are no longer valid for modeling of ultra-short devices with a gate length below 100 nm because they cannot predict hot carrier effects such as velocity overshoot quantitatively. Recently, the cellular automaton (CA) approach [5] has been developed as a discrete variant of the Monte Carlo (MC) technique [6,7]. So far, several tests of its applicability to sub-μm device modeling have been carried out successfully [5–9]. We introduce a new implementation of the CA method in this paper that allows precise control and efficient suppression of statistical errors in the CA. The central point is to replace the probabilistic treatment of the electric field used in [5] by a deterministic hopping of the particles in a three dimensional and periodic k-space. Furthermore, to take advantage of the high intrinsic speed of the CA, we have adapted the multigrid method to general device geometries in order to obtain a fast and efficient Poisson solver. These new developments have been implemented in our CA device simulator to study transport in planar doped barrier field effect transistors (PDBFET) with a gate length below 50 nm. We demonstrate that this new ultra-short Si-based device yields a high transconductance and transit-time frequency.

II. New developments in the Cellular Automaton approach for device simulations

Recently [5], the full Boltzmann equation (BE) has been transformed into a CA, where the kinetic terms of the BE are replaced by hopping probabilities in such a way that the equation of motion are fulfilled on the average for an ensemble of quasi particles. In an explicit procedure, the drift term of the BE has been transformed into probabilistic field scattering rates. This corresponds to a substitution of the free flight by a random walk. For very high electric fields, this procedure leads to artificial diffusion effects on the k-space lattice. For a *periodic* momentum discretization, this statistical error can be estimated analytically as follows. The CA-scattering probability to nearest neighbor sites due to the electric field $E(R)$ at lattice site R is given by [5] $P_E = e\Delta t E(R)/(\hbar\Delta k)$, where Δk is the lattice constant of the periodic k-space lattice, Δt the timestep, e the elementary charge and $\hbar$ the Plank constant. A particle moves in one time step with probability P_E to one of the nearest neighbor cells and remains in the cell with the probability $1-P_E$. Associated with this random walk, there is a diffusion in k-space given by $D_{art} = \Delta k^2 P_E(1-P_E)/(2\Delta t)$ which causes an artificial enhancement of the kinetic energy, the entropy and the longitudinal diffusion in real space of the system. In principle, this error can be reduced by a sufficiently small lattice constant Δk but this becomes impractical in a three dimensional momentum space.

We now show that it is possible to transform the drift term of the BE into a new *deterministic*

scattering rule of the CA that completely suppresses this statistical error. The main point is to replace the probabilistic scattering rate by a discrete free flight. We derive this scattering rule by calculating the number of time steps N a particle needs to change its momentum by an amount equal to the lattice constant Δk in k-space. To illustrate the procedure, we restrict ourselves to one dimension; the generalization to more dimensions is straightforward. Integration of the semiclassical equation of motion $\dot{k} = eE(r(t))/\hbar$ gives

$$\Delta k = k(t+\tau) - k(t) = \frac{e}{\hbar} \int_{t}^{t+\tau} dt' E\big(r(t')\big) \ . \tag{1}$$

Let us denote the initial time by $t = t_0$, and assume that $\tau = N\Delta t = t_N - t_0$ and set the real space position at time t_i equal to lattice vector $R(t_i)$. The discrete version of Eq. (1) reads

$$\Delta k = \sum_{i=0}^{N} \Delta t \frac{e}{\hbar} E(R(t_i)) \ , \tag{2}$$

which is a condition for N and yields a deterministic scattering rule for the electric field: A particle remains in its k-cell for N time steps and hops subsequently into its nearest neighbor cell. Consequently, this procedure confines the statistical error to one k-cell. With this procedure, only of the order of 10^3 3–D k-cells are required for a nonparabolic band structure up to 2 eV. The lattice we have chosen is a hexagonal close-packed structure where each cell has twelve nearest neighbors. The restriction to nearest neighbors transforms the drift-term of the BE into a *local* interaction on momentum cells, in complete analogy to the treatment of the real-space diffusion-term of the BE [5].

Importantly, we found that the new implementation of the CA does not require more computer time per iteration than our earlier two–dimensional implementation [5] even though it is significantly more accurate.

III. Multigrid-solver for the Poisson equation

An important component of self consistent device simulations is an efficient solver for the Poisson equation (PE). For a sub-micron device with high doping such as a Si MOSFET, for example, we find that the solution of the PE dominates the total computer time and constitutes a bottleneck of the overall simulation if we employ the standard SOR (successive over–relaxation) algorithm. We have therefore developed an iterative multigrid PE solver to provide an efficient coupling to the CA.

The basic idea of the multigrid approach [10,11] is to improve an approximate solution of the PE on many length scales simultaneously. Indeed, the major deficiency of the SOR is to reduce errors only on the length scale of the smallest grid. The SOR method tends to reduce local errors within a few iterations but often shows "critical slowing down" for the global, long-wavelength errors. In contrast, the multigrid method shows the same rapid convergence on all wavelengths.

We have implemented the multigrid method for general device geometries with irregular two-dimensional grids, allowing for any type of electrostatic boundary condition. For rectangular grids, we use a "zebra"-type line relaxation scheme along both x- and y-directions[11]. When the ratio of the grid spacing in x- and y-direction lies in the range $0.85 \le \Delta_x/\Delta_y \le 1$, we employ instead a checkerboard point relaxation method [11].

We find this multigrid solver to be 5 to 10 times faster than the SOR and consequently to provide a significant speed-up. The gain in computer time is the larger the more complex the boundary conditions are. This algorithm thus eliminates the bottleneck posed to the simulation by the PE and offers the attractive possibility to solve the CA and the PE concurrently in an asynchronous way since both methods

are iterative and require comparable computer time per iteration. We are currently developing decoupling schemes for the solution of BE and PE in the time-domain and in the space domain by slicing.

IV. Transport simulations of PDBFETs

In order to test the present new CA, detailed calculations of Si-PDBFET's [12] have been carried out. This transistor is a vertically grown variant of a Si-MOSFET that contains a δ-p^+-layer in the intrinsic region between the contacts instead of a homogeneous p-buffer (Fig. 1 (a)). Typical gate lengths that can be achieved are 50 nm or smaller. Consequently, one may expect a high transconductance and other short-channel effects in such a device. In the present simulations, we used a gate length of 50 nm and have varied the thickness of the p^+-layer from 5 to 20 nm. We found that a doping concentration of the δ-layer up to $5x10^{18}cm^{-3}$ guarantees that no free holes are present to deteriorate the device performance.

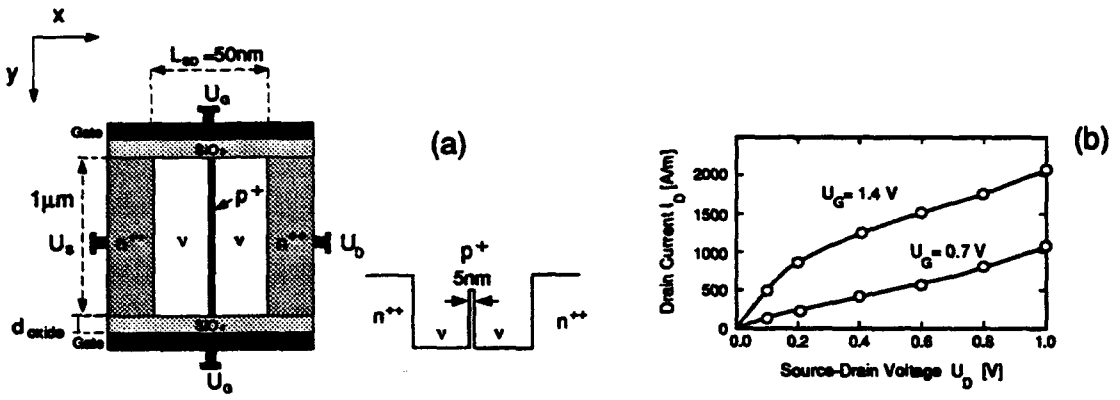


FIG 1. (a) Geometry of a vertically grown 50 nm planar-doped barrier FET. The n^{++} doping concentration is $10^{19}cm^{-3}$, the δ-buffer has a width of 5 nm and a maximum doping concentration of 5 x $10^{18}cm^{-3}$. (b) Typical calculated drain-current characteristics for two gate voltages U_G=0.7 V and 1.4 V. The high channel conductance is indicative of pronounced short channel behavior.

Fig. 1 (b) depicts the computed drain characteristics of a PDBFET with a 5nm δ-layer. The results show typical short-channel effects. In particular, the drain current does not saturate at higher drain voltages. This is due to the fact that the drain current cannot be efficiently controlled by the gate. In addition, velocity overshoot already appears at low drain voltages, as shown for a bias point at U_D=0.2 V and U_G=1.4 V in Fig. 2 (b). The corresponding longitudinal electric field is plotted in Fig 2 (a). The dashed line denotes the field in the bulk diode, which forces the electrons to remain in the n^{++} regions, whereas the full line shows the field in the inversion channel. In contrast to the continously increasing field in the inversion channel of an ultra-short MOSFET, the field in the two intrinsic regions is nearly homogeneous and has a magnitude of approximately 20 kV/cm. In the narrow p-buffer, on the other hand, there is a strongly inhomogeneous field that causes velocity overshoot of the carriers. For higher drain voltages, the field maintains its high value from the p-layer through the complete i-zone up to the n^{++} region of the drain contact. This leads to velocity overshoot nearly over the total channel length.

Our calculations predict a very high transconductance of at least 1000 mS/mm and a maximum transient time frequency of about 200 GHz. In contrast to short channel MOSFETs, where the junction fields reach values of more than 700 kV/cm, the significantly lower electric fields in the studied PDBFET's cause impact ionization to be considerable reduced.

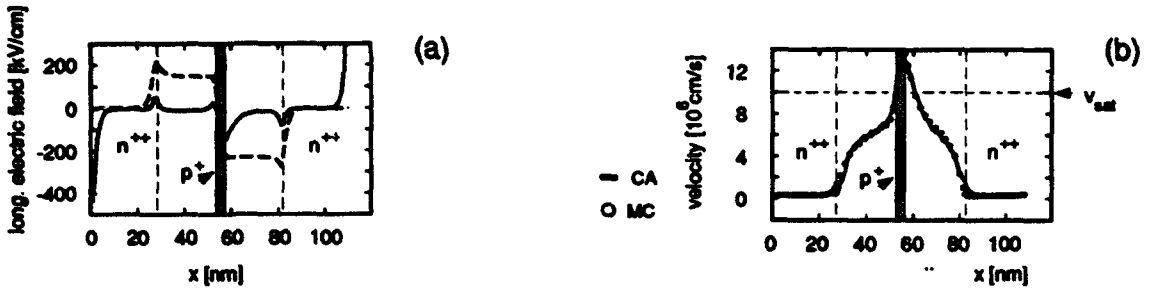


FIG 2. (a) Longitudinal electric field for a bias point at U_G=1.4 V and U_D= 0.2 V (other parameters as in Fig. 1). The field in the inversion channel (full line) is approximately constant in the intrinsic regions and exhibits a sharp maximum in the p-layer. The field in the bulk diode (dashed line) confines the electrons to the n^{++} regions. (b) Comparison of CA and ensemble Monte Carlo simulations for the vertically averaged drift velocity of the electrons. The agreement is excellent. For the chosen bias, velocity overshoot only occurs close to the p-buffer. v_{sat} denotes the saturation velocity.

V. Conclusion

We have presented new improvements in the cellular automaton approach for high field transport in semiconductors. A deterministic rule for the electric field in the CA leads to a dramatic reduction of the statstical errors in a fully three dimensional k-space discretization. In very good agreement with Monte Carlo results, we demonstrated the high speed capability of Si-PDBFET's resulting in a transconductance of 1000 mS/mm, a transit-time frequency of 200 GHz and no relevant influence of impact ionisation compared to short channel MOSFETs.

+Permanent address: Dipartimento di Ingegneria Elettronica, Universitá di Roma "Tor Vergata", I-00133 Rome, Italy

References

1. S. Selberherr, "Analysis and Simulation of Semiconductor Devices", (Springer, Wien, 1984)
2. W. L. Engl, editor, "Process and Device Modeling" (Elsevier, Amsterdam, 1986)
3. K. Hess, J. P. Leburton, U. Ravaioli, "Computational Electronics" (Kluwer Academic, Norwell, MA, 1991)
4. W. Haensch, "The Drift Diffusion Equations and its Applications in MOSFET Modeling", (Springer, 1991, Vienna)
5. K. Kometer, G. Zandler and P. Vogl, *Phys. Rev.* B 46, 1382 (1992)
6. M. Fischetti and S. E. Laux, *Phys. Rev.* B 38, 9721 (1988)
7. C. Jacoboni and P. Lugli, "The Monte Carlo Method for Semiconductor Device Simulation", (Springer, 1989, Vienna)
8. G. Zandler, A. Di Carlo, K. Kometer, P. Lugli, P. Vogl and E. Gornik, *IEEE Electron Dev. Letters* 14, 77 (1993)
9. G. Zandler, A. Rein, M. Saraniti, P. Vogl, P. Lugli, *Proceedings of the 23rd ESSDERC*, to appear in *Microelectronic Engineering*, (1994).
10. A. Brandt, "Guide to Mulitgrid Development," in Multigrid Methods, edited by W. Hackbusch and U. Trottenberg (Springer Lecture Notes in Mathematics No. 960) (Springer Verlag, New York, 1982), p.220
11. W. Hackbusch, *Multi-Grid Methods and Applications* (Springer-Verlag, New York, 1985)
12. H. Gossner, I. Eisele and L. Risch, *Jap. Journal of Appl. Phys.*, (1994).

FULL BAND S-MATRIX SIMULATION OF ELECTRON TRANSPORT IN SI

Carl Huster and Mark Lundstrom
1285 Electrical Engineering, Purdue University, W. Lafayette, IN 47907
(317) 494-3372 huster@ecn.purdue.edu lundstro@ecn.purdue.edu

Abstract

The scattering matrix approach to device simulation is extended to include the full bandstructure of silicon as evaluated by the emperical psuedopotential method. Both a determinsitic and a fast stochastic solution technique are presented. Results are given for bulk silicon and for a non self-consistent structure. Descretization of the Brillouin Zone is idetified as an key issue. It is concluded that one dimensional full band simulations are feasible on contemporary workstations.

I. Introduction to the Scattering Matrix Approach

The scattering matrix approach (SMA) is a deterministic technique for solving the Boltzmann Transport Equation. The problem is set by dividing the device into thin slabs and resolving the carrier fluxes between them into modes in momentum space as shown in fig. 1. The thickness of the slabs is typically 10 to 100A and need not be uniform throughout the device. The maximum slab thickness in a particular region of the device is determined by the spatial resolution requested by the user or the maximum allowable potential drop across a slab. Transport across each slab is treated by a scattering matrix. A column of a scattering matrix relates an incident mode in momentum space to all exiting modes in momentum space. The scattering matrices are generated by a Monte Carlo experiment. A number of carriers in a particular incident mode in momentum space are shot at thin slab of semiconductor. The exiting distribution of them in momentum space is mapped onto a column of the scattering matrix.

A self-consistent solution is achieved by a three step iterative process. First, the scattering matrices appropriate for the electric fields throughout the device are determined. This is done by interpolating between matrices in a library. Second the electron and hole fluxes are solved for. Thirdly the new potential distribution is determined by solving Poisson's equation for the new charge distribution. These steps are repeated until convergence has been achieved.

Presently we have two techniques available for solving for the fluxes. The first is a deterministic technique using Gauss-Seidel iterations. It operates by sweeping from left to right through the device and back. At each slab the exiting fluxes are updated based on the current guess at the incoming fluxes. One cycle through the device is a Gauss-Seidel iteration. It should be pointed out that Gauss-Seidel techniques are known to converge slowly and that acceleration techniques do exist. The second technique is a stochastic technique that we refer to as scattering matrix Monte Carlo. In the field of neutron transport the same idea is referred to as condensed history Monte Carlo. The technique operates by following a single flux through the device. Instead of the incoming flux being split to all coupled exiting modes, one is selected at random by treating the column of the scattering matrix as a probability distribution. The flux is then transmitted or reflected to the appropriate mode. The advantage of this technique is that it is extremely fast compared to the deterministic technique. The disadvantage is that statistical noise is present in a solution generated this way.

II. Full Band Work

We consider the full band work done to date to be a feasibility study. The work has focused on the issues of how to discretize the Brillouin Zone and how to deal with the large sparse matrices that result. The problem of discretization actually has two parts. First, due to the symmetry of the Brillouin Zone, only a fraction of it must be discretized. In the case of Monte Carlo simulations this fraction is typically 1/48th. For the SMA, the fraction that must be discretized is larger and depends on the real space geometry, crystal orientation, and electric field direction. If the electric field and the transport direction are both along the same <100> direction then 1/8th of the Brillouin Zone must be discretized (fig. 2). If they lie in the same <111> direction then 1/6th of the zone is required. It should be noted though that the first two conduction bands of silicon are treated, so the required fraction of the zone must be discretized twice, once for each band.

The second part of the discretization problem is how to discretize the required portion of the Brillouin Zone. This does not have such a clear cut solution. For the proof of concept work we have used a straightforward cubic discretization in the components of the wave vector. 7413 cubes were used to discretize the irreducible 1/48th of the Brillouin Zone. From using this discretization we have learned that there are problems with it. Since the quantities of interest are the energy and velocity of the carriers, a discretization in the wavevector is not well suited to resolving these quantities. Second, since a mode boundary cannot appear at a constant energy, it is difficult to get the band to band transitions resolved correctly. We are currently investigating other discretization schemes, but the current work is sufficient to establish the viability of the full band S-matrix simulation.

The discretization in the components of the wavevector mentioned above results in very large sparse matrices. Our matricies are generated using the full band Monte Carlo code from the University of Illinois at Urbana Champaign [4]. A typical matrix is approximately 93000 elements on a side, 0.015% full and requires approximately 15MB of storage. A simple extrapolation significantly overestimates the memory requirements for a device simulation due to the existence of a memory saving technique known as splitting. With this and some other techniques in place, the memory requirements for a 1-D device simulations should be only about 35 MB.

III. Results

Thus far two types of simulations have been demonstrated using the full band matrices. The first is of bulk silicon with an electric field applied. Due to its speed and simplicity this is usually the first test of a new scattering matrix. The simulation is done by imposing periodic boundary conditions on a single slab or matrix and iterating until convergence is achieved. The periodic boundary conditions are equivalent to and infinite chain of such slabs. Results are presented here for electric fields of 300kV/cm (fig. 3) and 1kV/cm (fig. 4). Notice that at 300kV/cm both the SMA and the scattering matrix Monte Carlo closely match the direct Monte Carlo results. This is expected since the scattering matrices were generated with the same Monte Carlo program that produced the results. Notice however that the deterministic solution technique (SMA) resolves the distribution accurately for several orders of magnitude beyond the stochastic techniques (direct Monte Carlo and scattering matrix Monte Carlo). At 1kV/cm the results do not match so well due to the poor discretization at low energies and the problems it has with coupling to the second band.

Results using the SMA and scattering matrix Monte Carlo for a non-self consistent model structure are presented in fig. 5. This structure features three electric field regions of 1kV/cm, 300kV/cm and 1kV/cm with periodic boundary conditions applied. Notice that both the velocity overshoot and undershoot are well resolved. Also notice that the low field saturation velocity is about four times higher than it should be. One factor of two is due to our discretization of momentum space. The other factor of

two is due to the fact that the Monte Carlo program was designed for studying high field transport and does not accurately reproduce low field results. There are two primary differences between the two results. First, since the scattering matrix Monte Carlo technique is stochastic, it shows statistical noise. Second, the scattering matrix Monte Carlo takes 57 seconds to run on an RS6000/580 while the SMA takes 2500 seconds using a Gauss-Seidel iteration accelerated by a technique known as fine mesh rebalancing.

IV. Conclusions

We have four important conclusions about the full band scattering matrix approach. First, the results presented here demonstrate that one dimensional full band self consistent simulations are feasible on contemporary workstations. Second, discretization of the Brillouin Zone is a key issue that must be addressed. Third, the scattering matrix Monte Carlo technique offers a very fast way of doing Monte Carlo simulations. Finally, the results suggest that two dimensional full band simulations should be feasible on high end workstations or parallel computers.

[1] A. Das and M. S. Lundstrom, "A Scattering Matrix Approach to Device Analysis," *Solid-State Electron.*, Vol. 33, pp. 1299-1307, 1990.

[2] M.A. Alam, M. A. Stettler, and M. S. Lundstrom, "Formulation of the Boltzmann Equation in Terms of Scattering Matrices", *Solid-State Electronics*, Vol. 36, pp. 263-271, 1993.

[3] M. A. Stettler and M. S. Lundstrom, "Self-consistent scattering matrix calculation of the distribution function in semiconductor devices.," *Appl. Phys. Lett.*, Vol. 60, pp. 2908-2910, 1992.

[4] K. Hess, ed., *Monte Carlo Device Simulation: Full Band and Beyond*, Kluwer Academic, Boston (1991).

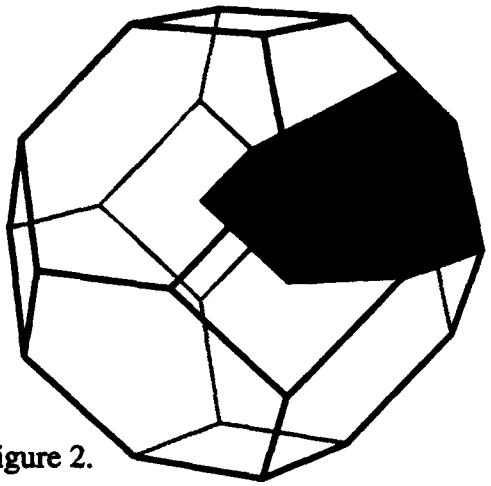
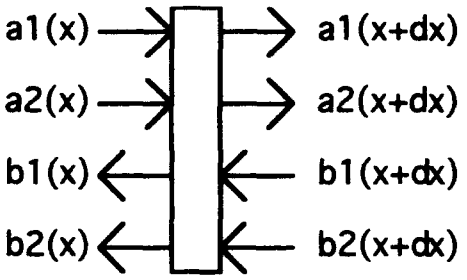


Figure 2.



$$\begin{bmatrix} a1(x+dx) \\ a2(x+dx) \\ b1(x) \\ b2(x) \end{bmatrix} = \begin{bmatrix} t11 & t21 & r'11 & r'21 \\ t21 & t22 & r'21 & r'22 \\ r11 & r21 & t'11 & t'21 \\ r21 & r22 & t'21 & t'22 \end{bmatrix} \begin{bmatrix} a1(x) \\ a2(x) \\ b1(x+dx) \\ b2(x+dx) \end{bmatrix}$$

Figure 1.

Fig. 1 Section of the Brillouin zone discretised for one-dimensional transport simulation. The electric field is assumed to lie along a <100> direction. Symmetry considerations dictate that 1/8th of the BZ, or 6 irreducible wedges are needed.

Fig. 2 Definition of fluxes and the full band scattering matrix. Subscripts refer to the band, 1 or 2, and each flux is an M X 1 vector, where M=22239.

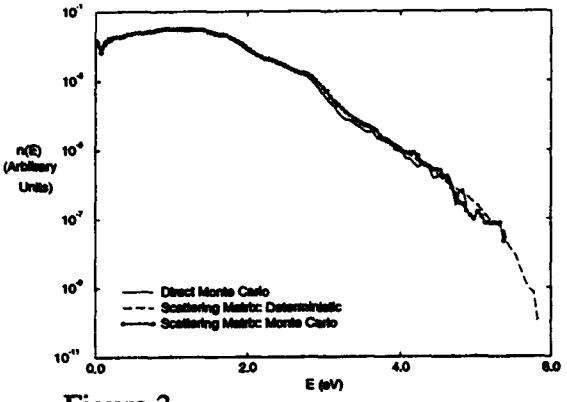


Figure 3.

Fig. 3 Computed energy distributions for electrons in bulk Si at an electric field of 300kV/cm. For comparison, we also display the results of a direct Monte Carlo simulation using the same program used to evaluate the S-matricies [4]. For the S-matrix approach, we show results for a deterministic solution procedure as well as for Monte Carlo solution procedure.

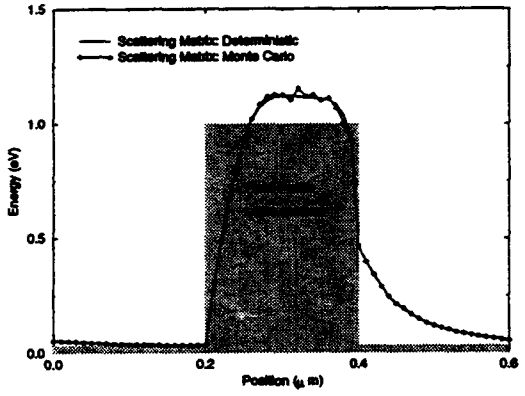
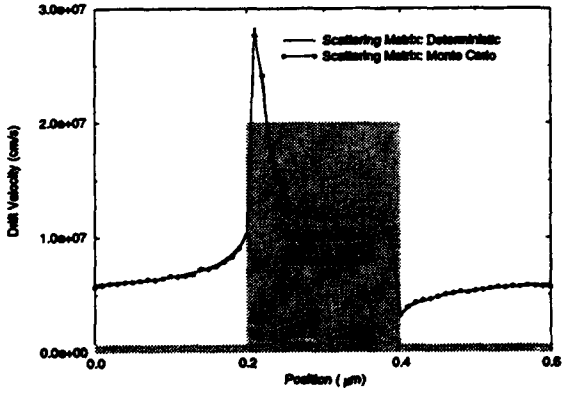


Figure 4.

Figure 5.



Figs. 4&5 Computed average energy and velocity vs. position for electron transport in a model Si device consisting of a low-high-low field profile. Again, results are shown for both the deterministic and Monte Carlo solution techniques. The Monte Carlo S-matrix solution took 57sec on an IBM RS-6000/580 workstation, most of which was overhead involving loading of the matricies.

AN ALTERNATIVE SOLUTION OF THE BOLTZMANN EQUATION: THE "SCATTERED PACKET METHOD"

J.P. NOUGIER, L. HLOU, P. HOULET, J.C. VAISSIERE, L. VARANI
Centre d'Electronique de Montpellier, Université Montpellier II
34095 Montpellier Cédex 5, France

Abstract

We present an original technique for the solution of the Boltzmann equation in bulk semiconductors: the *Scattered Packet Method*. This method intends to combine advantages and to overcome shortcomings of the direct solutions of the Boltzmann equation and of the Monte Carlo methods. The detailed procedure of the *Scattered Packet Method* is described and applied to the case of p-type silicon. The results obtained for first and second order transport parameters are found to be in excellent agreement with classical methods.

I. INTRODUCTION

Classical methods of solution of transport equations in semiconductors, such as the Monte Carlo (MC) method and the direct solution of the Boltzmann equation (BE), have shown their efficiency to provide transport coefficients. Anyway both methods have some inherent shortcomings. Direct solutions of the BE calculate distribution functions with high accuracy but don't take into account fluctuations. MC methods, due to the stochastic nature of the procedure and the limited number of carriers involved, meets with difficulties in calculating with high accuracy quantities on a hydrodynamic time scale such as the transient response of drift velocity and energy, small signal coefficients, etc. In order to combine the advantages of the above two methods, we have developed a new technique called the *Scattered Packet* (SP) *Method*. In this communication, the SP procedure is discussed in detail. Then, the theory is applied to the case of the p-type silicon and the results are compared with those obtained through classical methods.

II. THEORY

We consider a volume of the k-space large enough so that the number of carriers outside it is negligible. This volume is a sphere of radius k_{max} which, in spherical coordinates with the polar axis along the applied electric field E, is described by: $0 \le k \le k_{max}$, $0 \le \theta \le \pi$, $0 \le \varphi \le 2\pi$. The bandshapes are taken spherical, so that the variable φ can be omitted due to the symmetry around the electric field.

We discretize this volume in meshes $D_L = D_{ij}$ centered in $\mathbf{k}_L = \mathbf{k}(k_i, \theta_j)$ and limited as follows:

$$k_i - \Delta k/2 \le k' \le k_i + \Delta k/2 \quad and \quad \theta_j - \Delta\theta/2 \le \theta' \le \theta_j + \Delta\theta/2 \tag{1}$$

The volume of the mesh D_L is:

$$V_L = \int_{\varphi=0}^{\varphi=2\pi} d\varphi \int_{\theta_j-\Delta\theta/2}^{\theta_j+\Delta\theta/2} \sin\theta d\theta \int_{k_i-\Delta k/2}^{k_i+\Delta k/2} k^2 dk \tag{2}$$

Relations (1) and (2) are of course modified at the boundary of the domain.

The ensemble of carriers located in the volume V_L are defined as the *packet* P_L. $n_L(t) = n(k_i, \theta_j, t)$ is the number of carriers contained in V_L at time t. The distribution function is simply given by $f_L(t) = n_L(t)/V_L$.

The purpose of the method is now to determine a time-independent evolution matrix B that gives $n(t+\Delta t)$ when applied to $n(t)$, i.e. in matrix form:

$$[n(t+\Delta t)] = [B][n(t)] \tag{3}$$

The matrix element B_{ML} of [B] is the transition probability from the cell number L to the cell number M during the time interval Δt. In order to increase the accuracy in the calculations of [B] we divide each mesh in

submeshes of volume V_h so that $V_L = \sum_h V_h$. The number of carriers contained in the subcell h is given by $n_h = n_L V_h / V_L$. Now, we want to determine the repartition of this sub-packet on the different meshes of the domain after a time-step Δt sufficiently short so that the probability of having more than one collision during Δt is negligible. In order to calculate this repartition, we use some kind of Monte Carlo procedure with constant time-step. Under the application of an electric field, the carriers of the subcell h make a free-flight of duration Δt which transfer them into another subcell j of centered vector $\mathbf{k}_j = \mathbf{k}_h + e\mathbf{E}\Delta t/\hbar$. Let p_0 be the probability to have no collision during Δt, p_1 the probability for an optical phonon absorption, etc. Therefore the number of carriers having no collisions is $n_j = p_0 n_h$ and these carriers are located in the cell N which contains the subcell j. The element B_{NL} of [B] is incremented by n_j. The number of carriers undergoing an optical phonon absorption is $n_{circle\,k} = p_1 n_h$. These carriers are scattered along a sphere of constant energy according to the angular repartition probability. In a mesh M of this sphere, the number of carriers is (for an isotropic interaction):

$$n_M = n_{circle\,k} \left[\int_{\theta_M - \Delta\theta/2}^{\theta_M + \Delta\theta/2} \sin\theta' d\theta' \right] \left[\int_0^{\pi} \sin\theta' d\theta' \right]^{-1} \qquad (4)$$

The element B_{ML} of [B] is then incremented by n_M. For anisotropic interactions Eq. (4) is slightly modified. The column L of [B] is filled when all the sub-packets of P_L have been scattered in k-space. By repeating the same procedure for all the initial packets we finally evaluate the evolution matrix [B] which satisfies Eq. (3). We notice that [B] depends only on the material, the carrier concentration, the lattice temperature, the electric field, the time-step and the number of meshes used for the discretization of the k-space.

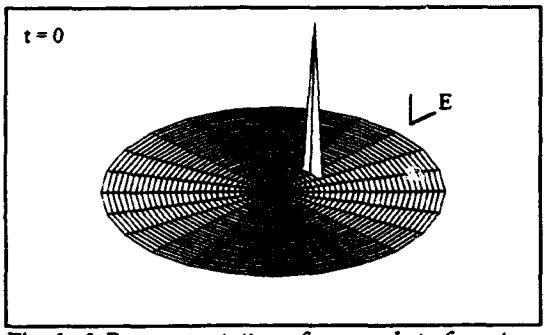


Fig. 1: *3-D representation of one packet of carriers centered in* $k = 10^9 m^{-1}, \theta = 0$. *Calculations are performed for the case of p-Si with Na=0, T=300 K and E=20 kV/cm.*

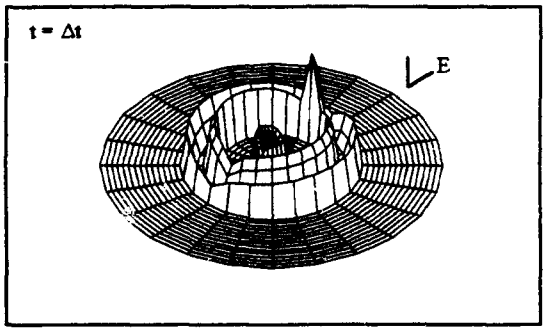


Fig. 2: *Evolution during one time-step* $\Delta t = 1fs$ *of one packet initialy centered in* $k = 10^9 m^{-1}, \theta = 0$. *Calculations are performed for the case of p-Si with Na=0, T=300 K and E=20 kV/cm. The vertical scale is not linear in order to enhance small values of n(k,Δt).*

To better clarify for the reader the repartition of a packet in k-space, let us consider one packet P_L centered in $k = 10^9 m^{-1}$ and $\theta = 0$ at time t=0 as shown in Fig. 1. By construction, the column L of the matrix [B] represents the packet repartition at time Δt when the initial number is equal to unity and located in $\mathbf{k}_L$. We have drawn on Fig. 2 n(t=Δt) which is proportional to the column L. The peak corresponds to the carriers that have been only displaced by the electric field without being scattered, and the circle at the same $|\mathbf{k}|$ refers to carriers which have undergone an elastic scattering. The inner and outer circle correspond to carriers having emitted or absorbed an optical phonon, respectively.

To study the transient regime, at time t=0 the number of carriers in the state $\mathbf{k}_L$ is n_L chosen equal to the thermodynamic value of the number of carriers of this state. Then, using Eq. (3) n(k,Δt) is computed and so on untill the stationary regime is obtained.The resulting algorithm (similar to the lattice-gas cellular-automaton method [1]) is physically equivalent to an ensemble MC method, using a considerable amount of carriers and without any use of random number.

III. CALCULATION OF THE VELOCITY AUTOCORRELATION FUNCTION

By neglecting cross correlation terms between velocities of different carriers, the autocorrelation function of the fluctuations of the drift velocity can be written as follows:

$$C(t) = \overline{\sum_{i=1}^{N} (v_i(t) - \overline{v_d})(v_i(0) - \overline{v_d})} \qquad (5)$$

where N is the total carrier number, $v_i(t)$ the velocity of the carrier i at time t, $\overline{v_d}$ the drift velocity.

Let us define by $F_{LM}(t)$ the ensemble of carriers leaving the state $\mathbf{k}_L$ at an initial time and reaching the state $\mathbf{k}_M$ at time t. By construction, these ensembles verify the two following properties:

(i) they constitute a partition of the whole system,

(ii) the carriers belonging to a given ensemble have the same initial and final velocity: $v(0)=v_L$ and $v(t)=v_M$, where v_L and v_M are the projections along the electric field of the carrier velocities in state L and M.

Therefore, the correlation given by Eq. (5) can be obtained by summing elemenary contributions from each family $F_{LM}(t)$:

$$C(t) = \sum_L \sum_M C_{LM}(t) \qquad (6)$$

with

$$C_{LM}(t) = n_{LM}(t)(v_L - \overline{v_d})(v_M - \overline{v_d}) \qquad (7)$$

where $n_{LM}(t)$ is the carrier number of the family $F_{LM}(t)$. At time t=0 the number of carriers in the state $\mathbf{k}_L$ is n_L chosen equal to its steady state value in the applied electric field E and $n_{L'\neq L}$ is taken equal to zero. Using the SP method $n_{LM}(t)$ and $C_{LM}(t)$ are computed. The number of simulations is equal to the number of meshes in k space (about 500). An acceleration technique described in Ref. [2] is used in order to reduce the CPU time. The diffusion coefficient can then be obtained taking the Fourier transform of Eq. (6).

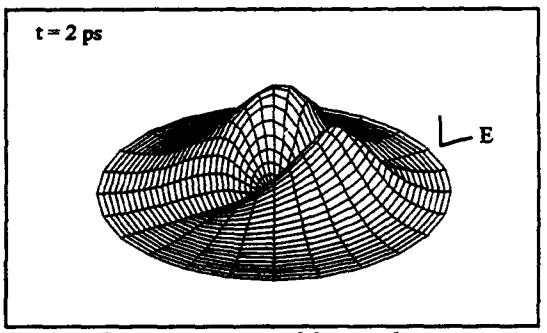


Fig. 3: *3-D representation of the steady-state carrier population. Calculations are performed for the case of p-Si with Na=0, T=300 K and E=50 kV/cm.*

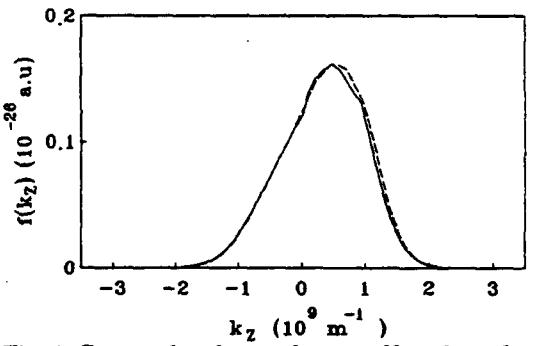


Fig. 4: *Carrier distribution function $f(k_z)$ along the electric field in p-Si with Na=0, T=300 K and E=20 kV/cm. The dashed line refers to the direct solution of the BE and the solid line to the SP method.*

IV. RESULTS

We present the results obtained for the first and the second order transport coefficients in the case of a p type silicon at T=300 K. The microscopic model is based on a single spherical nonparabolic-band and considers scattering with acoustic, impurity and non-polar optical phonon mechanisms as described in Ref. [3].

Figure 3 shows the steady-state number of carriers given by the SP method. The number of carriers in the cells located near k=0 (centre of the plot) and along the direction of the field is small, due to $k \sin\theta \cong 0$ (cf Eq. (2)).

Figure 4 shows the steady-state distribution function $f(k_z)$ along the electric field for an acceptor concentration Na=0. We observe an excellent agreement with results obtained from the direct solution of the BE [3]. The same agreement is found for the drift velocity as a function of time which is reported in Fig. 5, for three different electric fields.

Therefore we have shown that the SP method keeps the accuracy of the direct solution of the BE in calculating first order transport parameters.

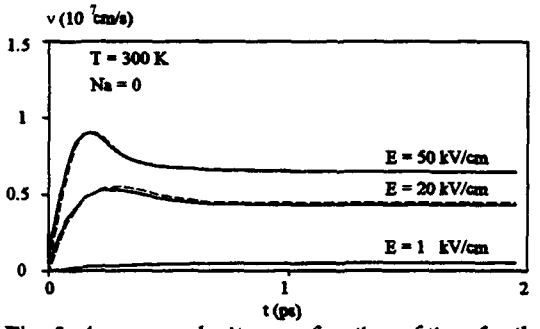


Fig. 5: *Average velocity as a function of time for the reported electric fields in p-Si with Na=0 and T=300 K. The dashed line refers to the direct solution of the BE and the solid line to the SP method.*

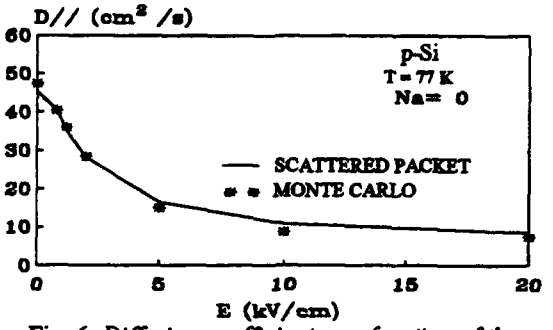


Fig. 6: *Diffusion coefficient as a function of the electric field in p-Si with Na=0 and T=77 K. The stars refer to the MC method and the solid line to the SP method*

As concerning second order transport parameters, Fig. 6 shows the low frequency diffusion coefficient as a function of the electric field. Also in this case the agreement between results obtained by the MC method [5] and the SP method is excellent.

Using about 500 cells, in order to compute precisely the 500*500 matrix elements B_{ML}, we compute (see Eq.(4)) how many particles from each of the 62500 subcells of each cell L are scattered in each of the 500 cells M. Due to the energy conservation, with about 16 steps in θ, this requires 62500*17*500 computations for each scattering mechanism. With optical absorption and emission, acoustical and impurity scatterings, the calculation of the matrix [B] takes about 30 minutes, the stationary regime of the distribution function is obtained after few seconds for a time-step of 1 fs and the correlation functions after few minutes on an IBM 3090.

V. CONCLUSIONS

We have presented a new technique to simulate carrier transport in bulk semiconductors based on an original numerical solution of the Boltzmann Equation. The results obtained for the distribution function, the first and second order transport coefficients have been found to be in excellent agreement with classical methods. The advantages of this method can be summarized as follows: a procedure closed to the corpuscular reality, a high accuracy calculation of distribution functions and fluctuations within a reasonable CPU time.

Developments of the *Scattered Packet Method* towards the device simulation and the study of electronic noise seems to be promising.

ACKNOWLEDGMENTS

This work have been performed within the *European Laboratory for Electronic Noise* (ELEN) and supported by the CEC through the contracts ERBCHRXCT920047 and ERBCHBICT620162. Partial support from the *Centre de Compétences en Calcul Numérique Intensif* (C3NI) and the italian *Consiglio Nazionale delle Ricerche* (CNR) is gratefully acknowledged.

REFERENCES

[1] K. Kometer, G. Zandler and P. Vogl, Phys. Rev. **B46**, 1382 (1992).

[2] L. Hlou, Thèse de Doctorat, Université Montpellier II (France), 1993 (available upon request).

[3] J.C. Vaissiere, Thèse de Doctorat ès Sciences, Université Montpellier II (France), 1986 (available upon request).

[4] J.P. Aubert, J.C. Vaissiere, J.P. Nougier, J. Appl. Phys. **56**, 1128 (1984)

[5] T. Kuhn, L. Reggiani, L. Varani, D. Gasquet, J.C. Vaissiere, J.P. Nougier, Phys. Rev. **B44**, 1074 (1991).

Monte Carlo Simulation of Impact ionization processes in silicon

K. Taniguchi, M. Yamaji, T. Kunikiyo, Y. Kamakura, H. Takenaka, M. Morifuji, K. Sonoda and C. Hamaguchi

Department of Electronic Engineering, Osaka University
Suita, Osaka, 565 Japan

Abstract

The energy-dependent impact ionization rate in silicon is derived by a first-order perturbation theory. The scattering rate is close to that obtained experimentally with a soft x-ray photo-emission spectroscopy. The reasonableness of the calculated results is also supported by the close agreement of simulation results to available experimental data. Key features of the derived impact ionization rate are (1) impact ionization rate for initial electrons with energy below 3 eV shows strong anisotropy, and (2) energy-dependent impact ionization rate has a large power exponent. For practical device simulation, we also derived a simple analytical expression of impact ionization rate under exponentially varying electric field conditions.

I. Introduction

The down-scaling of MOSFETs induces higher electric fields in the channel because the power supply voltage has been scaled less aggressively than device geometries. In deep submicron MOSFETs, the number of hot carriers is expected to increase quite rapidly, leading to the degradation of device characteristics. In order to ensure long term operation of MOSFETs, it is essential to model the behaviors of high energy electrons.
Device simulators developed in the last decade now make it possible to reproduce device characteristics but none of these can accurately estimate the distribution of high energy carriers. Note that high energy carriers directly affect long term reliability of devices as well as electron injection efficiency in FLASH memory. For more predictable T-CAD, high energy carriers have to be simulated more accurately. This requires the use of precise physical models instead of a traditional parameter fitting approach. Among several physical models, an impact ionization rate model is far more important in order to estimate the distribution of high energy carriers. The aims of this paper are twofold: (1) to derive impact ionization rate theoretically and a simple impact ionization model applicable for practical device simulation and (2) to verify the reasonableness of the derived impact ionization rate through the comparison with available experimental data.

II. Theory of impact ionization

Impact ionization process is a kind of electron-electron interaction taking place at the presence of high electric field. When a high energy electron collides with a valence electron, it gives up its kinetic energy to the valence electron which is ionized over the band gap. Thus two conduction electrons and a hole are left.

Theory of impact ionization process in silicon has been well established[1]. The impact ionization rate is obtained from the Fermi's golden rule.

$$S_{I.I.}(1,2 \rightarrow 1',2') = \frac{2\pi}{\hbar}\left[|M_a|^2 + |M_b|^2 + |M_a - M_b|^2\right] \times \delta(\varepsilon_1 + \varepsilon_2 - \varepsilon_{1'} - \varepsilon_{2'})$$

Where 1 and 2 denote the conduction and valence electron states before scattering, while 1' and 2' the conduction electron states after scattering. We used the Coulomb potential as an interaction Hamiltonian. The direct matrix element M_a is given in the forms as

$$M_a = \left\langle \phi_1'(r_1)\phi_2'(r_2) \left| \frac{e^2}{4\pi\epsilon(q,\omega)|r_1 - r_2|} \right| \phi_1(r_1)\phi_2(r_2) \right\rangle$$

To calculate the matrix elements, we need (1) wave functions of electrons involving impact ionization process and (2) a frequency- and wave vector-dependent dielectric function. After expanding the Coulomb potential into Fourier series, the direct matrix element is given by

$$\begin{aligned} M_a \;=\; & \sum_{G_1,G_2,G_{1'},G_{2'}} \frac{e^2}{q^2 V \epsilon(q,\omega)} A_{1'}^*(G_{1'},k_{1'}) A_{2'}^*(G_{2'},k_{2'}) A_1(G_1,k_1) \\ & \times A_2(G_2,k_2)\delta(-k_{1'} + k_1 - k_{2'} + k_2 + G) \end{aligned}$$

In the calculation, both momentum and energy among the particles are conserved through the delta functions. Also, both the normal and umklapp processes were took into account. Final wave vector-dependent impact ionization rate was derived from the integration over an eight-dimensional k space numerically. In our calculations, we used 113 plane waves and 15 reciprocal lattice vectors.

III. Calculated impact ionization rate

Figure 1 shows the calculated impact-ionization rate as a function of initial electron energy. The calculated results scatter in a rather wide range for initial electrons with kinetic energy below 3 eV because of its strong anisotropic nature. The anisotropy diminishes with increasing the electron energy. In Fig. 1, several reported analytical impact ionization rates[2-5] are also plotted. Although they differ in three orders of magnitudes, they all reported that the calculated impact ionization coefficients agree with experimental data. This means that none of these impact ionization rates have not been well verified. In other words, simulated results could be fitted to any experimental data by simply adjusting phonon scattering rates.
Recently Cartier et al. reported a new combined experimental and theoretical effort to find the electron-hole pair production rate[6]. This was achieved by using soft X-ray photoemission spectroscopy and by performing Monte Carlo simulations. The thin solid curves shown in Fig. 1 is the results of their experiment. Although there exist some undulations in the curves due to an artificial fitting to three sets of Keldysh formula[7], our calculated data are essentially the same as theirs.

For more practical use, we derived analytical form of impact ionization rate averaged over all initial electron states with a given energy.

$$S_{I.I.}(\varepsilon) = 1.0 \times 10^{11}(\varepsilon - 1.1eV)^{4.6}$$

Compared with the Keldysh form with power exponent of two, the new isotropic impact ionization rate has much larger power exponent of 4.6, indicating soft impact ionization threshold.

Figure 2 shows the impact ionization rates based on other first principle calculations[8-11]. The overall trend of the calculated results shows similar characteristics since the impact

ionization rate simply reflects the energy band structure of silicon. However, one point to note is that there still exists about one order of magnitude discrepancy among the reported values. The reasons for the discrepancy are not clarified yet. There may be several reasons: choices of the integration method, pseudopotential form factors used, dielectric function, energy allowance used in the numerical calculation.

IV. Comparison with experimental results

(1) *Transient impact ionization using anisotropic and isotropic I.I. rates*

In order to study the reasonableness of the derived isotropic impact ionization rate, we simulated the number of impact ionization events under non-steady state conditions by using isotropic and anisotropic scattering rates. No appreciable differences between the two cases indicate that it is reasonable to use isotropic impact ionization rate instead of more elaborate anisotropic ones without losing any physical meaning. In addition, the use of the isotropic impact ionization rate greatly improves the efficiency of Monte Carlo simulation in terms of memory capacity as well as computational time.

(2) *Impact ionization coefficient*

The number of impact ionization events produced by one carrier per unit length is defined as impact ionization coefficient, which varyies with the electric field as Aexp(-B/E). The open circles shown Fig. 3 are the simulated impact ionization coefficient under constant electric field. Good agreement with the experimental data demonstrates the validity of the impact ionization rate derived.

(3) *Non-local impact ionization coefficient in exponentially varying electric field*

In MOSFETs electric field in the velocity saturation region changes so rapidly that the channel electrons are no more equilibrium. According to simulation studies, the electric field in MOSFETs was found to vary exponentially with distance from the pinch-off point toward the drain. The data points in Fig. 4 show calculated impact ionization coefficients under exponentially varying electric field conditions. All the points below the solid line are the data calculated under increasing electric field condition, while the solid points above the solid line are the data for decreasing electric field conditions. Simulated characteristics length, λ, of the velocity saturation region is simply expressed with the gate oxide thickness and junction depth.

Figure 4 shows two interesting features: (1)under the increasing electric field, non-local impact ionization coefficient is significantly smaller than that in equilibrium state and (2)at lower electric field, the decrease of the impact ionization rate becomes more pronounced.

(4) *Simulation of substrate current*

Figure 5 shows the substrate currents calculated using two different impact ionization models. The solid line shows the experimental data.

A conventional drift-diffusion simulation using the local impact ionization coefficient overestimates substrate current. However, the use of the non-local impact ionization coefficient shown in Fig. 4 results in a reasonable agreement with experimental data. Note that even quite simple device simulator based on the drift-diffusion model can predict impact ionization current correctly once the non-local impact ionization coefficient is implemented in a device simulator.

(5) *Quantum yield*

We simulated quantum yield to further verify the derived impact ionization rate. By using p-ch MOSFET, electrons are injected from the gate electrode to the silicon substrate through the gate oxide. If the injected electrons has enough energy to make impact ionization, some of them create electron-hole pairs in the silicon substrate. The generated holes are collected into the inversion layer while electrons flow to the substrate electrode. By measuring the electrode currents, the average number of generated electron-hole pairs per injected electron is calculated. This is the definition of quantum yield.

Figure 6 shows the calculated quantum yields using the MC simulation together with experimental data[13-15]. The calculated data agree quite well with the data reported by Takagi and Toriumi. The discrepancy among the reported quantum yields may originate from the calibration error of the injected electron energy.

(6) *Transient impact ionization*

We investigated the anisotropic impact ionization in Si MOSFETs reported by Takagi[16]. The devices used for their experiment were n-ch Si MOSFETs with single drain structures on the (100) surface. The measured data shows the gate voltage dependence of the substrate current with different channel directions. They found that the substrate current along the 45° off the ⟨011⟩ direction, meaning ⟨001⟩ direction, becomes larger than that along 0 or 90°. The anisotropy decreases with increasing substrate current. We simulated impact ionization process with similar device structures.

The calculated impact ionization agrees well with the experimental data; the anisotropy diminishes with increasing ionization rate.

V. Conclusions

We theoretically derived impact ionization rate. The energy dependence of the rate was found to be the same as the pair generation rate obtained from the soft x-ray photo-emission spectroscopy. The reasonableness of the models is also supported by the close agreement of the simulation results to available experimental data. Key features of the derived impact ionization rate are (1) impact ionization rate for initial electrons with the energy below 3 eV shows strong anisotropy which directly reflects energy band structure of silicon, (2) energy dependent impact ionization rate shows large power exponent, indicating the soft threshold of impact ionization, (3) we demonstrated that both isotropic and anisotropic impact ionization rates lead to the same results. This means that for more practical Monte Carlo simulation we can use the energy dependent impact ionization rate instead of more complicated wave-vector dependent impact ionization.

Using the Monte Carlo simulation, we derived a simple analytical expression of non-local impact ionization coefficient under exponentially varying electric field conditions. We demonstrated that the newly derived impact ionization coefficient well reproduces substrate current of MOSFETs even in a drift-diffusion device simulation.

Reference

[1] E. O. Kane, *Phys. Rev.*, **159**, 624 (1967).

[2] J. Y. Tang and K. Hess, *J. Appl. Phys.*, **54**, 5139 (1983).

[3] M. V. Fischetti and S. E. Laux, *Phys. Rev.*, **B38**, 9721 (1988).

[4] R. Thoma, H. J. Peifer, W. L. Engl W. Quade, R. Brunetti and C. Jacoboni, *J. Appl. Phys.*, **69**, 2300 (1991).

[5] Th. Volgelsang and W. Hansch, *J. Appl. Phys.*, **70**, 1493 (1991).

[6] E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely, *Appl. Phys. Lett.*, **62**, 3339 (1993).

[7] L. V. Keldysh, *Sov. Phys.*, **JETP 37**, 509 (1960).

[8] N. Sano and A. Yoshii, *Phys. Rev.*, **B45**, 4171 (1992).

[9] J. Bude, K. Hess and G. J. Iafrate, *Phys. Rev.*, **B45**, 10958 (1992).

[10] J. Bude and K. Hess, *J. Appl. Phys.*, **72**, 3554 (1992).

[11] J. Wang and K. F. Brennan, *J. Appl. Phys.*, **75**, 313 (1994).

[12] Y. Kamakura, H. Mizuno, M. Yamaji, M. Morifuji, K. Taniguchi, C. Hamaguchi, T. Kunikiyo and M. Takenaka, *J. Appl. Phys.*, **75**, 3500 (1994).

[13] C. Chang, C. Hu and R. W. Brodersen, *J. Appl. Phys.*, **57**, 302 (1985).
[14] D. J. DiMaria, T. N. Theis, J. R. Kirtley, F. L. Pesavento, D. W. Dong, and S. D. Brorson, *J. Appl. Phys.*, **57**, 1214 (1985).
[15] S. Takagi and A. Toriumi, *Extended Abstracts of the Japan Society of Applied Phyics, Spring Meeting*, 1993, Abstract No.31p-ZX-10.
[16] S. Takagi and A. Toriumi, *Tech. Dig. of International Electron Devices Meeting*, p.711, 1992.

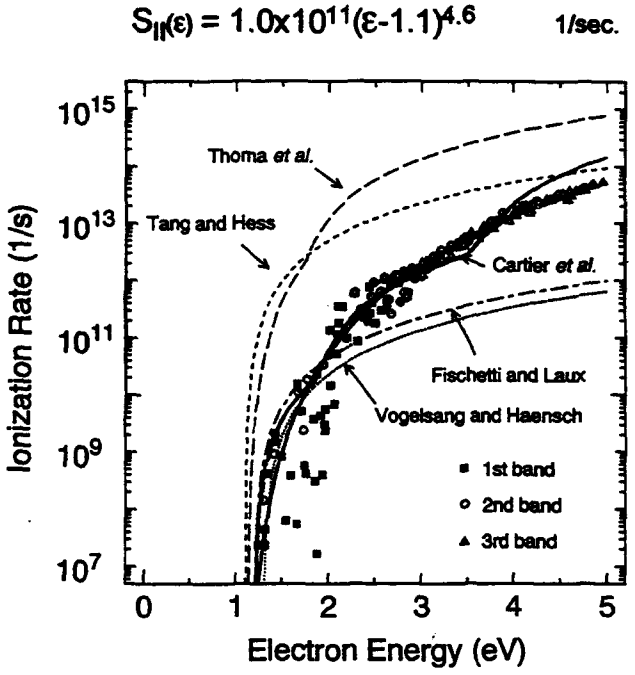


Figure 1. Calculated impact ionization rates. Bold solid line represents a best fitted curve to the calculated impact ionization rates. The other four lines except the solid lines represent impact ionization rate expressed by Keldysh formula. The thin solid curves show the impact ionization rate fitted to a set of three Keldysh formula.

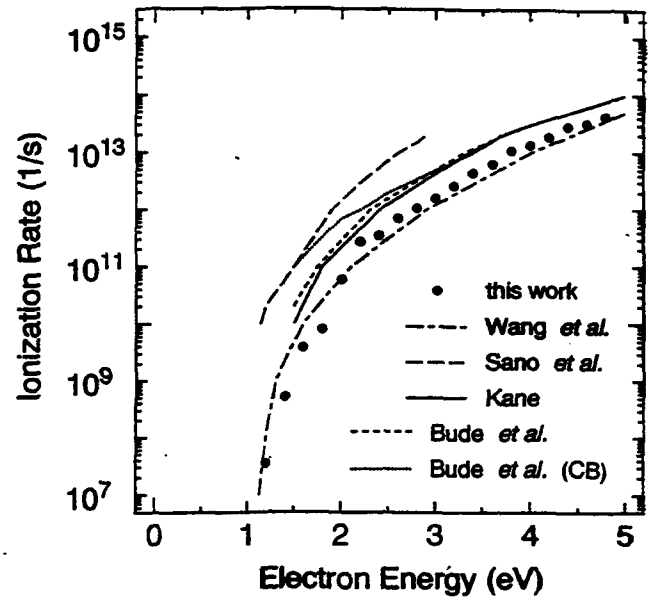


Figure 2. Comparison of reported impact ionization rates averaged over all initial electron states as a function of energy measured from the bottom of the conduction band.

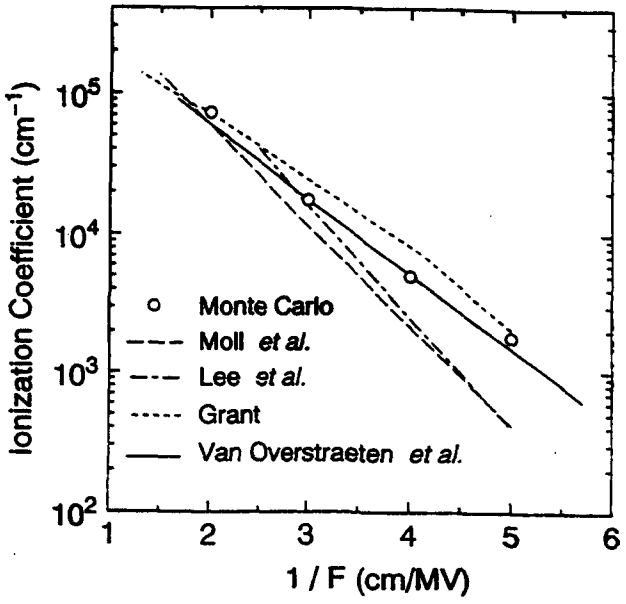


Figure 3. Calculated impact-ionization coefficient as a function of reciprocal electric field with reported experimental data.

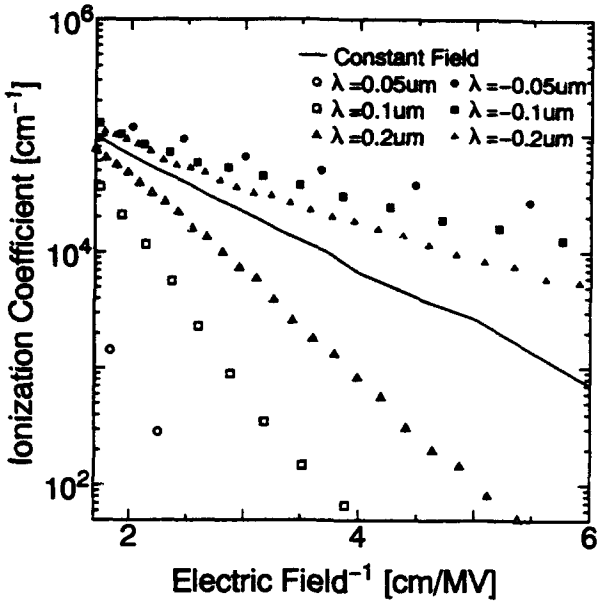


Figure 4. Calculated impact ionization coefficients under exponontially varying electric field conditions.

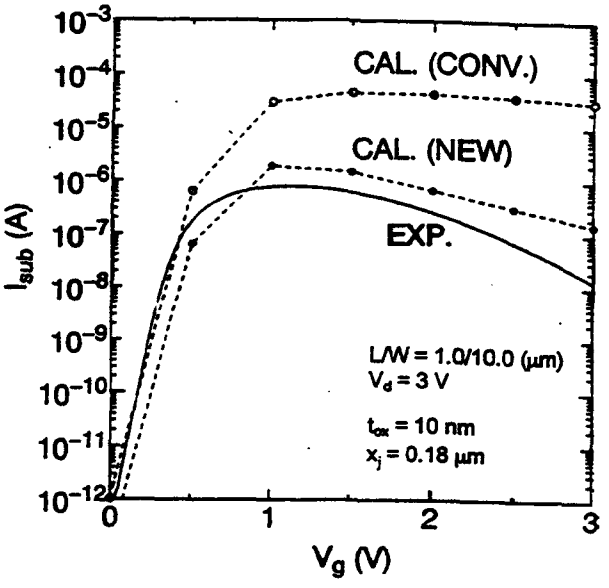


Figure 5. Simulated substrate current together with experimental result at V_D=3.0V for an n-ch MOSFET with the channel length of 1.0 μm. Gate oxide thickness of 10 nm and channel doping concentration of $1.2x10^{17}/cm^2$.

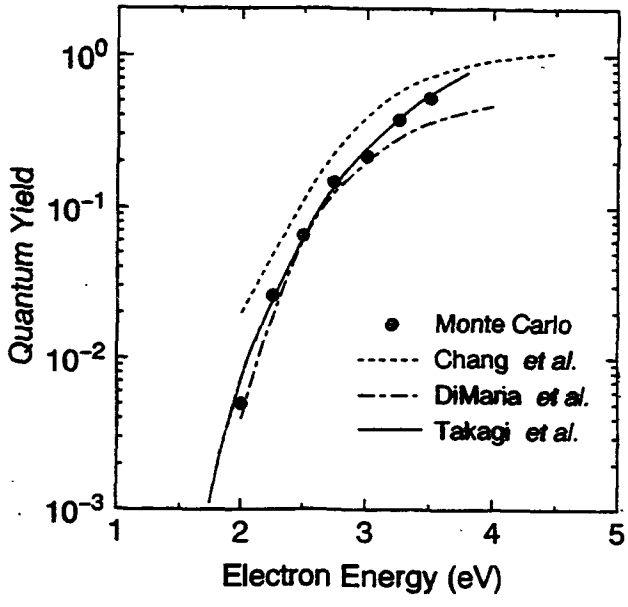


Figure 6. Quantum yield at room temperature as a function of electron energy. Open circles show the calculated results. Curves are the experimental results.

MONTE CARLO SIMULATION OF NONEQUILIBRIUM ELECTRON-PHONON SYSTEM IN QUANTUM WIRES

R. Gaška[1)], R. Mickevičius[1)], V. Mitin[1)], and Michael A. Stroscio[2)]

[1)] *Department of ECE, Wayne State University, Detroit, MI 48202*

[2)] *U.S. Army Research Office, P.O. Box 12211*
Research Triangle Park, NC 27709

Abstract

The ensemble Monte Carlo method for simulating relaxation of the nonequilibrium electron-phonon system in quasi-one-dimensional quantum wires is presented. Employing this method we have found that nonequilibrium (hot) phonon effects in quantum wires are well pronounced for electron concentrations higher than 10^5 cm^{-1} and depend strongly on the energy distribution of excited electrons. Two opposite effects of nonequilibrium phonons in two different relaxation stages have been revealed. The buildup of hot phonons leads to the substantial reduction of the electron gas cooling rate for $t > 0.5$ ps due to strong reabsorption of nonequilibrium phonons. In contrast, the very initial relaxation stage ($t < 0.5$ ps) is faster in the presence of hot phonons.

I. INTRODUCTION

When electrons are heated well above the lattice temperature they relax via cascade emission of phonons and drive the phonon system out of equilibrium. Hot phonons, in turn, affect the entire relaxation dynamics of hot electron gas. Moreover, the quasi-one-dimensional (1D) nature of electrons and optical phonons in quantum wires (QWIs) results in some specific peculiarities of hot phonon buildup that strongly modify nonequilibrium electron-phonon dynamics in QWIs [1]. Therefore, it is necessary to simulate coupled nonequilibrium electron-phonon system self-consistently [1,2].

In this paper we present ensemble Monte Carlo simulation results of the relaxation of 1D nonequilibrium electron-phonon system in QWIs after short pulse excitation. The accurate allowance for the peculiarities of coupled nonequilibrium 1D electron-phonon system within the Monte Carlo technique is discussed.

II. PECULIARITIES OF 1D ELECTRON-PHONON SYSTEM

In general, the phonon wave number is defined by the energy and momentum conservation equations and is given by:

$$q = \sqrt{k^2 + k'^2 - 2kk'\cos\theta}\,, \tag{1}$$

where k is the electron wave number before scattering, $k' = \sqrt{k^2 \pm 2m^*\omega_o/\hbar}$ is the electron wave number after absorption (sign +) or emission (sign −) of the optical phonon of frequency ω_o, and θ is the angle between electron wave vectors before and after scattering. Due to optical phonon quantization and the resultant 1D momentum conservation in quantum wires, electrons can emit or absorb optical phonons with wave vectors which are strictly defined by the electron momentum and the phonon energy. In 1D structures there are just two final states for scattered electrons: forward scattering with $\cos\theta = 1$ or backward scattering with $\cos\theta = -1$. Consequently, there are two possible phonon wave vectors available for emission (and two for absorption) by any single electron:

$$q_{min} = |k - k'|, \quad q_{max} = k + k'. \tag{2}$$

In contrast, in quasi-two-dimensional (2D) quantum wells (or bulk materials) due to existence of additional degree(s) of freedom, $\cos\theta$ can take any value in the range $(-1, +1)$, so that there is an

entire range of a phonon q values from $|k - k'|$ to $k + k'$ available for electron interactions. Fig. 1 depicts q_{min} and q_{max} as a function of electron energy. The dashed area shows the region of available phonon modes for electron to interact with in bulk (3D) or 2D systems. In 1D systems electrons can emit optical phonons only with wave numbers represented by the curve surrounding this dashed area. It is evident from Fig. 1 that nonequilibrium phonon distributions generated by electrons with different energies in 2D and bulk systems overlap. In contrast, electrons in QWIs with different energies generate nonequilibrium phonons in different q-space points. Hence, unlike in 2D and 3D systems nonequilibrium phonons in QWIs can be reabsorbed only by the electrons that have generated them. Consequently, in bulk materials and quantum wells the reabsorption rate for any single electron generally depends on the total phonon population, whereas in QWIs it depends on the occupation number of just two modes with q_{min} and q_{max}. On the other hand, the total phonon population is proportional to the total electron concentration, while the occupation number of single modes is determined by the concentration of electrons with certain energies (*energetic density of electrons*). If the electron energy distribution spreads and the total concentration remains the same, the energetic density of electrons decreases. Therefore, the total nonequilibrium phonon population remains constant because more nonequilibrium phonon modes are amplified, but the occupation number of each mode decreases. As the result the reabsorption rate in 2D and 3D systems does not change with the spread of electron energy initial distribution (it depends on total phonon population), whereas the reabsorption rate in 1D systems decreases (it depends on particular phonon occupation numbers). We come to the fundamental conclusion that, ***unlike in bulk materials and quantum wells, in QWIs hot phonon effects become more pronounced when narrowing energy distribution of excited (injected) hot electrons.***

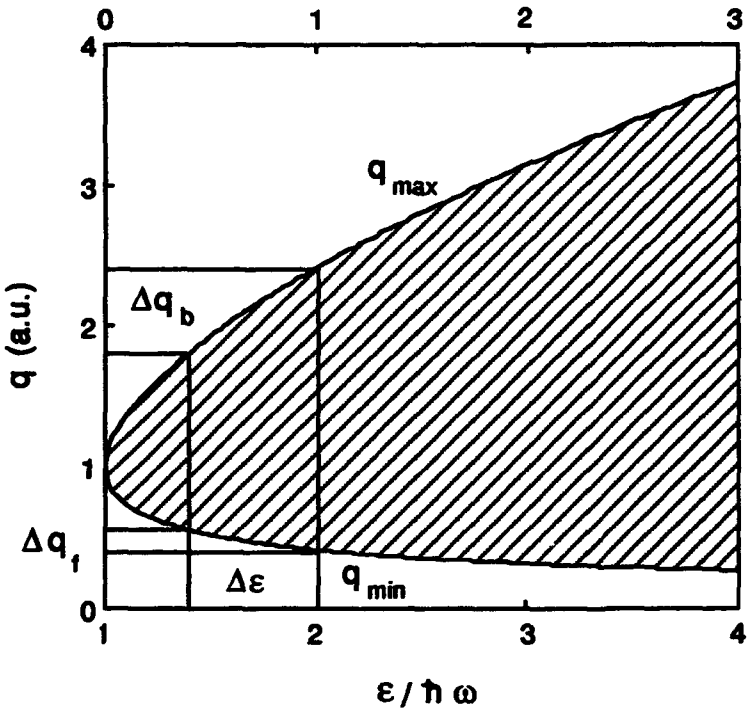


Fig. 1. Minimum and maximum phonon wave numbers versus electron energy normalized to phonon energy. The lower scale applies to emission and the upper for to absorption of phonons. Dashed area shows the range of phonon modes for electron scattering in 2D and 3D systems. The Δq_f and Δq_b depict the spread in nonequilibrium phonon distribution caused by electron energy spread $\Delta\epsilon$ for forward and backward scattering, respectively.

These peculiarities of hot phonon buildup in QWIs must be taken into account in numerical calculations. In Monte Carlo simulations of 3D and 2D nonequilibrium electron-optical phonon systems, the mesh interval for the phonon occupation number Δq is not a crucial parameter, provided that the interval is much less than the q-space region populated by nonequilibrium phonons: $\Delta q \ll q_{max} - q_{min}$. This region is sufficiently large (see Fig. 1) so that above condition is easily satisfied. However, in 1D systems, as we already mentioned above, there are just two single phonon

modes (q_{min} and q_{max}) available for a single electron to interact with. Therefore, when dealing with near-monoenergetic electron excitation virtually coherent phonon modes are amplified. Fig. 1 illustrates how energy broadening of excited electrons leads to the spread of phonon distribution in q-space. One can see that even substantial electron energy broadening ($\Delta\epsilon > \hbar\omega/2$) results in very narrow q-space region for hot phonons generated in forward scattering, Δq_f. Therefore, we generally cannot satisfy the condition $\Delta q \ll \Delta q_{f,b}$ because the fundamental limit of accuracy of determination of phonon wave vector sets the lower limit $\Delta q \geq 2\pi/L_x$, where L_x is the length of a QWI. For a QWI of the length $L_x = 10$ microns this limit is $2\pi/L_x \approx 6 \times 10^3\ cm^{-1}$.

III. MODEL AND METHOD

In our simulations we consider rectangular GaAs QWI embedded in AlAs. We have assumed infinitely-deep potential well for electrons. The hot electron energy dissipation model includes electron interactions with confined longitudinal optical (LO), localized surface (interface) optical (SO) phonons, and inelastic interaction with bulk-like acoustic phonons. We start the simulation of electron relaxation after the initial excitation by a short pulse with a duration of 0.1 ps. We have not simulated electron relaxation in coherent regime ($t < 100\ fs$) which requires a quantum mechanical description. Instead, we have focused our attention on the time range $t > 0.1\ ps$ when electrons can be treated semiclassically [3-6]. We do not take into account the electron-hole interaction. The initial state of electron relaxation accounts for the broadening of the electron energy distribution due to two effects: (i) uncertainty in electron initial energy due to the short electron lifetime at the excited level and (ii) spectral broadening of the exciting pulse with duration of the order of 0.1 ps [7]. In accounting for these effects we assume that they both lead to a Gaussian distribution of electron energy at $t = t_0$ [8], which corresponds to the end of the excitation pulse. We vary the excitation energy ϵ_{ex}, which corresponds to the center of a Gaussian distribution, as well as, $\Delta\epsilon$, the half-width of this distribution.

Hot phonon thermalization due to the decay of optical phonons into acoustic phonons is taken into account by recalculating N_q for every mesh interval at the end of each time step. For simulations we have used the bulk value of the phonon thermalization time $\tau_{ph} = 7\ ps$. We have not taken into account the increase in the acoustic phonon population as a result of the decay of nonequilibrium optical phonons. The reason for this is that acoustic phonons in a QWI embedded in surrounding material with similar elastic properties (GaAs in AlAs in our case) can easily penetrate through GaAs/AlAs interfaces and escape from the QWI. Therefore, we have excellent thermal conductivity and the QWI should not be heated much more than the whole GaAs/AlAs structure. Given that the surrounding AlAs is sufficiently massive, the increase in temperature would be negligible even if the QWI strongly radiates acoustic phonons.

IV. MONTE CARLO SIMULATION RESULTS

We have found that hot phonon effects in QWIs are well pronounced for electron concentrations of the order of $10^5\ cm^{-1}$ and depend strongly on the energy distribution of excited electrons. We have considered various half-widths of Gaussian electron distribution ranging from the extremelly narrow of 4 meV to the broad but still less than optical phonon energy of 30 meV. Figure 2 illustrates electron cooling dynamics in a $150 \times 250 \mathring{A}^2$ QWI at $T = 30\ K$ after initial electron excitation at an energy 4.5 times the LO phonon energy for two extreme limits of Gaussian electron distribution half-widths. For comparison, we plot the electron relaxation dynamics without nonequilibrium optical phonons. One can see that hot phonons lead to a substantial reduction of the electron gas cooling rate for $t > 0.5\ ps$ due to strong reabsorption of nonequilibrium phonons [1]. In contrast, the very initial relaxation stage ($t < 0.5\ ps$) is faster in the presence of hot phonons. As we have already mentioned above, hot phonon effects in both relaxation stages are more pronounced for narrow electron distributions (see Fig. 2). Hence, the higher nonequilibrium phonon populations are created (4 meV), the faster is the very initial relaxation stage.

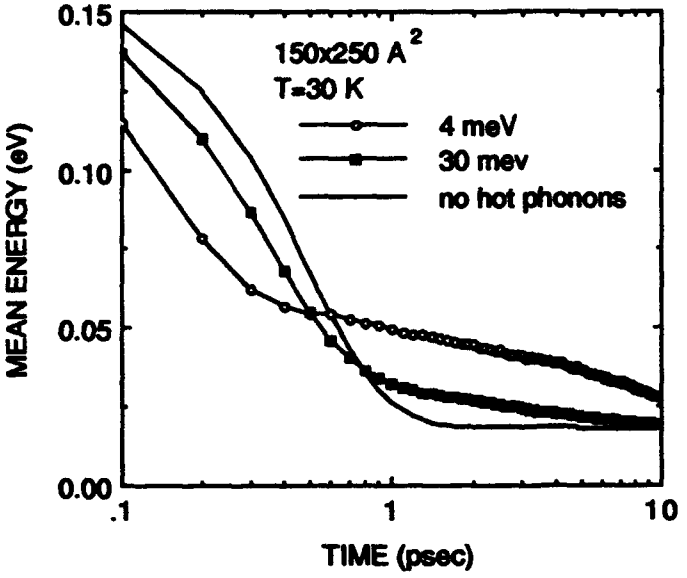


Fig. 2. Mean electron energy as a function of time after initial electron excitation at an energy equal to 4.5 times the LO phonon energy, for two initial widths of electron energy distribution. Electron concentration is $n = 10^5 cm^{-1}$ and lattice temperature is $T = 30\ K$. Solid curve describes the energy evolution for the case of an equilibrium phonon distribution. Results apply to the case of a single-subband QWI neglecting SO phonons.

This effect can be understood if one first considers the temperature dependence of the relaxation rate. At high temperatures both the emission and absorption rates are higher. This leads to fast energy redistribution of excited electrons. The cooling rate of electrons which emit optical phonons increases and that of electrons which absorb phonons decreases because of the $\epsilon^{-1/2}$ energy dependence of 1D density of states and scattering rates. The increase, however, is faster than the decrease due to the same $\epsilon^{-1/2}$ function. Therefore, the total 1D electron gas cooling rate increases when the electron energy redistributes due to emission and absorption of optical phonons. Hence, the very initial electron cooling rate in QWIs increases when increasing the lattice temperature, provided that electrons are excited well above optical phonon energy and thermal equilibrium energy. To observe an appreciable temperature effect on the relaxation rate it is necessary that phonon occupation number be greater than 1. Under phonon equilibrium such occupation numbers could even be unachievable in a solid state. However, due to strong buildup of nonequilibrium phonons at high excited electron concentrations the occupation number for certain phonon modes may be considerably higher than 1. This is why the initial relaxation is faster for higher nonequilibrium phonon occupations and thus, for narrower initial electron energy distributions.

Acknowledgement: The work was supported by the U.S. Army Research Office, the National Science Foundation, and the Center for Compound Semiconductor Materials of the University of Illinois at Urbana-Champaign.

1. R. Mickevičius, R. Gaška, V. Mitin, M. A. Stroscio, and G. J. Iafrate, to be published in Semicond. Sci. Technol. 1994.
2. L. Rota, F. Rossi, P. Lugli, and E. Molinari, to be published in Semicond. Sci. Technol. 1994.
3. A. V. Kuznetsov, Phys. Rev. B 44, 8721 (1991).
4. T. Kuhn and F. Rossi, Phys. Rev. B 46, 7496 (1992).
5. C. J. Stanton, A. V. Kuznetsov and C. S. Kim, to be published in Semicond. Sci. Technol. 1994.
6. J. Schilp, T. Kuhn, and G. Mahler, to be published in Semicond. Sci. Technol. 1994.
7. R. Gaška, R. Mickevičius, V. Mitin, M. A. Stroscio, and G. J. Iafrate, H. L. Grubin, to be published in J. Appl. Phys. 1994.
8. J. P. Leburton and D. Jovanovic, Semicond. Sci. Technol. 7, B202 (1992).

MICROSCOPIC SIMULATION OF ELECTRONIC NOISE IN SEMICONDUCTOR UNIPOLAR STRUCTURES

Luca Varani
Centre d'Electronique de Montpellier,
Université Montpellier II, 34095 Montpellier Cedex 5, France

L. Reggiani
Dipartimento di Fisica, Università di Modena,
Via Campi 213/A, 41100 Modena, Italy

Tilmann Kuhn
Institut für Theoretische Physik, Universität Stuttgart,
Pfaffenwaldring 57, 70550 Stuttgart, Germany

Tomás González, Daniel Pardo
Departamento de Física Aplicada, Universidad de Salamanca,
Plaza de la Merced s/n, 37008 Salamanca, Spain

Abstract

We present a microscopic analysis of electronic noise in semiconductor unipolar structures based on Monte Carlo simulations of the carrier motion self-consistently coupled with a Poisson solver. Current and voltage noise operations are applied and their respective representations discussed. As applications we consider the cases of homogeneous resistors, n^+nn^+ structures, and Schottky-barrier diodes. As a general result, noise spectroscopy is proven to be a source of valuable information to investigate and characterize transport properties of semiconductor materials and devices.

I. INTRODUCTION

The primary quantity which describes electronic noise is the spectral density of current (voltage) fluctuations $S_I(f)$ $[S_V(f)]$. It can be measured more or less directly in different ranges of the frequency f and microscopically interpreted from the calculation of its theoretical counterpart which is the associated correlation function $C_I(t)$ $[C_V(t)]$. This methodology has recently led to the development of a noise-spectroscopy which has proven to be very fruitful for investigating transport properties of materials and devices [1]. In this paper we deal with the problem of how simulating electronic noise from a microscopic point of view. To this end, we make use of the Monte Carlo (MC) technique which, by naturally incorporating all the microscopic noise sources, has recently emerged as a very powerful method. The main issues which will be addressed are: (i) to present a general theory and the algorithms for the calculation of the current and voltage spectral densities; (ii) to investigate systems with increasing degree of complexity; (iii) to decompose the obtained spectra in terms of their sources and spatial contributions.

II. THEORY

In studying electronic noise two different modes of operation, which are mutually exclusive, can be used: current noise operation and voltage noise operation. In the former, the voltage drop at the terminals of the device is kept constant in time and the current fluctuations in the external circuit are analyzed. In the latter the current in the device is kept constant in time and the voltage fluctuations at its terminals are analyzed. Both modes are of interest since, as it will be shown in the following, they provide different and complementary information.

From the Wiener-Khintchine theorem [2] it is:

$$S_X(f) = 2\int_{-\infty}^{+\infty} exp(i2\pi ft) C_X(t) dt \tag{1}$$

$$C_X(t) = \overline{\delta X(t')\delta X(t'+t)} \tag{2}$$

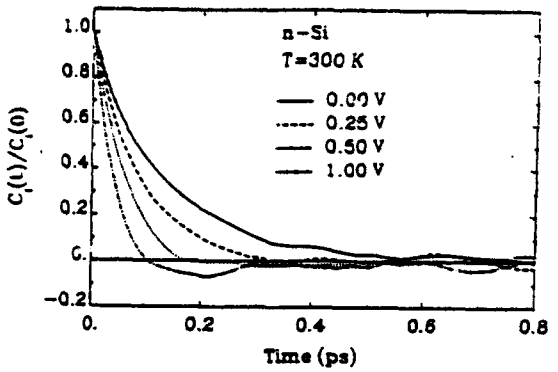


Fig. 1 - *Autocorrelation functions of current fluctuations for the different applied voltages reported. Calculations refers to a Si homogeneous structure with* $n = 10^{17}\ cm^{-3}$, $L = 0.6\mu m$ *at* $T = 300\ K$.

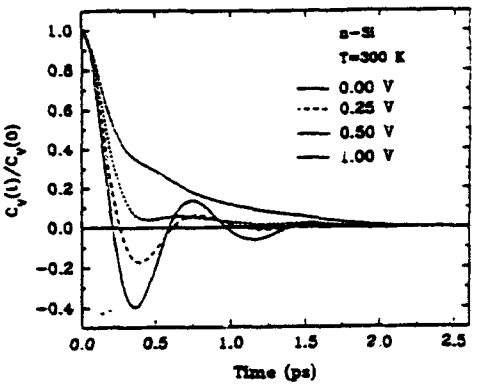


Fig. 2 - *Autocorrelation function of voltage fluctuations in the same structure and conditions as Fig. 1.*

where $\delta X(t) = X(t) - \overline{X}$ is the fluctuation of X around the average value $\overline{X}$, X being the current or the voltage. The problem is to provide a microscopic calculation of $C_X(t)$.

III. MICROSCOPIC CALCULATION

Under current noise operation the total current $I(t)$ as measured in the outside circuit is calculated as [3]:

$$I(t) = \frac{e}{L}\sum_{i=1}^{N_T(t)} v_i(t) = \frac{e}{L} N_T(t) v_d(t) \tag{3}$$

where e is the absolute value of the electronic charge, L the length of the sample, v_i the instantaneous value of the velocity component in the field direction of the i-th carrier, $N_T(t)$ the total number of carriers which are instantaneously present in the sample and $v_d(t) = [1/N_T(t)]\sum_{i=1}^{N_T(t)} v_i(t)$ is the drift-velocity.

Under voltage noise operation $I(t) = I_0$ and the time derivative of the voltage drop at the contacts $\Delta V(t) = [V(L,t) - V(0,t)]$ is calculated as [3]:

$$\frac{d}{dt}\Delta V(t) = \frac{L}{A\epsilon_0\epsilon_r}\left[\frac{e}{L}\sum_{i=1}^{N_T(t)} v_i(t) - I_0\right] \tag{4}$$

where A is the cross-sectional area of the sample, ϵ_0 the vacuum permittivity and ϵ_r the relative static dielectric constant of the background medium. The instantaneous voltage drop between the terminals can be obtained from a numerical integration of Eq. (4) over time.

In practice, the determination of $C_X(t)$ is performed from the knowledge of the time series $I(t)$ $[\Delta V(t)]$ as calculated from an ensemble MC simulation eventually coupled with a self-consistent Poisson solver, and taking appropriate boundary conditions concerning carrier injection-extraction from the contacts of the device. To this end, the total simulation, neglecting the initial transient, is recorded on a time-grid of step-size Δt. Then, by defining the time length in which the correlation function should be calculated as $m\Delta t$, with m integer, the correlation function is obtained as:

$$C_X(j\Delta t) = \overline{X(t')X(t'+j\Delta t)} = \frac{1}{M-m}\sum_{i=1}^{M-m} X(i\Delta t)X[(i+j)\Delta t] \tag{5}$$

with $j = 0,1,...,m$; $M > m$. Typical values are: $M = 50\ m$, $m = 100$. The corresponding $S_X(f)$ is determined by Fourier transformation.

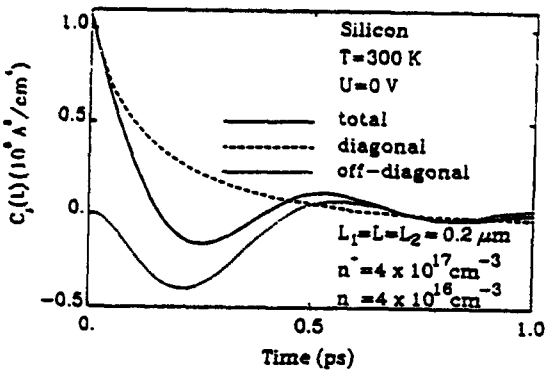


Fig. 3 - *Autocorrelation function of current-density fluctuations at equilibrium for a Si n^+nn^+ structure at $T = 300\ K$ with $n^+ = 4 \times 10^{17}\ cm^{-3}$, $n = 4 \times 10^{16}\ cm^{-3}$, and length $0.20 - 0.20 - 0.20\ \mu m$, respectively.*

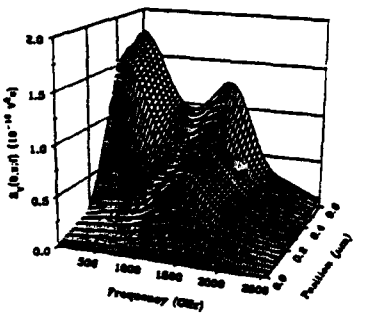


Fig. 4 - *Spectral density of voltage fluctuations as a function of frequency and position at equilibrium for a Si n^+nn^+ structure at $T = 300\ K$ with $n^+ = 10^{17}\ cm^{-3}$, $n = 10^{16}\ cm^{-3}$, and length $0.20 - 0.20 - 0.20\ \mu m$, respectively.*

IV. APPLICATIONS

In this Section we report the results obtained by the MC technique applied to different systems with increasing degree of complexity.

1. Resistor

The system we consider is a submicron Si resistor of length $L = 0.6\ \mu m$ with a donor concentration $n = 10^{17}\ cm^{-3}$ at 300 K. Figure 1 shows the current correlation function calculated at increasing applied voltages where its faster decay is associated with the onset of hot-carrier conditions. The presence of a negative part in $C_I(t)$ is attributed to the coupling between energy and velocity relaxation processes [3]. Figure 2 shows the voltage correlation functions $C_V(t)$ for the same resistor. At low voltages plasma and differential dielectric-relaxation times are responsible for the oscillatory and damping behaviors, respectively. At increasing applied voltages the sub-ohmic behavior of the current-voltage characteristics implies a significant increase of the dielectric relaxation time which, by becoming longer than the plasma time, washes-out the oscillations.

2. n^+nn^+ structure

The system we consider is a submicron Si n^+nn^+ structure at 300 K with two abrupt homojunctions. According to Ref. [3], the total correlation function can be decomposed as the sum of a diagonal and an off-diagonal contribution which are shown in Fig. 3. The former, describing the autocorrelation of the single particle-velocity, is responsible for the exponential decay. The latter, being associated with correlations due to the long-range Coulomb interaction, is responsible for an oscillatory behavior related to the plasma frequency of the n^+ and n regions.

Figure 4 shows the spectral density of the voltage fluctuations between $x = 0$ and the position x as function of x and frequency for the same structure of Fig. 3 but with $n = 10^{16}\ cm^{-3}$ and $n^+ = 10^{17}\ cm^{-3}$. Here, the different influence of each region in the structure is clearly emphasized. At low frequencies, most of the noise is originated in the n region due to its larger resistance. When going to higher frequencies, the contribution to the spectral density coming from the n region decreases, while that of the n^+ regions increases, reaching its maximum value near the associated plasma frequency (1275 GHz). At this frequency it can be clearly observed that the only contribution to the spectral density comes from the n^+ regions.

3. Schottky-Barrier diode

The system we consider is a one-dimensional GaAs $n^+ - n - metal$ structure at $300K$. The height of the barrier considered in the simulation is 0.735 V, which leads to an effective built-in voltage between the n region of the semiconductor and the metal of 0.640 V.

Figure 5 shows $S_I(f)$ at increasing applied voltages. The complexity of the spectrum is under-

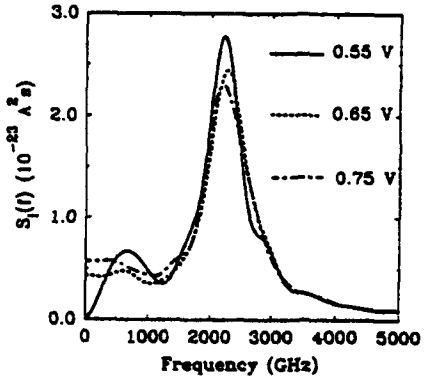


Fig. 5 - *Current spectral-density as a function of frequency for a GaAs Schottky barrier diode at* $T = 300\ K$ *with* $n = 10^{16}\ cm^{-3}$, $n^+ = 10^{17}\ cm^{-3}$ *and length of each region of* $0.35\ \mu m$.

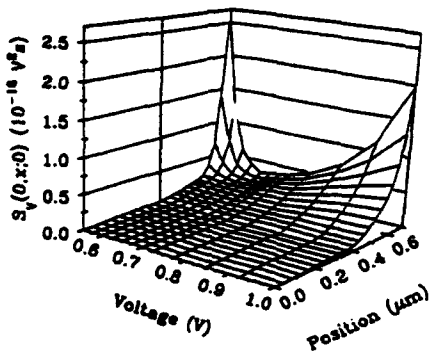


Fig. 6 - *Low-frequency value of the spectral density of voltage fluctuations as a function of position and mean voltage in the same structure as Fig. 5. The semiconductor-metal contact is at* $x = 0.7\ \mu m$.

stood on the basis of a strong coupling between fluctuations in carrier velocity and the self-consistent electric field. Two peaks are observed, one in the region below $10^3\ GHz$ and another at about $2.2 \times 10^3\ GHz$. The first is attributed to carriers that have insufficient kinetic energy to surpass the barrier and return to the neutral semiconductor region. The second originates from the coupling between fluctuations in carrier velocity and in the self-consistent field due to the inhomogeneity introduced by the $n^+ - n$ homojunction as discussed in Section 2.

Figure 6 shows a spatial analysis of the low-frequency value of the voltage spectral density. For voltages lower than $0.640\ V$ shot-noise is dominant [2], and most of the noise arises in the depletion region close to the barrier. At increasing voltages, thermal noise associated with the series resistance prevails, and the noise becomes spatially more distributed, mainly originating from the n region of the device. Finally, at the highest voltages, the presence of hot carriers and intervalley mechanisms in the n region leads to the appearance of an excess noise.

V. CONCLUSIONS

We have presented a theoretical simulation of electronic noise in semiconductor materials and two-terminal devices. Calculations are based on the Monte Carlo technique which, to include fluctuations of the self-consistent electric field, is coupled with a Poisson solver. The current representation, by allowing a decomposition in terms of different noise contributions, is found to provide useful information on the nature of the noise sources. The voltage representation, by allowing a spatial analysis to be carried out, is found to provide a local information on the strength of the noise sources. We believe that the generality of the approach here proposed, besides providing a rigorous basis for the interpretation of noise-spectroscopy measurements, still leaves wide possibilities of implementation for the analysis of more complicated systems.

ACKNOWLEDGMENTS

This work has been performed within the European Laboratory for Electronic Noise (ELEN) supported by the Commission of European Community through the contracts EKBXCT920047 and ERBCHBICT920162. Partial support from the SA-14/14/92 project by the Consejería de Cultura de la Junta de Castilla y León and by the Italian Consiglio Nazionale delle Ricerche (CNR) is gratefully acknowledged.

REFERENCES

[1] Proceedings of the 12^{th} *International Conference on Noise in Physical Systems and 1/f Fluctuations*, edited by P. H. Handel and A. L. Chung (AIP, New York, 1993).
[2] A. Van der Ziel, *"Noise in Solid State Devices and Circuits"* (Wiley, New York, 1986).
[3] L. Varani and L. Reggiani, Rivista Nuovo Cimento, in press.

TIME-STEP STABILITY FOR SELF-CONSISTENT MONTE CARLO DEVICE SIMULATION

P. W. Rambo and J. Denavit
*University of California, Lawrence Livermore National Laboratory**
P. O. Box 808, Livermore, CA 94551

ABSTRACT

An important numerical constraint on self consistent Monte Carlo device simulation is the stability limit on the time step imposed by plasma oscillations. The widely quoted stability limit for the time step between Poisson field solutions, $\Delta t<2/\omega_p$ where ω_p is the plasma frequency, is specific to the leapfrog particle advance used in collisionless plasma simulation and does not apply to typical particle advance schemes used for device simulation. We present a stability criterion applicable to several algorithms in use for solid state modeling; this criterion is verified with numerical simulation. This work clarifies the time step limitation due to plasma oscillations and provides a useful guide for the efficient choice of time step size in Monte Carlo simulation. Because frequent solution of the Poisson equation can be a sizable computational burden, methods for allowing larger time step are desirable. The use of advanced time levels to allow stability with $\omega_p\Delta t>>1$ is well known in the simulation of collisionless plasmas; we have adapted these implicit methods to semiconductor modeling and demonstrated stable simulation for time steps larger than the explicit limit.

I. TIME STEP STABILITY

One important constraint on self consistent simulations of both solid state devices and plasmas is numerical stability of plasma waves. This limitation imposes a maximum on the allowed time step interval between Poisson field solutions Δt, relative to the plasma frequency ω_p, and is particularly important for simulations of devices with high carrier concentrations, such as found in heavily doped contact regions. Motivated by analysis of numerical schemes for plasma simulation, many authors have quoted the stability limit $\omega_p\Delta t<2$ [1]. This limit of $\omega_p\Delta t<2$, however, is specific to the leapfrog particle advance used in plasma simulation and is generally not applicable to algorithms used for solid state device simulation. In contrast to the leapfrog algorithm which is centered and advances the particles with a fixed time increment equal to the time step between field solves, Δt, solid state simulations typically use non-centered algorithms with a particle time step $\delta t\neq\Delta t$. Furthermore the particle time step is often picked stochastically based on mean free collision times determined not only by physical parameters but also by details of the numerical implementation such as self-scattering.

The numerical stability of typical algorithms used for Monte Carlo device simulation has been investigated in Ref. [2]. This analysis is applied to a variety of algorithms in different regimes; here we confine our discussion to the case of $\delta t<<\Delta t$ which is generally applicable to any scheme which only uses the electric field at the old time level in advancing the particles. This situation could correspond to the case of high collision rate (perhaps due to a large self scatter rate), or simply an attempt to ensure very accurate particle orbits. In this limit the particle advance approximates an exact orbit. Then the numerical solution corresponds to solving the Boltzmann equation exactly between times t^n and $t^{n+1}=t^n+\Delta t$ with the time independent electric field, $E(t^n)$. A dispersion relation is obtained, which shows the surprising result that in the absence of collisions, instability occurs for all time steps. In practice, collisions allow stable simulation for finite time step. The appropriate collision rate is the rate of momentum transfer ν_c, defined by the first velocity moment over the Boltzmann collision operator, and may be related to the mobility by $\nu_c=e/\mu m^*$ with e and m^* the electron charge and effective mass . The amount of collisionality required to offset the tendancy for growth is determined by the threshold for stability (zero growth, $\gamma=0$),

$$\left(\frac{\nu_c}{\omega_p}\right)_{\gamma=0} = \left(\frac{1-e^{-\alpha}-\alpha e^{-\alpha}}{1-e^{-\alpha}}\right)^{1/2}, \text{ for } \alpha \equiv \nu_c \Delta t \le \alpha_c$$

$$= \left(\frac{\alpha}{2}\frac{1-e^{-\alpha}}{1+e^{-\alpha}}\right)^{1/2}, \text{ for } \alpha \equiv \nu_c \Delta t \ge \alpha_c \qquad (1)$$

and plotted in Fig. 1. Stable solutions lie above the stability threshold shown as a solid line in Fig. 1. For values of the collision rate below this threshold, unstable growth is present. α_c=3.72 corresponds to the point *C* in Fig. 1 where the threshold crosses the boundary between complex and real roots denoted by a dotted line. An approximate expression for the stability limit, valid for $\gamma\Delta t \ll 1$ and $\nu_c \Delta t \ll 1$, is given by,

$$\omega_p \Delta t \le 2\,\nu_c / \omega_p \qquad (2)$$

This condition is plotted as a straight dashed line in Fig. 1 and comparison with the exact threshold (solid line) shows that it remains a good approximation out to values of $\omega_p\Delta t$ approaching unity.

A number of simulations have been performed to explore the stability boundary in the space of ν_c/ω_p vs. $\omega_p\Delta t$ shown in solid line in Fig. 1. The code used allows multiple nonparabolic, elliptic bands and scattering processes appropriate for simulation of GaAs or Si. The particle advance is performed as described by Hockney and Eastwood with δt picked randomly based upon the total scattering rate Γ which includes self scattering. A grid with uniform spacing Δx is used; interpolation from the particles to the grid uses standard linear weighting, and the Poisson equation is solved directly without spatial smoothing. Results from simple simulations which closely confo~n to the analysis are shown in Fig. 1. For these runs a single spherical parabolic band is used, and collisions correspond to elastic, isotropic scattering which is independent of energy. Simulation results are plotted as solid markers if unstable growth is observed, and as open markers if the run was observed to be stable ($\gamma\Delta t$<0.01). The circles are from simulations with $\Gamma=\nu_c$, while triangles represent simulations with $\Gamma=5\nu_c$. It can be seen that the stability condition implied by the boundary between solid and open markers is in reasonable agreement with the analysis, but

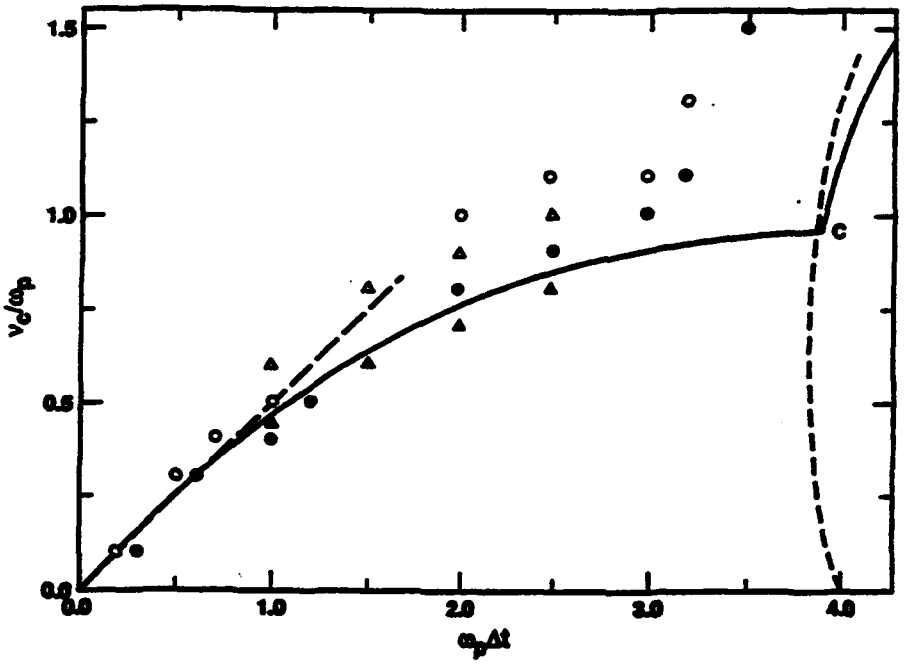


Fig. 1. Numerical stability as a function of collisionality, ν_c/ω_p, and time step between Poisson solves, $\omega_p\Delta t$. Markers represent results of simulations: open for stable, closed for unstable.

indicates a slightly more stringent stability condition; this is due to the effects of finite temperature which are neglected in the analysis.

As an example relevant to realistic device simulations, consider GaAs at a doping density of $N_D=1.0\times10^{17}$ cm^{-3}; assuming the electron density is equal to the doping density we have $\omega_p=1.0\times10^{13}$ s^{-1}. At a lattice temperature of T=300 K, the low field mobility is μ=5.3×10^3 cm^2/V·s; this corresponds to an effective collision frequency $\nu_c\approx5.0\times10^{12}$ s^{-1}. Then $\nu_c/\omega_p\approx0.25$, and the stable time step limit is predicted to be $\omega_p\Delta t\approx0.5$ or $\Delta t\approx2.6\times10^{-14}$ s. At a lattice temperature T=77 K, $\mu\approx9.2\times10^3$ cm^2/V·s corresponding to $\nu_c\approx2.9\times10^{12}$ s^{-1}. Then $\nu_c/\omega_p\approx0.15$ and the stability limit is $\omega_p\Delta t\approx0.3$ ($\Delta t\approx1.5\times10^{-14}$ s). Simulations of the two cases described above (T=300 K, $\omega_p\Delta t$=0.5; T=77 K, $\omega_p\Delta t$=0.3) performed using realistic models for GaAs show weak instability; stability requires somewhat smaller values of the time step consistent with the effect of finite pressure. Unstable runs were observed to saturate by heating the electrons. In some cases, mobilities were noticably reduced and significant numbers of electrons promoted to the upper valleys.

II. LARGE TIME STEP ALGORITHM

Frequent solution of the Poisson equation to resolve plasma oscillations can be a sizable computational burden. Caution suggests that the condition presented above not be approached to closely, since finite pressure effects slightly lower the stability limit. Additionally, near the stability limit unphysical heating of the carriers may be a more insidious effect than the catastrophic instability which occurs well above the limit. The necessity of using advanced time levels for numerical stability with $\omega_p\Delta t>>1$ has long been known in the case of collisionless plasma simulation, and stable large time step simulations have been achieved using time-implicit methods [3].

The key ingredient for large time step stability is to advance the particles using the *advanced* electric field E^{n+1}, such as

$$x^{n+1}=x_0+\beta\Delta t^2\frac{qE^{n+1}}{m^*}, \tag{3}$$

where x_0 depends only on quantities at the past time level t^n and β is the implicitness parameter. Because the new field depends on the new particle positions through the solution to the Poisson equation, however, an implicit solution for the electric field is required. The implicit field equation may be found by writing the Poisson equation at the new time level, and linearizing the charge density with respect to perturbations due to the advanced field,

$$\frac{\partial E^{n+1}}{\partial x}=\frac{4\pi}{\varepsilon}\rho^{n+1}\equiv\frac{4\pi}{\varepsilon}\{\rho_0(x_0)+\delta\rho\}. \tag{4}$$

The perturbation to the charge density, $\delta\rho$, may be expressed in terms of the perturbation to the particle position δx by,

$$\delta\rho=-\frac{\partial}{\partial x}(\rho_0\delta x)=-\frac{\partial}{\partial x}(\rho_0\frac{q}{m^*}\beta\Delta t^2E^{n+1}). \tag{5}$$

Substituting into Eq. (5), and rearranging, the field equation becomes

$$\frac{\partial}{\partial x}\left\{(1+\chi)\frac{\partial\phi}{\partial x}\right\}=-\frac{4\pi}{\varepsilon}\rho_0\quad,\quad\chi\equiv\frac{4\pi}{\varepsilon}\rho_0\frac{q}{m^*}\beta\Delta t^2=\beta(\omega_p\Delta t)^2, \tag{6}$$

with χ the effective susceptibility due to the partial advance of the particles to x_0. Strict implementation of such a scheme requires writing these equations with the spatial derivatives replaced by finite differences generalized to include the interpolation between the grid and particles. This leads to a matrix system for the new electric field which is completely consistent with the particle push but has a larger stencil than the original explicit system. Simplified differencing (and reduced computational stencil) can be obtained by simply writing Eq. (2) in finite difference form; this is appropriate if $\omega_p\Delta t$ is not too large.

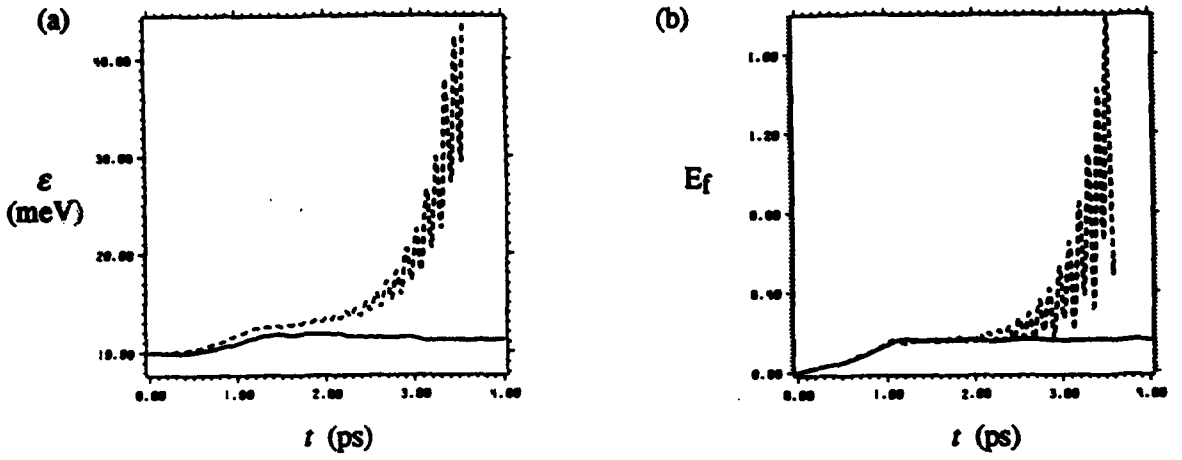


Fig. 2. Time history of (a) average particle energy and (b) electrostatic field energy (arbitrary units) from two diode calculations: explicit $\omega_p\Delta t$=0.42 (dashed line) and implicit $\omega_p\Delta t$=2.24 (solid line).

We have adapted these implicit plasma techniques to semiconductor modeling and demonstrated stable simulation for $\omega_p\Delta t$ larger than the limit given in Section I. At each time step, particles which undergo one or more collisions ($\delta t<\Delta t$) are advanced explicitly. These collisional particles contribute only to ρ_0 and not to the susceptibilities. Particles which do not undergo a collision ($\delta t\geq\Delta t$) are advanced implicitly by performing a partial push and accumulating the necessary susceptibilities. Then the implicit field equation is solved and the positions of the implicit particles corrected, completing the time step. Calculations of a submicron GaAs diode [4] have been performed as a realistic test. The diode is composed of a 0.25 μm undoped active layer between 0.35 μm n^+-layers doped at density N_D=2.0×10^{17} cm^{-3} (ω_p≈2.7×10^{13} s^{-1}). The simulation model is substantially the same as used by Tomizawa et al., and for their time step, Δt=1.0×10^{-14} s, similar results are obtained. Figure 2 shows time histories of the average particle energy and electrostatic energy from simulations with an applied voltage of 0.25 volts (rising from zero in 1.0 ps) and lattice temperature of 77 K. An explicit calculation with Δt=1.5×10^{-14} s is unstable (as predicted above for $\omega_p\Delta t$≈0.42), while in contrast, the implicit calculation (β=0.75) with Δt=8.0×10^{-14} s ($\omega_p\Delta t$≈2.24) is well behaved. The increase in time step which can be realized is limited, because as Δt is increased, the fraction of particles which are treated implicitly decreases. As $\Gamma\Delta t$ becomes of order unity, most of the particles are treated explicitly, and the stability limit of Eq. (1) becomes effective. Although the time savings in one dimension is modest, appreciable gains in multi-dimensional simulation might be realized because of the increased computational burden of the Poisson solve.

* This work was supported by the U. S. Department of Energy at Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

[1] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*. New York: Adam Hilger, 1988.

[2] P. W. Rambo and J. Denavit, "Time Stability of Monte Carlo Device Simulation," IEEE Trans. on Computer-Aided Design 12, 1734 (1993).

[3] J. Denavit, "Time Filtering Particle Simulations with $\omega_p\Delta t$>>1," J. Comput. Phys. 42, 337 (1981); A. B. Langdon, B. I. Cohen, and A. Friedman, "Direct Implicit Large Time-Step Particle Simulation of Plasmas," J. Comput. Phys. 51, 107 (1983).

[4] K. Tomizawa, Y. Awano, N. Hashizume, and M. Kawashima, "GaAs n^+-$i(n)$-n^+ diode," IEE Proc. 129, 131 (1982).

MONTE CARLO STUDIES OF HOT ELECTRON GENERATION IN SCALED MOSFETs

A. Duncan, C.H. Lee and U. Ravaioli
Beckman Institute and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

Abstract

In this work, we report on a study of submicron MOSFET structures, performed with a full band Monte Carlo simulator. In order to obtain a systematic understanding of the hot electron effects in scaled structures, we investigate a series of devices with scaled geometry, both with constant and scaled bias.

I. INTRODUCTION

As the dimensions of integrated devices continue to shrink, investigation of hot electron effects becomes increasingly important to assess the influence of overshoot phenomena and reliability problems, like breakdown due to impact ionization, defect generation, and injection into the gate oxide. In strucures used for flash memory applications, it is important to control or even enhance the hot electron population. In order to investigate these effects in detail, it is necessary to introduce knowledge of the bandstructure in the model, because at the high electron energies involved, simple models of the band are inaccurate. Because of the massive computational resources needed [1,2], full band Monte Carlo applications for complete device structures have been possible only in relatively recent times, beginning with the development of the simulator DAMOCLES [3,4] at IBM, Yorktown Heights. In this work we report on a study of submicron MOSFET structures performed with a full band Monte Carlo simulator which includes the first two branches of the silicon conduction band.

II. MODEL

Knowledge of the bandstructure is necessary to accurately calculate the electron trajectories in real and in momentum space and to determine the density of states and therefore the scattering rates at high electron energies. Large tables store the information used to obtain the electron velocity and to relate energy and momentum for the determination of the final state after scattering. The inclusion of the band structure causes considerable numerical and memory overhead, but due to improvements in the solution techniques and to increased computational power, full band calculations are now possible on modern workstations.

In the simulator for this study, the band structure for Si is calculated using the empirical pseudopotential model of Cohen and Bergstresser [5]. All the necessary information is stored for the k points inside the irreducible wedge of the Brillouin zone and is mapped to all

the points of momentum space by using symmetry properties. The silicon model includes intravalley acoustic phonon scattering, F and G type X-X intervalley phonon scattering, X-L intervalley phonon scattering, ionized impurity scattering, and impact ionization. Ridley's statistical screening is used in the ionized impurity scattering calculation [6], and Bude's model for impact ionization is used [7]. The total scattering rate is adjusted so that at high energies, it follows the total density of states as implemented in [8].

The random flight times for the electron trajectories are generated using the vectorized ensemble constant time technique [9]. At the beginning of every iteration, the electron density in real space is evaluated using a 2-D cloud-in-cell scheme [10], and Poisson's equation is solved numerically using a simple vectorized relaxation scheme. Holes are included in the constant quasi-Fermi level approximation [3]. The boundary layer in the contacts are kept neutral by injecting the necessary number of electrons.

III. SIMULATION RESULTS

Both constant and variable bias scaling were applied to the test structure shown in Figure 1. Figure 2 shows the effect of constant bias scaling on the energy distribution at the drain/channel junction for V_{gs}= 2.5 V and V_{ds}= 3.0 V. The doping is increased and all device dimensions within the silicon are decreased by a constant factor as the device is scaled. The oxide thickness is decreased by the square root of this factor. Figure 3 shows the energy distribution when the bias is scaled with the square root of the channel length.

IV. CONCLUSION

Full band Monte Carlo is a valuable tool for studying high energy effects in scaled MOSFETs. We have demonstrated how the method can be used to generate the electron energy distribution for such devices. Analysis of the energy distribution can serve as a guideline to determine scaling rules and to assess the necessary level of statistical enhancement to study the energy tails. Future work will include the simulation of more realistic MOSFET structures and the calculation of gate and substrate currents using a stratification technique to enhance the high energy tails [11].

ACKNOWLEDGEMENT

This work was supported by the Computational Scinece Graduate Fellowship Program of the Office of Scientific Computing in the Department of Energy and by the Joint Services Electronics Program, grant N00014-90-J-1270.

References

[1] K. Hess, ed., *Monte Carlo Device Simulation: Full Band and Beyond* (Kluwer Academic Publishers, Norwell, MA, 1991).

[2] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer-Verlag, New York, 1989).

[3] M.V. Fischetti and S.E. Laux,Phys. Rev. B **38**, 9721 (1988).

[4] S.E. Laux, M.V. Fischetti and D.J. Frank, IBM J. Res. Develop. **34**, 466 (1990).

[5] M.L. Cohen and T.K. Bergstresser, Phys. Rev. **141**, 789 (1966).

[6] B.K. Ridley, J. Phys. C **10**, 1589 (1977).

[7] J. Bude, Ph.D. Dissertation, University of Illinois Urbana-Champaign, 1992.

[8] J.Y. Tang, Ph.D. Dissertation, University of Illinois Urbana-Champaign, 1983.

[9] U. Ravaioli, in *Monte Carlo Device Simulation: Full Band and Beyond*, edited by K. Hess (Kluwer Academic Publishers, Norwell, MA, 1991), pp. 267-284.

[10] R.W. Hockney and J.R. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981).

[11] C.H. Lee, U. Ravaioli, K. Hess, C. Mead, and P. Hassler, "Simulation of MOS Floating Gate Memory Device with a Full Bandstructure Monte Carlo Approach," submitted for publication.

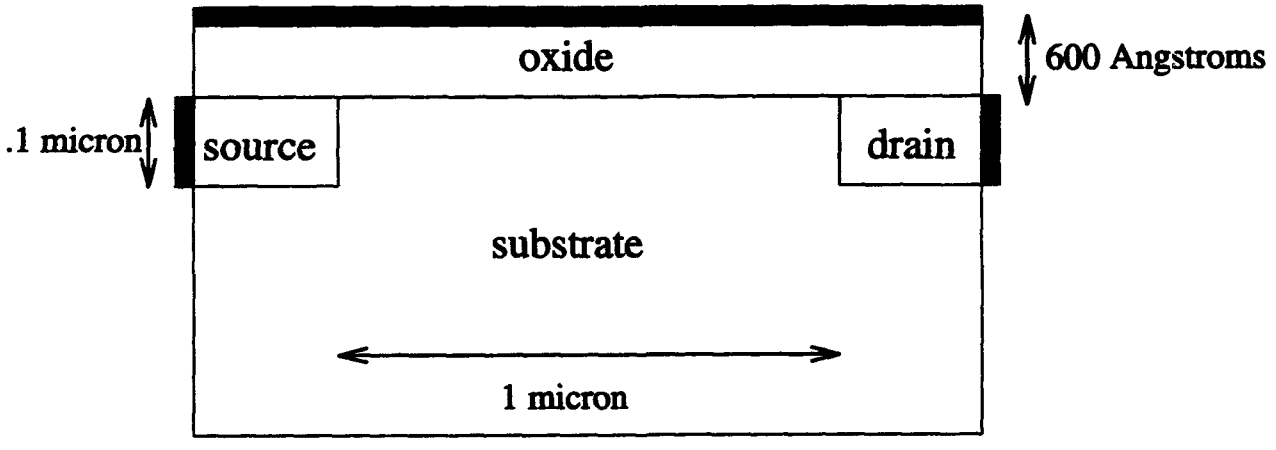


FIG 1.Schematic of device that is scaled by variable and constant bias scaling. Drain and source doping is $N_D = 10^{19} cm^{-3}$ and substrate doping is $N_A = 10^{16} cm^{-3}$. $V_{sub} = V_s = 0$ V.

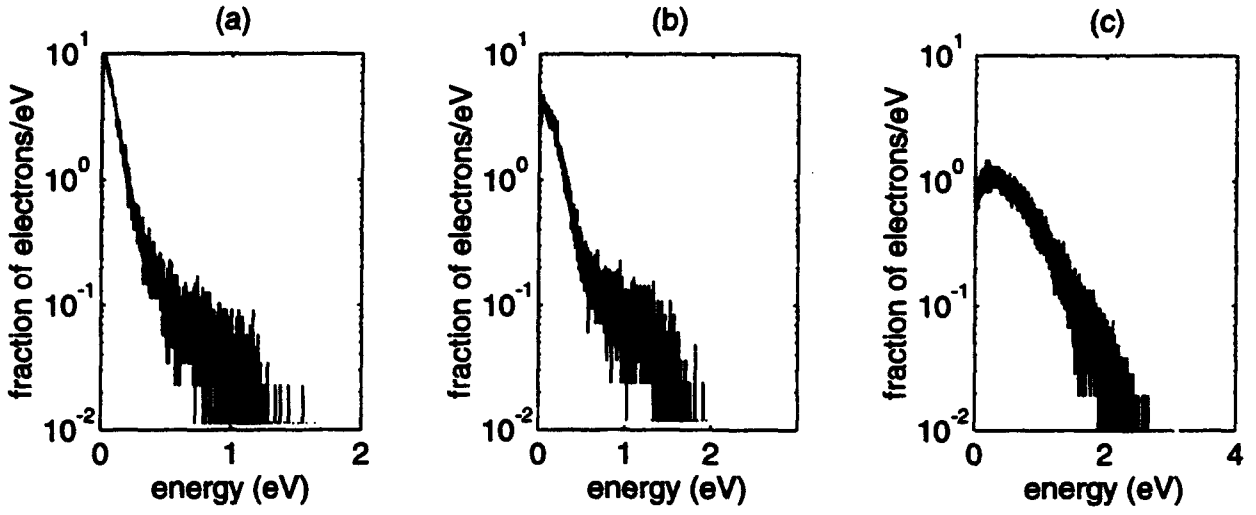


FIG 2. Energy distribution at drain/channel junction for devices with channel lengths of (a) 1μm, (b) $.5\mu$m, and (c) $.25\mu$m with constant biases of $V_{gs} = 2.5$ V and $V_{ds} = 3.0$ V

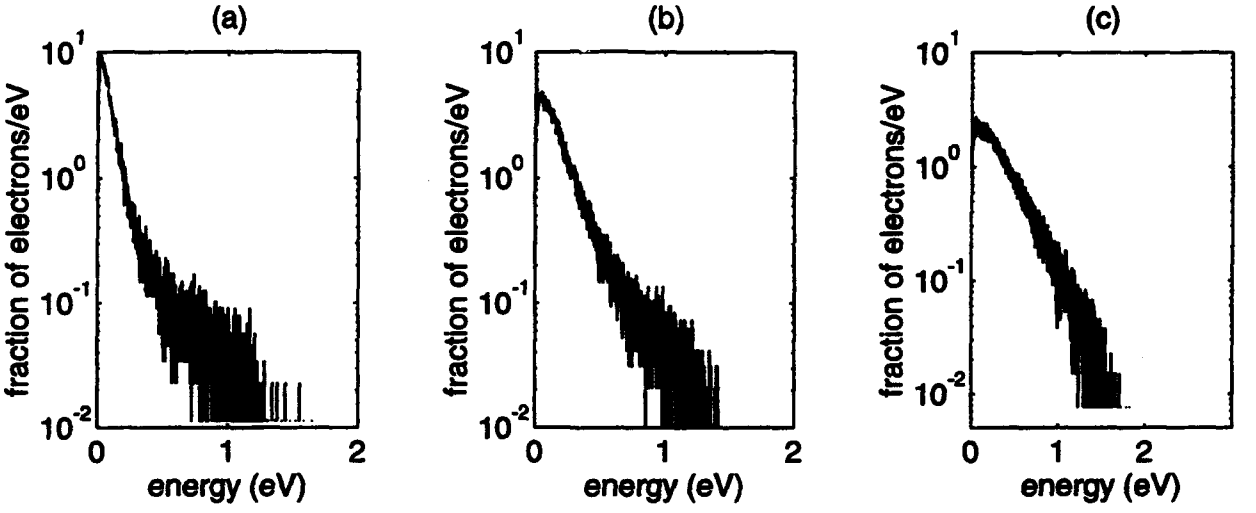


FIG 3. Energy distribution at drain/channel junction for devices with channel lengths of (a) 1μm, (b) $.5\mu$m, and (c) $.25\mu$m with scaled biases.

A REJECTION METHOD FOR SELECTION OF SCATTERED STATES

William S Lawson
*University of California, Lawrence Livermore National Laboratory**
P. O. Box 808, Livermore, CA 94551

ABSTRACT

A rejection method is presented that sidesteps much of the labor necessary in the usual techniques for choosing a scattered state after an electron-phonon collision with full band structure. The phonon wave number is chosen randomly, then tested to see if the resultant collision will satisfy energy conservation to within some accuracy. If not, the collision is rejected, and if so, then the wave number is adjusted in order to enforce energy conservation more precisely. The price one pays is in a high rejection rate. If the cost of a rejection is small, however, this rejection rate can be tolerated. This method will not compete with analytical models (near valley minima), but may outperform the more usual techniques. Accuracies of a few percent are practical.

I. INTRODUCTION

In Monte Carlo simulations, a time-consuming problem is the calculation of the phonon scattering rate for a given particle, and the selection of the scattered state given that a scatter has occurred. Typically, these two problems are dealt with separately. The scattering rate is usually tabulated in some fashion, and the scattered state is found randomly taking into account the surface within the Brillouin zone on which energy is conserved, and the density of states at each point on that surface.

One advance made early in deciding when to scatter a particle was the invention of the Null-Scatter, or Rejection method. In this method the scattering rate is assumed to be some simple function of the wave number (usually a step function) that is always larger than the true scattering rate. The true scattering rate *at the time of the scatter* can then be applied by rejecting some fraction of the intended scatters, *i.e.*, not scattering the particle. This technique has become almost universal.

Choosing a scattered state for a particle is a more difficult task, since momentum and energy must be conserved, implying that the scattered particle must lie on a complex surface within the Brillouin zone. Furthermore, the local density of states varies over that surface, and the scattering rate is proportional to the density of states. At least two groups have addressed this problem with reasonably accurate band structures. My interpretation of their techniques is the following: Fischetti and Laux [1], divide the Brillouin zone into small cubical cells, locate all the cells through which the energy-conserving surface passes, weight each cell according to the density of states, the scattering rate and the area of the energy-conserving surface contained within the cell, then choose a cell randomly according to these weights. The scattered state is chosen from the plane within the cell that approximates the energy-conserving surface. Yoder *et al.* [2], also choose such cubical cells randomly, and by a similar method. They then choose the scattered state by a rejection method: states are chosen randomly within the cell with a *uniform* distribution; if energy is not conserved within a given tolerance, the state is rejected and another state chosen at random until a suitable state is found. The present technique shares some features with that of Yoder *et al.*, but dispenses with the cells.

II. THE METHOD

The scattering rate from wave number $\mathbf{k}$ to wave number $\mathbf{k}'$ can be written in the form

$$S(\mathbf{k},\mathbf{k}') = f(\mathbf{k},\mathbf{k}')g(\mathbf{p})\delta(E - E' \pm E_p)$$

where $\mathbf{p} = \pm(\mathbf{k}' - \mathbf{k})$, is the phonon wave number, E is the initial particle energy, E′ is the final particle energy, and E_p is the phonon energy. Usually f does not vary strongly with $\mathbf{k}$ and $\mathbf{k}'$, and can be approximated by a constant — if this is not good enough, the rejection method can be used to correct the scattering rate. In the present work it is assumed that f is constant. For Polar optical phonons (most important for Gallium Arsenide), $g(\mathbf{p}) \propto 1/p^2$ for small p.

The basis of the present method is to approximate the scattering rate as

$$S(\mathbf{k},\mathbf{k}') \approx f(\mathbf{k},\mathbf{k}')g(\mathbf{p}) \cdot \frac{1}{2\Delta E} H(\Delta E - |E - E' \pm E_p|)$$

where H is the Heavyside step function, and ΔE can be virtually anything, as long as it is small enough to give an accurate measure of the local density of states. States are chosen at random over some domain known to encompass the energy-conserving surface (allowing for the energy mismatch ΔE), and those within the energy tolerance ΔE are accepted, while those that do not are rejected, and treated as null scatters. States that are accepted are then moved onto the energy-conserving surface using a second-order Newton's method (using both the first and second derivatives).

This technique has the advantages that no energies, densities of states, or surface areas must be computed and tabulated within cubical cells, nor need the actual scattering rate be calculated. Furthermore, the method is not tied to a mesh, and so its speed does not suffer when meshes are refined. It has the obvious disadvantage that a high rate of rejection is likely, so that fast evaluation of the energy is critical to minimize the time spent on rejections. (A mesh *is* used in the evaluation of the band energy.)

The accuracy of the method is limited by the value of ΔE. If one is not close to a critical energy (a valley, saddle point or maximum), the density of states is smooth, and it is easy to see that the error in the effective density of states will be of order ΔE^2. If one is close to a critical energy, one must either be sure that the resulting inaccuracy is of little consequence (as for maxima and saddle points, or when $\Delta E << kT$), or ensure that ΔE is small enough. For a parabolic ellipsoidal valley, one can show that as long as E – ΔE is above the energy of the valley minimum, the density of states is reasonably accurate. This fortuitous fact makes it possible to use this method near valleys, but since an analytic technique can be used near any but the X valleys in GaAs (which are not close to parabolic or ellipsoidal), it would probably not be the method of choice. Nonetheless, it has been implemented here.

Several techniques can be used to increase the speed of the method. For low energies one can reduce the k-space volume to be sampled to the regions around the valleys. This is especially easy when the scattering is independent of p, since then choosing p uniformly is equivalent to choosing $\mathbf{k}'$ uniformly, and any shape volume in which a wave number can easily be chosen with uniform probability density can be used; in particular, ellipsoids bounding the valleys can be used, eliminating most of the volume of the Brillouin zone. In this case, acceptance rates of 20%-50% can be achieved near valleys, with ΔE proportional to the energy above the valley minimum. At higher energies (sampling from the entire Brillouin zone), with ΔE = 0.2 eV, acceptance rates of 10%-20% are typical (this is about as large as one can reasonably make ΔE).

When g is not constant, as in polar optical phonons, it is necessary to work in the phonon Brillouin zone. A fair amount of work is necessary to locate and surround the valleys (with a conical section of a spherical shell, in this case), making the method unattractive for low-energy scattering when analytical methods are applicable. This technique was used here as a proof-of-principle, regardless of its attractiveness. Another difficulty in the polar optical case is that for small p, g changes so rapidly that one must limit ΔE to less than about 1/5 of the phonon energy for reasonable accuracy. This would be a fatal restriction if it were necessary for ΔE to be independent of p. By making ΔE a function of p (in this work $\Delta E \propto \sqrt{p}$ for intra-valley and full-band scattering, constant for inter-valley scattering), ΔE can be made small only for small p. The result of this technique is an overall acceptance rate of between 4% and 6%, with the high-energy acceptance rate being on the low end.

The limits on the accuracy of this method might seem to be a drawback, but a fundamental limit exists in the accuracy of the semi-classical model. The source of this limit is in the application of Fermi's Golden Rule in computing the scattering rate. The Golden Rule assumes that a long period of time is available during which the interaction (a scatter in this case) may occur. In fact, the mean

scattering time is itself a limit on the time available for interaction. Heisenberg's uncertainty principle is applicable here, and indicates that the selection rules for momentum conservation and energy conservation will only hold to within limits determined by the mean free path, and the scattering time (and phonon life-time) respectively, according to $\Delta k \geq 1/2\lambda$ and $\Delta E \geq \hbar/2\tau$. This should not be interpreted as non-conservation of energy, of course, but as an indeterminacy in the band energy. For most semiconductors at high fields, the scattering rate implies an uncertainty in the carrier energy of almost 0.1 eV; of course, at low fields the scattering rates are much lower, and the bands are accordingly very narrow. This implies that the accuracy of the semi-classical model is probably limited to a few percent at high fields, and it would be pointless to attempt to reduce simulation errors below this level.

Once a scattered state has been accepted, it must be brought to the energy-conserving surface. To do this quickly requires at least the first derivatives of the energy with respect to the wave number, and preferably the second derivatives. In the present work a cubic interpolation scheme was used to compute the energy as a function of wave number. This scheme required most of those derivatives, and could interpolate them as well. The accuracy of interpolated derivatives used in Newton's method was of some concern, but the scheme has worked extremely well, with only one or two iterations necessary in most cases. If four iterations do not suffice, the state is discarded (this is very rare). One must also specify a direction in which the wave number will be moved in order to bring it to the energy-conserving surface. In general, this direction was taken to be the gradient *at each iteration*, but when the wave number was in a valley, it was constrained to move in a line through the center of the valley.

III. RESULTS

Simulations were performed with the first conduction band of Gallium Arsenide. The band structure was calculated using the pseudopotential method with the form factors of Cohen and Bergstresser [3]. Ten thousand particles were advanced in momentum, but not in space, with a uniform and constant electric field. Only polar optical phonon scattering was allowed, but both absorption and emission were included. The phonon temperature was 0.026 eV, and all phonons were assumed to have an energy of 0.0355 eV. No attempt was made to properly scale the time, electric field, velocity or scattering constant, but the time scale is a consequence of the electric field and the scattering constant, and the velocity is purely a diagnostic, so no arbitrary physical parameters are introduced. The simulation was run on a CRAY-YMP, but was not vectorized (much vectorization should be possible, but was avoided for simplicity).

The drift velocity versus electric field in the 100 direction is shown in Figure 1 (arbitrary units, lines delimit rough statistical error bounds). Acoustic phonon scattering is important for scattering at moderately large energies, so the saturation level should not be taken too seriously, but negative differential resistivity is clearly exhibited.

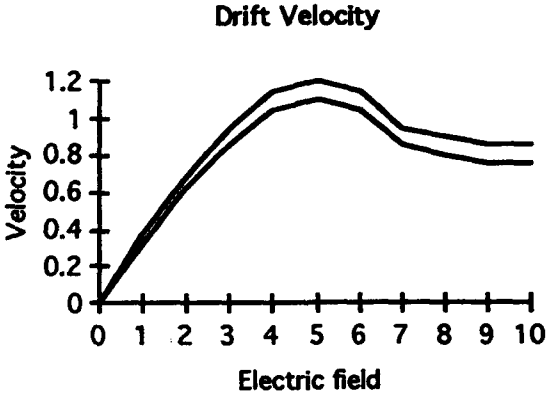


Fig. 1. Drift velocity versus field, polar optical scattering only.

Numerous tests were performed to ensure that the various parts of the test code were performing properly. These included scattering of an ensemble of particles from a single initial state, scatter plots of randomly chosen states from limited regions of the Brillouin zone and zero-field runs at various temperatures, as well as scatter plots of the actual simulation runs. These tests cannot guarantee that the code contains no errors, but serious error is unlikely and no potential errors are likely to affect the timing, which is of the most interest in assessing the usefulness of the scheme.

At low electric fields (with few particles outside the Γ valley) the mean computer time per actual scatter, including the cost of null scatters, is 1.6 ms. At higher electric fields (in the saturated regime), the mean time per scatter is 2.6 ms. The time required for a rejection outside the valleys is 70 μs, while an acceptance requires 140 μs. Within the valleys, a rejection requires 105 μs (an acceptance 175 μs), which is more than compensated by the higher acceptance rate for valley scatters. An evaluation of the electron energy requires 15 μs, and an evaluation of the energy with all its first and second derivatives requires 30 μs. For reference, a scalar multiply on the CRAY YMP requires 50 ns. All these times are for no vectorization at all. Obviously vectorization would improve the performance.

IV. SUMMARY

The rejection scheme for selection of scattered states presented here has been successfully implemented, and may be competitive when analytic methods are not applicable. When the scattering rate does not depend directly on the phonon momentum, the scheme is reasonably efficient. The scheme has a higher rejection rate when the scattering is not uniform in the phonon momentum. The high rejection rate may be compensated by the computational simplicity of the method and the relative speed of rejections.

The scheme can be extended to a complete and realistic scattering model, given the necessary data on band structures, overlap integrals and so on. The possible techniques for increasing the acceptance rate and improving the speed of rejections have not been exhausted.

* This work was supported by the U. S. Department of Energy at Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

REFERENCES

[1] M.V. Fischetti and S.E. Laux, "Monte Carlo Analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", in Physical Review B 38, 9721 (1988).

[2] P.D. Yoder, J.M. Higman, J. Bude and K. Hess, "Monte Carlo simulation of hot electron transport in Si using a unified pseudopotential description of the crystal", in Proceedings of Seventh International Conference on Hot Carriers in Semiconductors, Nara, Japan 1991, C. Hamaguchi and M. Inoue Editors (IOP Publishing Ltd. — Adam-Hilger, NY)

[3] M.L. Cohen and T.K. Bergstresser, "Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures", in Physical Review 141, 789 (1966).

NEW HIGHLY EFFICIENT METHOD FOR THE ANALYSIS OF CORRELATION FUNCTIONS BASED ON A SPHERICAL HARMONICS EXPANSION OF THE BOLTZMANN TRANSPORT EQUATION'S GREEN'S FUNCTION

Chr. Jungemann, P. Graf, G. Zylka, R. Thoma, and W.L. Engl
Institut für Theoretische Elektrotechnik
University of Aachen, Kopernikusstr. 16, 52056 Aachen, Germany

Abstract

A new method for a more efficient calculation of correlation functions for stationary and Markovian processes described by the Boltzmann transport equation is given. The conditional probability representing the dynamics of the system is expanded with spherical harmonics. In the resulting discrete system the evaluation of the correlation functions involves only matrix-vector operations which can be performed very efficiently. The results for different electric and magnetic fields agree very well with standard Monte Carlo results and the cpu time usage is about one order of magnitude smaller.

I. INTRODUCTION

Macroscopic balance equations derived from the Boltzmann transport equation (BTE) rely on transport coefficients which in part can not be directly determined from experiment [1,2]. A powerful tool to obtain these coefficients is the correlation function analysis of certain microscopic quantities by Monte Carlo (MC) simulations of stationary homogeneous bulk systems [3,4]. The diffusion constant for example can be obtained from the time integral over the velocity autocorrelation function [5]. Since MC simulations are cpu time intensive and the correlation functions are needed for various doping concentrations as well as electric and magnetic field conditions more efficient methods are required.

II. DESCRIPTION OF THE METHOD

Correlation functions are calculated with the joint distribution function $f(\vec{k},t,\vec{k}_0,t_0)$ which can be expressed by the conditional probability $P(\vec{k},t|\vec{k}_0,t_0)$ times the normalized distribution function $f(\vec{k}_0,t_0)$. $P(\vec{k},t|\vec{k}_0,t_0)$ describes the dynamics of the system and $f(\vec{k}_0,t_0)$ the state of the particle ensemble. The correlation function of the microscopic quantities $X(\vec{k})$ and $Y(\vec{k})$ is given by:

$$< X(t)Y(t_0) >= \frac{\Omega^2}{(2\pi)^6}\int d^3k \int d^3k_0 X(\vec{k})P(\vec{k},t|\vec{k}_0,t_0)f(\vec{k}_0,t_0)Y(\vec{k}_0), \qquad (1)$$

where Ω denotes the system volume. $P(\vec{k},t|\vec{k}_0,t_0)$ is the conditional probability (CP) that a particle started at time t_0 with wavevector $\vec{k}_0$ appears at time t with wavevector $\vec{k}$. Since our MC model includes only one particle scattering processes the CP is also the Green's function of the corresponding BTE. We investigate only stationary processes with constant electric and magnetic fields. Thus $P(\vec{k},t|\vec{k}_0,t_0)$ equals $P(\vec{k},t-t_0|\vec{k}_0,0)$. Since the process is Markovian the CP for the

discrete times $(i\delta t, 0)$ with $i \geq 1$ can be calculated as the i-times product of the CP for the time step δt utilizing the Chapman-Kolmogorov equation [6]:

$$P(\vec{k}, i\delta t|\vec{k}_0, 0) = \left[\frac{\Omega}{(2\pi)^3}\right]^{i-1} \int d^3k_{i-1} \cdots \int d^3k_1 P(\vec{k}, \delta t|\vec{k}_{i-1}, 0) \cdots P(\vec{k}_1, \delta t|\vec{k}_0, 0). \quad (2)$$

In our new approach the CP is discretized by a spherical harmonics expansion in the solid angles [7] and an equidistant discretization of the absolute values of $\vec{k}$ and $\vec{k}_0$. The spherical harmonics expansion is especially well suited for this problem because the investigated microscopic quantities involve only spherical harmonics up to the second order. The discrete CP has the form of a matrix and equation (2) reads:

$$\underline{\underline{P}}(i\delta t|0) = \left[\underline{\underline{P}}(\delta t|0)\right]^i. \quad (3)$$

The correlation function is now calculated by multiplying the CP matrix from the left and right with the vectors of $X(\vec{k})$ and $Z(\vec{k}_0) = f(\vec{k}_0, 0)Y(\vec{k}_0)$ as resulting from the discretization:

$$< X(i\delta t)Y(0) >= \underline{X}\left[\underline{\underline{P}}(\delta t|0)\right]^i \underline{Z}. \quad (4)$$

The direct calculation of the i-th power of the CP matrix being very cpu time intensive can be avoided by an iterative calculation of $\left[\underline{\underline{P}}(\delta t|0)\right]^i \underline{Z}$. In the first time step ($i = 1$) the product of the CP matrix and the vector on the right-hand side is performed. The result of this operation is again a vector. For the next time step this vector is multiplied with the CP matrix once more. Thus only matrix-vector operations have to be performed which are much faster than matrix-matrix operations.

The CP can be obtained from the BTE directly with a perturbation series expansion [9,10]. This formulation is equivalent to the MC method but direct numerical evaluation is very tedious and the MC method is normally preferred. But for zero electric field and a scattering rate $S(\epsilon)$ which depends only on energy ϵ the CP has a simple form. Truncating the perturbation series expansion after the second term and including particle number conservation results in the following expression for the CP (arbitrary magnetic field):

$$P(\vec{k}, t|\vec{k}_0, 0) = \exp[-S(\epsilon(\vec{k}))t]\delta\left(\vec{k} - \vec{k}_0 + \frac{e}{\hbar}\int_0^t d\tau \vec{v}(\tau) \times \vec{B}\right) \quad (5)$$

$$+\frac{\Omega}{(2\pi)^3}\int_0^t dt_1 \exp[-S(\epsilon(\vec{k}))(t - t_1)]S\left(\vec{k} + \frac{e}{\hbar}\int_{t_1}^t d\tau \vec{v}(\tau) \times \vec{B}\middle|\vec{k}_0 - \frac{e}{\hbar}\int_0^{t_1} d\tau \vec{v}(\tau) \times \vec{B}\right),$$

where $S(\vec{k}|\vec{k}_0)$ denotes the transition rate and $\vec{v}$ the particle velocity. Since only one scattering event has been considered in eq. (5), the time t must be chosen sufficiently small compared with the scattering rate. The extension to multiple scattering events for longer times t is straight forward. With eq. (5) the discrete formulation of the CP can be calculated analytically and the setup time is reduced by three orders of magnitude in relation to an evaluation of the CP by the MC method.

In the case of arbitrary electric and magnetic fields the MC method is used, because the discretization of the equation corresponding to eq. (5) for nonzero electric field is very difficult. The discrete CP is evaluated by simulating particles with different initial wavevectors k_0, θ_0, φ_0 (spherical coordinates) for the time δt using a MC method [8]. The initial absolute values of the wavevectors are given by the discretization. The initial angles of the particles θ_0 and φ_0 are

chosen from an equidistant grid between 0 and π and 0 and 2π, respectively†. Since the final absolute values of the particles wavevectors do not match the values given by the discretization, they are mapped onto the grid by linear interpolation. A coefficient of the discrete CP is obtained by summing over all particles with the appropriate k, k_0 multiplied with the proper spherical harmonics of the initial and final state. If the electric and magnetic field are parallel to the polar axsis, the CP has the following symmetry property:

$$P(k,\theta,\varphi,i\delta t|k_0,\theta_0,\varphi_0,0) = P(k,\theta,\varphi-\varphi_0,i\delta t|k_0,\theta_0,0,0), \tag{6}$$

due to the employed Si-model [8]. This effect can be exploited to reduce the variance of the MC method for the setup of the CP by integrating analytically over the angle φ_0 in the spherical harmonics expansion. The setup procedures and the calculation of the correlation functions are well suited for parallelization.

III. RESULTS

The CP matrix has been discretized with about 50 points for the absolute value of $\vec{k}$ and an expansion up to the fourth order involving 25 spherical harmonics. Thus the dimension of the CP matrix is about 1250. 13 microscopic quantities for X and 13 different microscopic quantities for Y have been investigated resulting in 169 auto- and cross-correlation functions. In Table I execution times are listed for the new method and a MC simulation on a 4D 480 SGI computer. The new method is at least eight times faster than the MC simulation. Figure 1 shows stationary expectation values calculated with the new method and MC simulation as a function of the applied field for undoped silicon at room temperature [8] and good agreement is found.

TABLE I. Comparison of execution times for the setup of a CP matrix based on a 4th order expansion, calculation of correlation functions, total time and standard MC simulation for three electric fields.

field	Setup	cor. fun.	total	MC
(kV/cm)	(s)	(s)	(s)	(s)
1	1131	200	1331	48568
10	1179	465	1644	12520
100	858	463	1321	13116

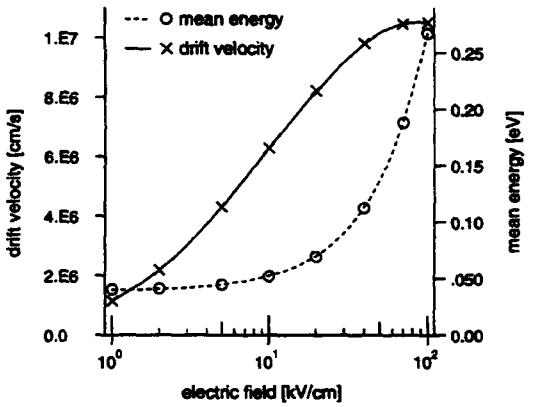


FIG. 1. Stationary expectation values for drift velocity and mean energy with the new method (lines) and MC simulation (symbols).

In Fig. 2 longitudinal and transversal velocity autocorrelation functions are shown and no difference is found between MC simulation and the new method based on an expansion up to the 3rd and 4th order. Results for other correlation functions involving 2nd order spherical harmonics (velocity is 1st order) showed differences between the expansion up to 3rd or 4th order. Therefore the 4th

† The particle weight is modulated in θ_0-direction satisfying the Simpson rule and in φ_0-direction satisfying a trapezoidal rule to ensure sufficient orthogonality of the spherical harmonics in the process of numerical integration with the MC method.

order expansion was used for which the results agreed very well with MC data. In the case of weak correlation the new method is superior to the MC simulation because of the high statistical noise of the MC method.

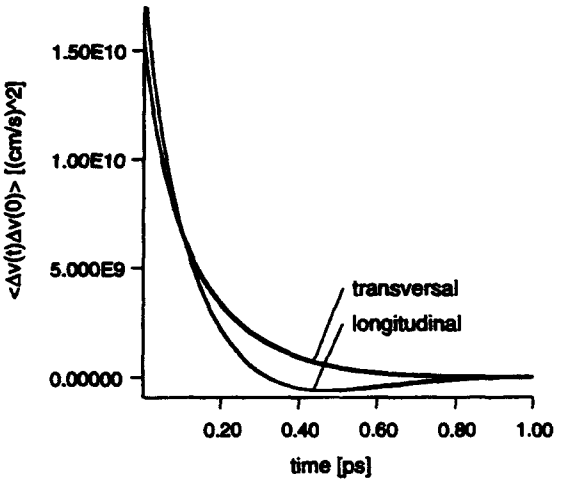


FIG. 2. Autocorrelation functions for longitudinal and transversal velocity calculated with MC and CP matrices based on 3rd and 4th order expansions for $10kV/cm$ $(300K)$.

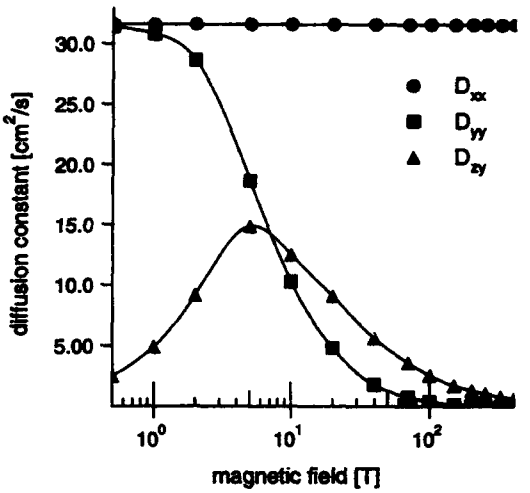


FIG. 3. Elements of diffusion constant tensor (D_{xx}, D_{yy}, D_{xy}) for a magnetic field in x-direction and zero electric field $(300K)$.

In Fig. 3 elements of the diffusion constant tensor for zero electric field are shown as a function of the magnetic field up to $400T$ neglecting quantization effects. The magnetic field lies parallel to the x-axsis which is the polar axsis. The setup time of the CPs was below three seconds using the analytical method mentioned above.

IV. CONCLUSION

A new method for the calculation of correlation functions has been developed which is much faster than the standard MC simulation and less noisy. It agrees very well with MC results for different electric and magnetic fields. The method is well suited for parallelization.

ACKNOWLEDGEMENTS

The financial support of the "Bundesministerium für Forschung und Technologie" under contract No. NT 2792 D1 is gratefully acknowledged.

[1] K. Bløtekjær, IEEE Trans. Electron Devices **ED-17**, 38 (1970)
[2] R. Thoma et al., IEEE Trans. Electron Devices **ED-38**, 1343 (1991)
[3] T. Kuhn et al., Phys. Rev. B **42**, 11132 (1990)
[4] R. Thoma and W.L. Engl, SISDEP Tech. Dig **4**, 185 (1991)
[5] D.K. Ferry and J.R. Barker, J. Appl. Phys. **52**, 818 (1980)
[6] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland Physics Publishing 1985)
[7] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer-Verlag, New York, 1982)
[8] R. Brunetti et al., Solid-State Electron. **32**, 1663 (1989)
[9] J. Zimmermann et al., Solid-Sate Electron. **26**, 233 (1983)
[10] H.-J. Peifer,Ph.D. thesis, Aachen University, 1992

TIME-DEPENDENT ANALYSIS OF THE COUPLED HOT-CARRIER–HOT-PHONON BOLTZMANN EQUATIONS

L. Hlou, J. C. Vaissiere, J. P. Nougier, L. Varani, P. Houlet
Centre d'Electronique de Montpellier,
Université Montpellier II, 34095 Montpellier Cedex 5, France

L. Reggiani
Dipartimento di Fisica, Università di Modena,
Via Campi 213/A, 41100 Modena, Italy

M. Fadel
Faculté des Sciences d'Agadir,
Agadir, Morocco

P. Kocevar
Institut für Theoretische Physik, Universität Graz,
Universitätsplatz 5, A-8010 Graz, Austria

Abstract

We present a detailed investigation of the transient transport regime in InP at room temperature based on an original method to solve numerically the coupled hot-phonon–hot-carrier time-dependent Boltzmann Equations. The method enables a study of the perturbation of the phonon distribution function induced by hot carriers and the corresponding modifications of the carrier distribution function. As a consequence of the high numerical accuracy of the method, the time behavior of the main transport parameters can be investigated in great detail.

I. INTRODUCTION

The influence of hot phonons on carrier transport parameters in polar semiconductors has been theoretically studied in relation with non-ohmic transport [1], laser photoexcitation [2], and noise phenomena [3]. As concerning the transient regime, the influence of non-equilibrium phonons has been recently investigated in Ref. [1] by means of a Monte Carlo simulation in n-type GaAs. However, a detailed and extensive investigation of the effect of a non-equilibrium phonon population on the carrier distribution function (CDF), phonon distribution function (PDF) and main transport parameters is still lacking in the literature. In this communication we present a detailed analysis of the transient transport regime in InP at room temperature under spatially homogeneous conditions. We take advantage of an original method to solve numerically the coupled hot-phonon–hot-carrier Boltzmann Equations (BE) in the time domain. The method enables a study of the perturbation of the longitudinal-optical (LO) PDF induced by hot carriers and the corresponding modifications of the CDF. As a consequence, the time behavior of the main transport parameters can be investigated in great detail with an accuracy far beyond other existing numerical methods.

II. THE SYSTEM OF COUPLED EQUATIONS

To take into account the perturbation of the LO phonon population, it is convenient to write the time dependent BE for the CDF $f(\mathbf{k},t)$ in the following form:

$$\frac{\partial f(\mathbf{k},t)}{\partial t} = \hat{C}_{nc} f(\mathbf{k},t) + \hat{C}_{po} f(\mathbf{k},t) - \frac{f(\mathbf{k},t)}{\tau_{po}(\mathbf{k})} \tag{1}$$

Here $\hat{C}_{nc}$ is the operator including the external field term and collisions with acoustic deformation potential, piezoelectric, impurity, intervalley and intravalley non-polar optical phonon scatterings; $\hat{C}_{po}$ is the input term for polar-optic LO-phonon scattering and $[f(\mathbf{k},t)/\tau_{po}(\mathbf{k})]$ the output term, $1/\tau_{po}(\mathbf{k})$ being the polar-optic scattering rate. The detailed expressions for the above operators

can be found in Ref.[4]. We remark that $\hat{C}_{po}$ and τ_{po} depend on the PDF and, as a consequence, Eq. (1) becomes non-linear since its solution requires the knowledge of the PDF.

The time dependent phonon BE gives the time variation of the PDF $N(\mathbf{q},t)$. This variation is the result of the balance between two terms: the former is associated with the phonon appearance and disappearance due to carrier emission and absorption and the latter is associated with the nonelectronic lattice relaxation of the phonons. Therefore, we write the phonon BE in the following form:

$$\frac{\partial N(\mathbf{q},t)}{\partial t} = \hat{C}_{ph}N(\mathbf{q},t) - \hat{D}_{ph}N(\mathbf{q},t) - \frac{[N(\mathbf{q},t) - N_L]}{\tau_L} \tag{2}$$

where $\hat{C}_{ph}$ and $\hat{D}_{ph}$ are the gain and loss operators related to emission and absorption of phonons by carriers, respectively, N_L is the thermal-equilibrium Bose-Einstein distribution and τ_L the non-electronic phonon relaxation time.

III. NUMERICAL SIMULATION

The main task is now to solve the system of coupled equations (1) and (2). To do that, we have devised the following procedure:

(i) The CDF and PDF at thermal equilibrium are introduced in Eqs. (1) and (2) thus calculating $[\partial f(\mathbf{k},t)/\partial t]_{t=0}$ and $[\partial N(\mathbf{q},t)/\partial t]_{t=0}$.

(ii) From the knowledge of these quantities we determine $f(\mathbf{k},\Delta t)$ and $N(\mathbf{q},\Delta t)$ by Taylor expansion.

(iii) The new CDF and PDF at time Δt are introduced in Eqs. (1) and (2) thus calculating $[\partial f(\mathbf{k},t)/\partial t]_{t=\Delta t}$ and $[\partial N(\mathbf{q},t)/\partial t]_{t=\Delta t}$.

(iv) From the knowledge of these quantities we determine $f(\mathbf{k},2\Delta t)$ and $N(\mathbf{q},2\Delta t)$.

(v) Steps (iii) and (iv) are iteratively repeated until the stationary regime is reached.

This selfconsistent procedure has been found to present some numerical problems. As a matter of fact, when solving the usual BE (not coupled with the phonon BE) the different operators which appear are independent of time: this enables their associated matrix to be calculated only at the beginning of the simulation. In the present case the operator associated with the LO-phonon interaction depends on the PDF, which is a time dependent quantity: therefore $\hat{C}_{po}$ must be recalculated at each time step. Furthermore, the inclusion of the equation for the time evolution of the PDF has been found to increase significantly the duration of the transient regime. This leads to the necessity of adopting some kind of optimization of the numerical algorithm in order to save computer time.

As concerning the carrier BE we notice that Eq. (1) can be written in matrix form as:

$$\left[\frac{\partial f}{\partial t}\right] = [C]_{nc}[f] + [C]_{po}[f] + \left[\frac{1}{\tau_{po}}\right][f] \tag{3}$$

where $[C]_{po}$ and $[1/\tau_{po}]$ are the matrices associated with the LO operator depending on time through the PDF, and $[C]_{nc}$ is the time-independent matrix associated with the operator $\hat{C}_{nc}$. The calculation of the first two matrices is the part of the program which requires most of the computer time. To try to avoid this problem we have verified that the increase in the time-duration of the transient is due to the slow time variation of the PDF. Therefore, we have calculated the time-dependent matrices ONLY when one value of $N(\mathbf{q},t)$ has undergone a significative variation (about 1 %).

As concerning the phonon BE, Eq. (2) can be written in the following form:

$$\frac{\partial N_i(t)}{\partial t} = \left[\sum_j E_{ij}f_j(t)\right] N_i(t) + \sum_j F_{ij}f_j(t) - \frac{N_i(t)}{\tau_L} + \frac{N_L}{\tau_L} \tag{4}$$

The coefficients E_{ij} and F_{ij} are independent of the CDF and therefore independent of time: this enables to calculate the matrices $[E]$ and $[F]$ only at the beginning of the simulation, thus saving a significant amount of computer time.

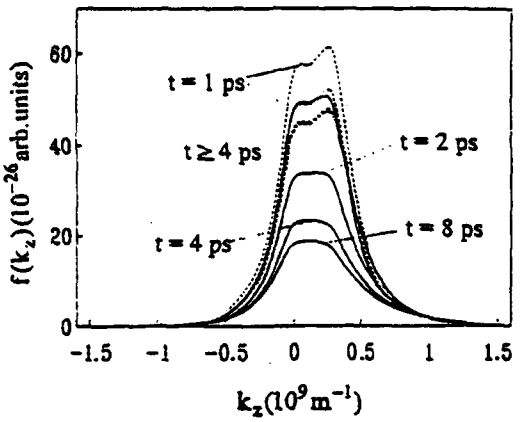


Fig. 1 - *Carrier distribution function* $f(k_z) = f(k_x = 0, k_y = 0, k_z)$ *along the electric field as a function of* k_z, *in the* Γ *valley of InP, for* $T_L = 300\ K$, $N_D = 10^{17}\ cm^{-3}$, $E = 10\ kV/cm$ *and the reported times. Solid lines: calculations taking into account hot-phonons; dashed lines: calculations assuming phonons to be at thermal equilibrium.*

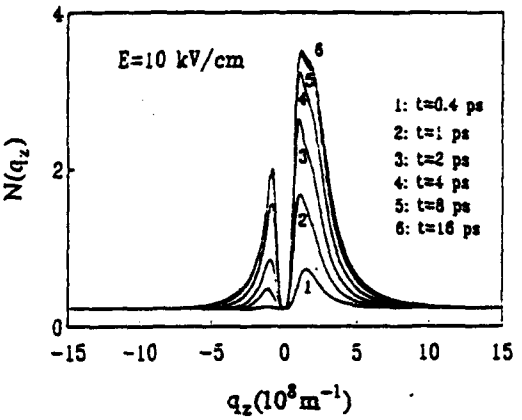


Fig. 2 - *Phonon distribution function* $N(q, \theta = 0)$ *along the electric field in InP, for* $T_L = 300\ K$, $N_D = 10^{17}\ cm^{-3}$, $E = 10\ kV/cm$ *and the reported times from the beginning of the transient.*

To evaluate the gain in computer time some tests have been performed on an IBM 3090. By using the optimization procedure here described we were able to reduce the CPU time of a typical simulation from 120 hours to about 6 hours, thus gaining a factor of about 20.

IV. RESULTS

The theory is applied to the case of n-type InP at a temperature $T_L = 300\ K$, a doping concentration $N_D = 10^{17}\ cm^{-3}$ and with the same material parameters as reported in the Appendix B of Ref. [4]. Two spherical and non-parabolic bands (one Γ and four equivalent L valleys) are taken into account. The simulation includes the following intra and intervalley scattering mechanisms: acoustic deformation potential (in elastic approximation), piezoelectric (in elastic approximation), polar optical, impurity (in Brooks-Herring model including a screened Coulomb potential) and non-polar optical intervalley. The LO-phonon relaxation time τ_L has been taken as 5.8 *ps*.

Figure 1 reports the CDF in the Γ valley, at different times from the beginning of the transient and for an abruptly applied electric field of 10 kV/cm. Since some time is required in order to perturb the PDF from its equilibrium value, for times shorter than 0.4 *ps* no significant difference is observed between the values of the CDF assuming phonons to be perturbed or at thermal equilibrium. For times longer than 0.4 *ps* the two CDF begin to differ one from each other, the difference becoming more evident at increasing times. We notice also that the presence of a non-equilibrium phonon population is responsible for an increase in the time duration of the transient.

Figure 2 reports the results for the PDF at different times from the beginning of the transient and for an electric field of 10 kV/cm. The appearance of the perturbation of the PDF is related to the displacement of the carriers in the high energy region [4]. At the beginning of the transient the CDF is displaced under the action of the electric field; as a consequence the PDF develops a peak at small positive phonon wave-vectors associated with the phonons emitted by carriers at high energy. This peak progressively increases with time due to the enhanced LO-phonon emission by the carriers.

Figure 3 reports the results for the drift-velocity at three electric fields of 5, 10 and 20 kV/cm. For very short times (≤ 0.4 *ps*) there is nearly no difference between the values of the drift-velocity obtained with and without hot-phonons. For an intermediate electric field of 10 kV/cm, a second overshoot is observed which is related to the perturbation of the PDF.

Figure 4 reports the results concerning the average carrier energy for the same electric fields

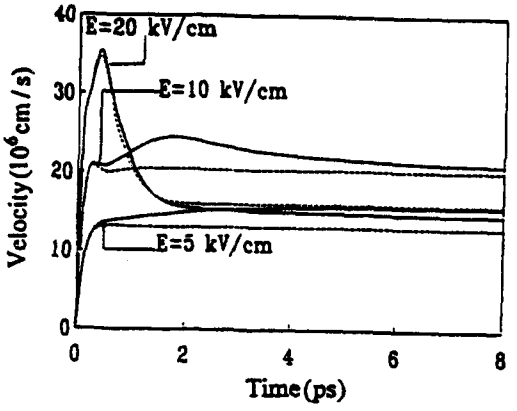


Fig. 3 - *Mean-carrier drift-velocity as a function of time in InP, for T_L = 300 K, N_D = 10^{17} cm^{-3}, and the reported electric fields. The solid lines refer to calculations taking into account hot-phonons and the dashed lines to calculations obtained assuming phonons to be at thermal equilibrium.*

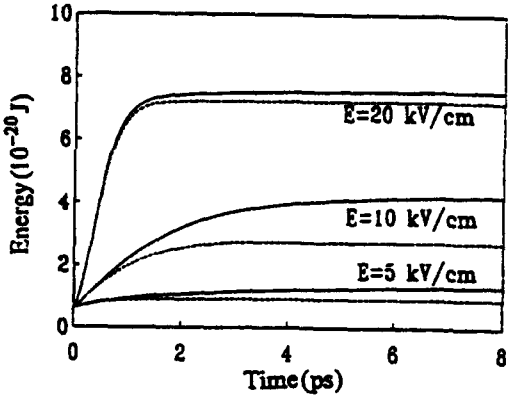


Fig. 4 - *Mean-carrier energy as a function of time in InP, for T_L = 300 K, N_D = 10^{17} cm^{-3}, and the reported electric fields. The solid lines refer to calculations taking into account hot-phonons and the dashed lines to calculations obtained assuming phonons to be at thermal equilibrium.*

as in Fig. 3. Also in this case we observe that the presence of non-equilibrium phonons is found to be responsible for modifications in the transient regime. As already noticed for the case of the drift-velocity, these modifications appear after a time required for the perturbation of the PDF to take place.

V. CONCLUSIONS

We have presented a detailed investigation of the transient transport regime in InP at room temperature. We take advantage of a numerically extremely efficient method for solving the coupled hot-phonon–hot-carrier Boltzmann Equations in both the linear and non-linear regime. The accuracy of the method is particularly evident during the transient which is fundamental to the performances of high-frequency semiconductor devices. The non-linearity introduced by the phonon disturbance is responsible for a great complexity in looking for a solution of the whole problem and noticeable effort has been made in order to optimize the code and obtain a reasonable CPU time. Modifications of the transient regime at low and intermediate electric fields are observed.

ACKNOWLEDGMENTS

This work has been performed within the *European Laboratory for Electronic Noise* (ELEN) and supported by the Commission of European Community through the contracts ERBCHRXCT920047 and ERBCHBICT920162. Partial support from the Italian *Consiglio Nazionale delle Ricerche* (CNR) and the *Centre de Competences en Calcul Numériques Intensif* (C3NI) is gratefully acknowledged.

REFERENCES

[1] M. Rieger, P. Kocevar, P. Lugli, P. Bordone, L. Reggiani and S.M. Goodnick, Phys. Rev. **39**, 7866 (1989).
[2] P. Lugli, P. Bordone, L. Reggiani, M. Rieger, P. Kocevar and S. M. Goodnick, Phys. Rev. **39**, 7852 (1989).
[3] P. Bordone, L. Varani, L. Reggiani, L. Rota and T. Kuhn, Appl. Phys. Lett. **63**, 1107 (1993).
[4] J.C. Vaissiere, J.P. Nougier, P. Fadel, L. Hlou and P. Kocevar, Phys. Rev. **46**, 13082 (1992).

NUMERICAL SOLUTION OF THE PERTURBED BOLTZMANN EQUATION IN FREQUENCY AND TIME DOMAINS

J. C. Vaissiere, J. P. Nougier, L. Varani, P. Houlet
Centre d'Electronique de Montpellier,
Université Montpellier II, 34095 Montpellier Cedex 5, France

L. Hlou
Faculté des Sciences de Kénitra,
Kénitra, Morocco

L. Reggiani
Dipartimento di Fisica, Università di Modena,
Via Campi 213/A, 41100 Modena, Italy

E. Starikov, P. Shiktorov
Semiconductor Physics Institute,
Goshtauto 11, 2600 Vilnius, Lithuania

Abstract

We present two original methods which yield the small-signal response around the d.c. bias in bulk semiconductors, using direct numerical resolutions of the perturbed Boltzmann equation. The first method operates in the frequency domain. An a.c. sinusoidal electric field perturbation superimposed to the d.c. field produces an a.c. perturbation of the distribution function which is computed at each frequency. The second method operates in the time domain. A step electric field perturbation is superimposed at time t=0 to the d.c. field. The resulting perturbations of the distribution function and of the average velocity are then computed as a function of time. These methods are applied to the case of holes in silicon at T=300 K under hot-carrier conditions and used to compute the differential-mobility spectrum.

I. INTRODUCTION

Small-signal response functions around the bias point are known to play a fundamental role in the investigation of hot-carrier transport and noise in bulk semiconductors. In the time domain they reflect both dynamic and relaxation processes inherent to the hot-carrier system and can be used for the detailed investigation of kinetic phenomena. In the frequency domain they provide the differential mobility spectrum which is necessary for several purposes, such as: to evaluate a possibility of amplification and generation, to calculate the gain or the absorption coefficients, to obtain the noise temperature using additionally the spectral density of velocity fluctuations, etc. To date the most comprehensive theoretical analysis of these phenomena is based on numerical solutions of the Boltzmann Equation (BE), typically by means of Monte Carlo simulations. However, together with evident advantages, the Monte Carlo method has also inherent shortcomings mainly related to the stochastic nature of the procedure: as a matter of fact, the standard Monte Carlo scheme meets difficulties in calculating with high accuracy quantities on a hydrodynamic time scale such as the small-signal kinetic coefficients. Other alternative methods deals with the steady state hot-carrier transport and often cannot be reformulated in terms of the time-dependent BE. In this communication, we present two original deterministic (as opposite to stochastic) methods which yield the small-signal response around the d.c. bias in bulk semiconductors, using direct numerical resolutions of the perturbed BE.

II. THEORY

The distribution function $f(\mathbf{k},t)$ of carriers in homogeneous nondegenerate semiconductors with a uniform external applied electric field $\mathbf{E}(t)$ is the solution of the time-dependent BE. In a constant electric field $\mathbf{E}_s$ of magnitude E_s, $f(\mathbf{k},t)$ takes the stationary value $f_s(\mathbf{k})$. If a small electric field $\delta\mathbf{E}(t)$ is superimposed on $\mathbf{E}_s$, it produces a variation of the distribution function $\delta f(\mathbf{k},t)$ which is the solution of the perturbed BE in time domain [1]:

$$\frac{\partial}{\partial t}\delta f(\mathbf{k},t) + \frac{e\mathbf{E}_s}{\hbar}\cdot\nabla_k \delta f(\mathbf{k},t) - C\delta f(\mathbf{k},t) = -\frac{e\delta\mathbf{E}(t)}{\hbar}\cdot\nabla_k f_s(\mathbf{k}) \qquad (1)$$

where $\hbar$ is the reduced Planck constant and C the collision operator.

1. Harmonic-Response Method

When the perturbation is sinusoidal [$\delta\mathbf{E}_{har} = \delta\mathbf{E}exp(i\omega t)$], the response is also sinusoidal [$\delta f(\mathbf{k},t) = \delta f(\mathbf{k},\omega)exp(i\omega t)$]. Then from Eq. (1) we obtain the perturbed BE in frequency domain [1]:

$$i\omega\delta f(\mathbf{k},\omega) + \frac{e\mathbf{E}_s}{\hbar}\cdot\nabla_k\delta f(\mathbf{k},\omega) - C\delta f(\mathbf{k},\omega) = -\frac{e\delta\mathbf{E}}{\hbar}\cdot\nabla_k f_s(\mathbf{k}) \tag{2}$$

From the knowledge of $\delta f(\mathbf{k},\omega)$ we obtain the Fourier transform $\delta\mathbf{v}(\omega)$ of $\delta\mathbf{v}(t)$ as:

$$\delta\mathbf{v}(\omega) = \left[\int \mathbf{v}(\mathbf{k})\delta f(\mathbf{k},\omega)d^3k\right]\left[\int f_s(\mathbf{k})d^3k\right]^{-1} \tag{3}$$

The complex quantities $\delta\mathbf{v}(\omega)$ and $\delta\mathbf{E}$ are linearly related through the a.c. differential mobility $\mu(\omega)$ as: $\delta\mathbf{v}(\omega) = \mu(\omega)\delta\mathbf{E}$.

By assuming a spherical symmetry of the band model the perturbation term $\delta f(\mathbf{k},\omega)$ can be written as $\delta f(k,\theta,\omega)$ where $k = |\mathbf{k}|$ and $\theta = (\mathbf{E},\mathbf{k})$. After discretization, the gradient and the collision operators in Eq. (2) appear as linear combinations of $\delta f(k,\theta,\omega)$. In practice, the computed quantity is $\delta f_E = \delta f(k,\theta,\omega)/\delta E$, represented by a column matrix $[\delta f_E]$ which has a real part $[\delta f_E]_{re}$ and an imaginary part $[\delta f_E]_{im}$ calculated as:

$$\begin{aligned}[\delta f_E]_{re} &= [A]\left([A]^2 + \omega^2[I]\right)^{-1}[g] \\ [\delta f_E]_{im} &= -\omega\left([A]^2 + \omega^2[I]\right)^{-1}[g]\end{aligned} \tag{4}$$

where the square matrix $[A]$ represents the discretized operator $[(e\mathbf{E}_s/\hbar)\nabla_k - C]$, the column matrix $[g]$ represents the discretized vector $(e/\hbar)\nabla_k f_s(\mathbf{k})$, and $[I]$ is the identity matrix. The unknowns on the left-hand side of Eq. (4) are easily obtained using standard numerical techniques (Gauss procedure). This method enables to use an arbitrary value of δE: indeed, since the computed quantity is δf_E the actual value of δE does not appear in Eq. (4). Furthermore, the solution of Eq. (4) requires a specific program. We remark also that the solution of Eq. (2) presents difficulties for low frequencies ($< 10^8$ Hz) because its associated determinant becomes small [2].

2. Impulse-Response Method

In this case we apply a step-like electric field perturbation, $\delta\mathbf{E}_{step}(t) = \delta\mathbf{E}u(t)$ where $u(t)$ is the step function $u(t) = 1$ if $t \geq 0$ and $u(t) = 0$ if $t < 0$. The step distribution response $\delta f_{step}(\mathbf{k},t)$ is then the solution of Eq. (1), and the step velocity response $\delta\mathbf{v}_{step}(t)$ is given by Eq. (3) where $\delta f(\mathbf{k},\omega)$ is replaced by $\delta f_{step}(\mathbf{k},t)$. To obtain the transient distribution function $\delta f_{step}(\mathbf{k},t)$, we first solve (using a direct method [3]) the transient BE in the constant field E_s, so calculating $f_s(\mathbf{k})$. Then we solve the transient BE in a constant field $E_s + \delta E$, with the initial distribution equal to $f_s(\mathbf{k})$, thus evaluating the transient $f(\mathbf{k},t)$. The step distribution response is then calculated by difference as $\delta f_{step}(\mathbf{k},t) = f(\mathbf{k},t) - f_s(\mathbf{k})$. Then $\delta\mathbf{v}(\omega)$ is calculated as:

$$\delta\mathbf{v}(\omega) = \int_{-\infty}^{+\infty} \frac{d\delta\mathbf{v}_{step}(t)}{dt} exp(-i\omega t)dt \tag{5}$$

Thus Eq. (5) provides a second method to obtain the a.c. differential mobility.

This method can be used by employing the same program developed for the direct solution of the BE [3] or the Scattered Packet Method [4] since the accuracy of these methods is sufficient to compute precisely $d\delta\mathbf{v}_{step}(t)/dt$. On the other hand, with respect to the harmonic-response method, it is necessary to take a value of δE large enough (typically between 1 and 10 % of E_s). This calculation can take advantage of an acceleration technique described in Ref. [5].

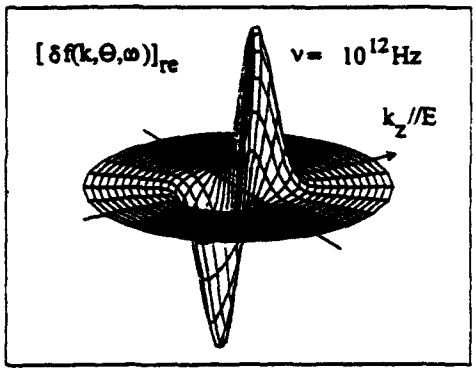


Fig. 1 - *3-D representation of the real part of the perturbation of the distribution function* $[\delta f(k,\theta,\omega)]_{re}$ *(harmonic-response method), in arbitrary scales, at frequency* $\nu = \omega/2\pi = 10^{12}$ *Hz, for holes in Si,* $T = 300$ *K,* $E_s = 10$ *kV/cm, corresponding to a perturbing field* $\delta E = 1$ *V/cm.*

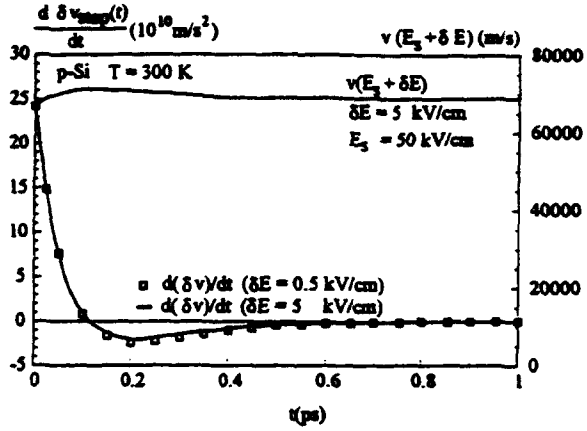


Fig. 2 - *Drift velocity (right scale) and time derivative of the transient response of the drift velocity (left scale). Calculations refer to holes in Si with* $T = 300$ *K,* $E_s = 50$ *kV/cm and the reported values of* δE.

III. RESULTS

The above procedures are used to calculate the small-signal response characteristics of holes in Si at T=300 K. The microscopic model is based on a single spherical nonparabolic-band and considers scattering with acoustic and non-polar optical phonon mechanisms as described in Ref. [6].

Figure 1 shows the real part of the perturbation of the distribution function $\delta f(k,\theta,\omega)$ calculated using the harmonic-response method [see Eq. (4)]. Each radial curve gives the variation of $\delta f(k,\theta,\omega)$ at a given value of the angle θ. In analogy with the Drude model for the a.c. conductivity, the real part describes the dissipative contribution which is in phase with the field while the imaginary part (here not reported) describes the optical contribution which is in quadrature with the field. Figure 2 reports the time dependence of the drift velocity when at time $t = 0$ a step electric field is superimposed to E_s. The same figure shows the time derivative of the transient response of the drift velocity for two different values of δE (we notice that, in order to compare the two curves, the reported values have been divided by $\delta E/(1\ V/m)$). The excellent agreement observed shows that a δE of few percents of E_s can be employed in order to compute the linear response of the system. Figure 3 shows the time-derivative of the velocity response-function $\delta v_{step}(t)$ (divided by $\delta E/(1\ V/m)$) whose Fourier transform gives $\delta v(\omega)$ according to Eq. (5). At time t=0, all curves have practically the same value of $[d\delta v_{step}(t)/dt]_{t=0} = e\delta E/m^*$, where m^* is the effective mass. The small changes at $t = 0$ are due to the non-parabolicity of the band. At zero and low electric fields, the shape of the velocity response- function is practically exponential with a characteristic time constant which corresponds to momentum relaxation. At higher fields the shape becomes more complicated by exhibiting a negative part which is understood as follows. At the initial stage of the velocity relaxation, carriers obtain extra velocity, since their initial momentum relaxation time τ_p is somewhat longer than that in the new steady-state. Then, the energy relaxation affects τ_p (i.e. τ_p becomes shorter) and this extra velocity is lost. Therefore, the energy relaxation is responsible for the negative contribution of the velocity response-function.

The harmonic and impulse response methods are further used to calculate the differential mobility spectrum which is reported in Fig. 4. The circles and the solid line show the a.c. mobility computed respectively with the harmonic- and the impulse-response method. The agreement between the two techniques is excellent, thus validating the present approach. In particular, from Fig. 4 significant deviations from the simple Drude slope of μ_r and μ_i are evidenced. This peculiarity is explained as follows. At zero and low d.c. electric fields the impulse velocity response decreases monotonously with increasing time (see Fig. 3), and the characteristic relaxation time involved is

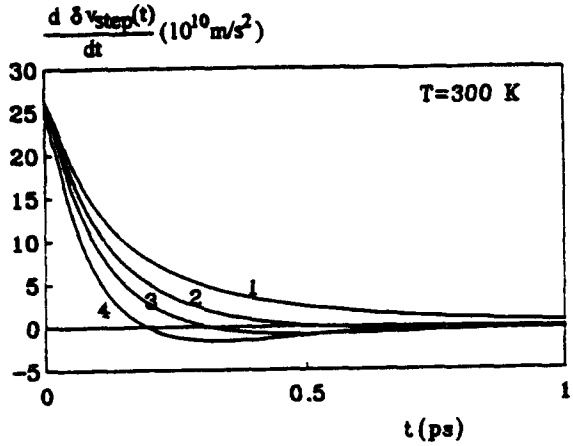


Fig. 3 - *Time derivative (divided by $\delta E/(1\ V/m)$) of the transient response of the drift velocity. Calculations refer to holes in Si with $T = 300\ K$, and $\delta E = 1\ V/cm$ for $E_s = 0$, and $\delta E = 0.1E_s$ otherwise. 1: $E_s = 0$; 2: $E_s = 5\ kV/cm$; 3: $E_s = 10\ kV/cm$; 4: $E_s = 20\ kV/cm$.*

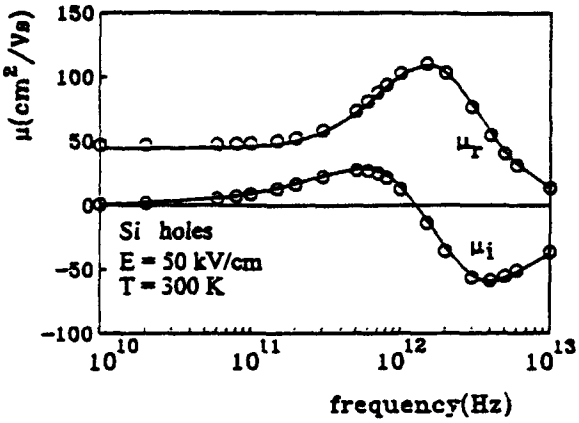


Fig. 4 - *Real part μ_r and imaginary part μ_i of the a.c. mobility for holes in Si at an applied d.c. electric field $E_s = 50\ kV/cm$. Circles : harmonic-response method with $\delta E = 1V/cm$; Solid line: impulse response-method with $\delta E = 0.1E_s$.*

then the momentum relaxation time. At higher fields, the energy relaxation time begins to play a role. This results in a negative value of $[d\delta v_{step}(t)/dt]$, which corresponds to a bump in μ_r. With increasing electric field, μ_r increases in the low frequency region, which implies a positive value of μ_i then decrease resulting in a negative value of μ_i.

IV. CONCLUSIONS

We have presented two methods for calculating the small-signal response around the d.c. bias in bulk semiconductors, using direct numerical resolutions of the perturbed Boltzmann equation. Both methods have been validated for the case of holes in Silicon and proven to give exactly the same results when used to compute the differential mobility spectrum. The harmonic-response method requires to perform a simulation for each frequency of interest while the impulse-response method gives directly the whole spectrum within one simulation. The methods are deterministic and therefore overcome the difficulties of the stochastic methods (such as Monte Carlo simulations) in calculating with high accuracy transport parameters on a hydrodynamic time scale.

ACKNOWLEDGMENTS

This work has been performed within the *European Laboratory for Electronic Noise* (ELEN) and supported by the Commission of European Community through the contracts ER-BCHRXCT920047 and ERBCHBICT920162. Partial support from the italian *Consiglio Nazionale delle Ricerche* (CNR) and the *Centre de Competences en Calcul Numériques Intensif* (C3NI) is gratefully acknowledged.

REFERENCES

[1] J. C. Vaissiere, J. P. Nougier, L. Varani, P. Houlet, L. Hlou, E. Starikov, P. Shiktorov and L. Reggiani, Phys. Rev., in press.
[2] J. P. Aubert, J.C. Vaissiere and J.P. Nougier, J. Appl. Phys. **56**, 1128 (1984).
[3] P.A. Lebwohl, P.M. Marcus, Solid State Commun. **9**, 1671 (1971).
[4] J.P. Nougier, L. Hlou, P. Houlet, J.C. Vaissiere and L. Varani, Proceedings of the 3rd *Int. Workshop on Computational Electronics*, Portland (1994).
[5] L. Hlou, Thèse de Doctorat, Universitè Montpellier II (France), 1993 (available upon request).
[6] J. C. Vaissiere, Thèse de Doctorat ès Sciences, Universitè Montpellier II (France), 1986 (available upon request).

SCALED ENSEMBLE MONTE CARLO STUDIES OF IMPACT IONIZATION

A. M. Kriman[a] and R. P. Joshi[b]
[a] *Elec. &Comp. Eng'ng & CEEM., Univ. at Buffalo NY, 14260 (kriman@acsu.buffalo.edu)*
[b] *Elec. &Comp. Eng'ng, Old Dominion Univ., Norfolk, VA, 23529 (rpj@ee.odu.edu)*

Abstract

We introduce a scaled ensemble Monte Carlo (SEMC) technique to simulate high field electron transport. This technique is designed to improve the accuracy of the phase-space statistics of the non-equilibrium carrier distribution. As recognized in weighted-ensemble Monte Carlo, increasing the number of simulated particles is inefficient, since it substantially increases the computation times without ensuring adequate representation of the sparsely populated regions of phase space. A scaled scheme is especially important for these transient simulations, since one cannot average the electronic trajectories over sufficiently long times.

The SEMC technique we propose overcomes this problem by redistributing the computational effort to weight the low-density regions of k-space more heavily. This is done through a formulation which uses an energy-dependent factor to scale the distribution functions and the scattered rates. The technique has the advantage of flexibility and simplicity in coding, is very similar to the traditional non-weighted approach, and doesn't involve particle-splitting. In the present work, the SEMC procedure is applied to simulate impact ionization for high-field transport, using a soft threshold. Since the impact ionization coefficient is strongly affected by the high-energy tail of the distribution function, we apply SEMC to evaluate this. In general, results depend sensitively on band structure and on coupling to other energy dissipation modes.

I. INTRODUCTION

The simulation of impact ionization is, in principle, not very different from the simulation of other scattering processes in semiconductors: an instantaneous, energy-dependent microscopic scattering rate $R(E)$ is determined used directly in Monte Carlo simulation; or a parameter α determined from $R(E)$ and the non-equilibrium distribution function is used in drift-diffusion simulation. The main practical difficulty is that the total scattering rate, and α, depend sensitively on the distribution function at very high energy. This is true even for the low fields where α is approximately field independent. Since only a small fraction of particles are found at the relevant high energies, ordinary Monte Carlo techniques produce results with large statistical uncertainties, and special modifications must be used which enhance the accuracy of the simulation in the low particle-density regimes. Here we describe the use of a new method, *scaled* ensemble Monte Carlo (SEMC), to perform this function.

Like most Monte Carlo treatments, the present one is based on an essentially classical description. In the absence of significant interparticle correlations, the system is described formally by a time-dependent single-particle distribution function $f(\mathbf{r},\mathbf{p};t) = f(\mathbf{x};t)$ [$\mathbf{x} = (\mathbf{r},\mathbf{p})$ is a phase space coördinate]. The distribution function obeys the Liouville equation:

$$\frac{\partial f}{\partial t} = \{H,f\} + \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}, \tag{1}$$

where $H = H(\mathbf{x})$ is the time-independent single-particle Hamiltonian, and $\{\cdot,\cdot\}$ is the Poisson bracket. In our simulations, we used a multi-valley Hamiltonian for electrons in GaAs bulk [2], with central Γ and satellite L and X valleys. In our band model, the different satellite minima communicate only via deformation scattering—the wave vector is unrestricted in principle, although the region of crystal momenta between minima is essentially unoccupied for energetic reasons. This standard model is augmented by a uniform electric field.

We make the usual assumptions that collisions or scattering events take place on time scales much shorter than the time between collisions, so effects such as collision broadening can be neglected. Further, the scattering events are in fact approximated as instantaneous, so that intracollisional field effects can also be ignored. In this approximation we treat acoustic and optical phonon scattering by deformation potential, and polar optical scattering. All important sources of potential and phonon scattering can be written in the form

$$\left(\frac{\partial f(\mathbf{x};t)}{\partial t}\right)_{\text{coll}} = -\Gamma(\mathbf{x};t)\, f(\mathbf{x};t) + \int \Gamma(\mathbf{x},\mathbf{x}';t) f(\mathbf{x}';t)\, d\mathbf{x}' \ . \tag{2}$$

II. SCALING FORMALISM

Our approach, which has been developed previously for the zero-field case, depends primarily on the observation that an energy- and time-dependent scaling factor commutes with noncollisional term on the right-hand side of (1):

$$s(H,t)\ \{H,f\} = \{H,s(H,t)f\} \ . \tag{3}$$

Thus, we define a *scaled* distribution function defined by

$$\hat{f}(\mathbf{x};t) \equiv s(H(\mathbf{x}),t)\ f(\mathbf{x};t) \ . \tag{4}$$

The principal advantage of multiplying by an energy-dependent factor arises from counting statistics. If we are interested in the distribution function in some region about the phase space point x, we consider as a function of time the number of simulation particles N in the vicinity of that point. The fractional error in $\hat{f}$ is then $1/\sqrt{N}$, which is the fractional error in $f \equiv s^{-1}\times\hat{f}$ as well. For regions of low phase-space density, N is proportional to $f(\mathbf{x};t)$. Thus, for example, to examine regions where the phase space density is down by seven orders of magnitude from the maximum, one needs many times of ten million simulation particles to keep the error from exceeding the estimate.

A standard solution to this problem — weighted EMC — defines distinct regions of phase space with different ratios of simulation points to actual particles (the ratio is constant within each region). When particles cross between regions, particle trajectories must be "split" (for multiple simulation) or decimated. If the density falls smoothly, there is no single efficient place to draw the boundary between regions, and many regions and multiple particle-splitting interfaces must be defined [3].

Our approach is based upon scaling the distribution function that is simulated by EMC, rather than upon a weighting the EMC simulation of an unscaled distribution function. The choice of s determines the trajectory density. In principle, these two approaches may be equivalent in particular cases. However, a weighted EMC eliminates the usual identification between individual initial particles and individual trajectories sampled (in what is the Monte Carlo integration of the Boltzmann equation). As a result, one loses the intuitive simplicity of regarding sampled trajectories as individual particles of a large ensemble. In a scaled EMC, on the other hand, one preserves a one-to-one correspondence of initial condition to trajectory, and it remains possible to regard the trajectories sampled as the actual paths of individual particles. As we describe below, however, in order to redistribute the statistical sampling weight, one pays the price that the trajectories do not follow the paths of ordinary particles.

III. TIME-DEPENDENT SCALING

Particle-number conservation imposes an important constraint on how $s(H(\mathbf{x}),t)$ is allowed to be chosen. By appropriate normalization, the total number N of simulation particles in the simulation of $\hat{f}$ is the integral of the scaled distribution function:

$$\hat{N}(t) = \int \hat{f}(\mathbf{x})\, d\mathbf{x} = \int s(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} \ . \tag{5}$$

If s is chosen to emphasize high-energy regions which have low density, then during a relaxation, thermalization will cause a transfer of (real) electrons to lower-energy regions where s is smaller. If s is not allowed a time-dependence to compensate, the total number of simulation particles of $\hat{f}$ must decrease — degrading the statistics in $\hat{f}$. Conversely, a heating mechanism would increase the number of simulation particles, improving statistical precision but possibly requiring undesirable computational expense. By allowing s to have a time-dependence, we accomplish in the time domain what the energy dependence of s accomplishes for the energy domain: reduce variations in particle number so that fractional errors can be kept at an acceptable level throughout the region of interest, with the least computational effort.

We chose a simple form for the joint energy-and-time dependence of s: we let them be *independent factors*. This can be written

$$s(H,t) = \exp\big(\alpha(t)+\gamma(H)\big) \ . \tag{6}$$

Furthermore, we let the function $\gamma(H) = -H \,/\, k_B T^{eff}$. This is appropriate for distributions f which are approximately characterized by effective temperatures below T^{eff} [Higher temperatures lead to normalization problems with $\hat{f}$.] This choice is also convenient computationally: a linear function γ implies that inelastic scattering rates (described below) are space-position-independent, and can be stored efficiently in look-up tables.

The time evolution of α is then defined specifically so that the total number of simulation points is constant.

$$\dot{\alpha} \ = \ \frac{-1}{\hat{N}} \int dx \ dx' \ G(x,x') \ \hat{f}(x';t) \ , \tag{7}$$

where

$$G(x,x') \equiv s(x) \ \Gamma^{I}(x,x') \ [s(x')]^{-1} \ = \ \exp\big[\gamma(H(x))-\gamma(H(x'))\big] \ \Gamma^{I}(x,x') \tag{8}$$

is essentially the unscaled total in-scattering rate, adjusted for the change in scaling factor experienced by a particle that undergoes a change in energy. (The I-subscript is explained below.) Equation (7) is a necessary condition for $\hat{f}$ to be simulable by scattering events alone, without independent generation and recombination (or other gain/loss) mechanisms. That it is sufficient is demonstrated constructively by finding the appropriate scattering rates.

Using (2), we find that $\hat{f}$ obeys a modified Liouville equation

$$\frac{\partial \hat{f}}{\partial t} \ = \ \{H,\hat{f}\} + \left(\frac{\partial \hat{f}}{\partial t}\right)_{coll} , \tag{9}$$

in which the "collision" or scattering term is defined by

$$\left(\frac{\partial \hat{f}}{\partial t}\right)_{coll} - \frac{1}{s}\left(\frac{\partial s}{\partial t}\right)\hat{f} = s\left(\frac{\partial f}{\partial t}\right)_{coll} . \tag{10}$$

Equations (2) and (10) specify completely the modified scattering term $(\partial \hat{f}/\partial t)_{coll}$. However, in order to implement a Monte Carlo time-evolution, one must determine scattering rates for the scaled problem which are analogous to the out-scattering rates $\Gamma(x;t)$ and the in-scattering rates $\Gamma(x,x';t)$ of the unscaled problem. There is some freedom in way this is done. One well-known degree of freedom is associated with self-scattering:

$$\left\{ \begin{array}{ll} \Gamma(x;t) & \rightarrow \Gamma(x;t) + \Delta\Gamma(x;t) \\ \Gamma(x,x';t) & \rightarrow \Gamma(x,x';t) + \Delta\Gamma(x;t) \ \delta(x-x') \end{array} \right\} , \tag{11}$$

where ordinarily $\Delta\Gamma$ is chosen to make the total out-scattering a positive constant. This is a kind of gauge transformation, in which the physically-significant total scattering rate is fixed, while unobservable components of the in- and out-scattering rates make the numerical implementation tractable. Another choice of gauge (the in-scattering gauge used in (8) and indicated by a superscript I) uses $\Delta\Gamma(x) = -\Gamma(x)$, to set the out-scattering to zero. In any case, the off-diagonal part of the out-scattering is a gauge invariant. Making some further transformations, we arrive at

$$\hat{\Gamma}(x,x') = G(x,x') + K(x) \ \frac{1}{\hat{N}} \hat{f}(x) \quad \text{for } x \neq x', \tag{12}$$

where

$$K(x) \equiv \int G(x',x) \ dx' \ . \tag{13}$$

The second term on the right-hand side of (12) leads to a kind of attractive interparticle scattering. This performs a rôle similar to that of trajectory iteration in weighted-EMC approaches: simulation particles entering critical regions are given greater weight, and are effectively caused to perform multiple traversals. However, in SEMC this weighting is implemented smoothly, rather than abruptly at the boundary of a region of interest, and it is accomplished with a fixed number of particles undergoing essentially ordinary scattering.

IV. SIMULATIONS

We have applied the SEMC technique to bulk GaAs semiconductor at 300 K. We used parameters (deformation potentials, phonon energies, band structure, etc.) that have been confirmed empirically in previous simulations [2]. The SEMC simulation used an effective temperature of 400 K. The system was allowed to reach a steady state under a uniform field. The figure below illustrates a typical result. The total particle density (Γ, L and X valleys) is plotted as a function of local energy (that is, for each particle, the energy is measured from the local position of the conduction band minimum). Results are shown for 70 kV/cm. The striking feature is the range of densities estimated—sixteen (16) orders of magnitude.

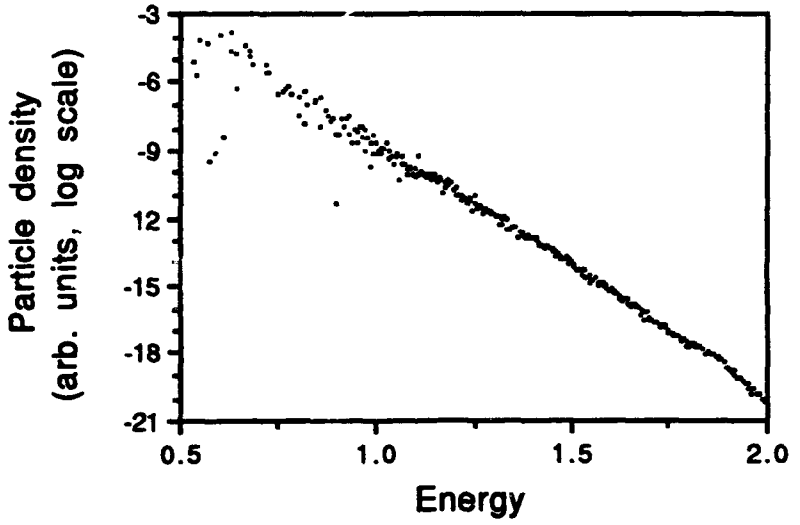


Fig. 1: Particle distribution function $f(E)$, including all valleys, as a function of local energy above the conduction band, in a 70 kV/cm field.

REFERENCES

[1] D. K. Ferry, *Semiconductors*, (Macmillan, 1991).

[2] M.-J. Kann, A. M. Kriman and D. K. Ferry, "Effect of Electron-Electron Scattering on Intervalley Transition Rates of Photoexcited Carriers in GaAs," Phys. Rev. B Vol. 41, No. 18, pp. 12659-12665, 15 June 1990.

[3] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, (Springer-Verlag, 1989).

USING THE RANDOM-K APPROXIMATION TO CONSISTENTLY INCORPORATE IMPACT IONIZATION INTO THE GENERALIZED EXPANSION METHOD FOR SOLVING THE MULTI-BAND BTE IN SILICON

Yu-Jen Wu and Neil Goldsman
Department of Electrical Engineering, University of Maryland, College Park, MD 20742

Abstract

We present here a consistent approach for determining impact ionization coefficients for a multi-band model in silicon. Using first-order perturbation theory and a random-k approximation, the impact ionization rate is determined to reflect the multi-band density of states in silicon. To account for the actual density of states, we have solved four coupled Boltzmann transport equations by combining a generalized Legendre polynomial expansion method with numerical techniques using finite differences and sparse matrices. Calculated values for the impact ionization coefficients agree very well with experiment for electrons in silicon, while being obtained in significantly less CPU time than required by analogous Monte Carlo calculations.

I. INTRODUCTION:

Fundamental analytical work on impact ionization was performed by Keldysh [1]. Since Keldysh presented his expression for impaction ionization rate, several analytical or quasi-analytical investigations of impact ionizations have been performed which use Keldysh's results [2, 3]. These investigations have provided considerable insight into impact ionization, and the nature of Boltzmann equation in semiconductors. However, in the past, analytical formulations were usually constrained to an energy range which is limited to silicon's first conduction band, and therefore did not reflect silicon's true density of states. Furthermore, while agreement with experiment has been obtained, the use of Keldysh's formulation for silicon has not been fully justified since, from a theoretical point of view, Keldysh's expression was derived for a single parabolic conduction band, which is not strictly appropriate for silicon.

In this work, we extend these quasi-analytical investigations to a larger energy range which reflects silicon's actual density of states. Furthermore, instead of using the Keldysh expression, we derive an expression for impact ionization rate which is consistent with the actual density of states in silicon. The expression for impact-ionization is based on the random-k formalism, initially introduced by Kane [4]. Once an expression for impact ionization is obtained, we then use it, in conjunction with a quasi-analytical method of solving the Boltzmann transport equation, to calculate impact ionization coefficients.

II. ANALYTICAL FORMULATION:

A. Impact Ionization Rate:

To calculate impact ionization rates which accurately reflect the density of states in silicon, we begin with the random-k approximation presented by Kane [4]. With this formulation, Kane found that if he required energy to be conserved, but neglected conservation of crystal momentum, he obtained virtually the same results for impact ionization rates as when he required both energy and crystal momentum be conserved. In other words, Kane found that the energy dependent impact ionization rate could be expressed accurately in terms of the density of states. Mathematically, this energy-conserving, random-k approximation gives the following expression for impact ionization:

$$\frac{1}{\tau_{ii}(\varepsilon)} = A \sum_{n_2,n_3,n_4} \int \rho_{n_2}(\varepsilon_2)\rho_{n_3}(\varepsilon_3)\rho_{n_4}(\varepsilon_4)\delta(\varepsilon + \varepsilon_4 - \varepsilon_2 - \varepsilon_3)d\varepsilon_2 d\varepsilon_3 d\varepsilon_4 \quad (1)$$

where ε_4, n_4 refer to the energy and band containing the valence band electron; ε_2, n_2 and ε_3, n_3 are the energies and bands of the final electron states; $\rho_{n_i}(\varepsilon)$ is the density of states function in band n_i; and A is an ionization coupling constant which can be determined empirically.

The sum over n_2, n_3, and n_4 in equation (1) accounts for permutations of possible final energy bands allowed by energy conservation. The range of the summations is determined by the band structure and the conservation of energy. To evaluate equation (1), we take values for the density of states functions, $\rho_i(\varepsilon_i)$, from the effective, multi-band silicon dispersion relation given by Brunetti *et. al.* [5]. This band structure averages the actual dispersion relation over spheres, thereby providing the actual density of states. By accurately providing the density of states, this averaging is consistent with the use of the random-k approximation.

After making the approximation that holes are generated mostly at the top of the valence band, and inserting limits consistent with the spherical band model, equation (1) becomes

$$\frac{1}{\tau_{ii}(\varepsilon)} = A\int_0^{\varepsilon-\varepsilon_g-1.75}\rho_1(\varepsilon_2)\rho_2(\varepsilon-\varepsilon_g-\varepsilon_2)d\varepsilon_2 + A\int_{\varepsilon-\varepsilon_g-1.75}^{1.75}\rho_1(\varepsilon_2)\rho_1(\varepsilon-\varepsilon_g-\varepsilon_2)d\varepsilon_2$$
$$+A\int_{1.75}^{\varepsilon-\varepsilon_g}\rho_2(\varepsilon_2)\rho_1(\varepsilon-\varepsilon_g-\varepsilon_2)d\varepsilon_2 \qquad for \quad \varepsilon-\varepsilon_g \geq 1.75 \tag{2}$$

$$\frac{1}{\tau_{ii}(\varepsilon)} = A\int_0^{\varepsilon-\varepsilon_g}\rho_1(\varepsilon_2)\rho_1(\varepsilon-\varepsilon_g-\varepsilon_2)d\varepsilon_2 \qquad for \quad \varepsilon-\varepsilon_g < 1.75 \tag{3}$$

where $\varepsilon_g = 1.12eV$ is the energy gap for silicon.

B. Impact Ionization Coefficients:

To determine the impact ionization coefficients, we used the following expression which accounts for the density of states in higher bands:

$$\alpha_n = \sum_{i=1}^{4}\frac{1}{v_d}\int_0^{\infty}\frac{1}{\tau_{ii}(\varepsilon)}f_0^{(i)}(\varepsilon)\rho^{(i)}(\varepsilon)d\varepsilon \tag{4}$$

where $f_0^{(i)}(\varepsilon)$ is the isotropic distribution function in band i; $\rho^{(i)}(\varepsilon)$ is the density of states function in the i^{th} band; and v_d is the silicon saturation velocity.

Before we can calculate the impact ionization coefficients, we must first solve the Boltzmann transport equation (BTE) to find $f_0^{(i)}(\varepsilon)$.

C. Formulation of the BTE:

To obtain $f_0^{(i)}(\varepsilon)$, we use the Legendre polynomial (LP) expansion method for solving the BTE. Previously, investigators used a state of the art LP approach for impact-ionization studies which incorporated a first order LP expansion that accounted for one or two conduction bands in Si [6]. In this work, we use a generalized approach for obtaining the distribution function to arbitrarily-high LP order. We also include the effects of silicon's higher conduction bands.

To solve the BTE to arbitrarily high order and for densities of states which correspond to higher conduction bands, we first write a steady-state, homogeneous Boltzmann equation for each of the four bands of the spherically averaged silicon dispersion relation [5]. We also include the effects of acoustic, intervalley, interband phonon scattering, and impact ionization. The equation for a specific band is coupled to the other equations through impact ionization and interband scattering.

Next, with the LP approach, we expand the distribution function in each band, $f^{(i)}(\vec{k})$, in terms of Legendre polynomials to an arbitrarily high order:

$$f^{(i)}(\vec{k}) = \sum_{n=0}^{\infty} f_n^{(i)}(\varepsilon)P_n(cos\theta) \tag{5}$$

where θ is the angle between $\vec{k}$ and electric field $\vec{E}$; $P_n(cos\theta)$ are the Legendre polynomial basis functions which provide the angular dependence of the distribution function; and $f_n^{(i)}(\varepsilon)$ represent the unknown coefficients of the basis functions.

Once the distribution function has been expressed in terms of the LP expansion, the goal is to find the unknown coefficients $f_n^{(i)}(\varepsilon)$. We find the coefficients by using a variation of the generalized LP expansion approach [7]. With this method, LP orthogonality and recurrence relations are used to generate a system of equations for the unknown coefficients, $f_n^{(i)}(\varepsilon)$. The powerful aspect of this technique is that the equations for all the coefficients, $f_n^{(i)}(\varepsilon)$, have identical forms and are thus automatically generated to arbitrarily high order. In addition, in order to reduce the number of equations required for solution, and also improve upon numerical characteristics, we substituted odd-indexed LP equations into even-indexed LP equations and reformulated the Boltzmann equation into a system of second-order, linear, differential-difference equations. The analytical form of the generalized equation for even-indexed Legendre coefficients $f_n^{(i)}(\varepsilon)$ in band i, can be written as:

$$
\begin{aligned}
&-\frac{2(eE)^2}{m_i^*}\left\{\alpha_{n-1}\alpha_{n-2}\left[\frac{\tau_i\gamma_i}{\gamma_i'^2}\frac{\partial^2 f_{n-2}^{(i)}(\varepsilon)}{\partial\varepsilon^2}+\left(\frac{\tau_i'\gamma_i}{\gamma_i'^2}-\frac{\tau_i\gamma_i\gamma_i''}{\gamma_i'^3}-(n-2)\frac{\tau_i}{\gamma_i'}\right)\frac{\partial f_{n-2}^{(i)}(\varepsilon)}{\partial\varepsilon}\right.\right.\\
&\quad\left.-\frac{n-2}{2}\left(\frac{\tau_i'}{\gamma_i'}-\frac{n}{2}\frac{\tau_i}{\gamma_i}\right)f_{n-2}^{(i)}(\varepsilon)\right]+\left[\alpha_{n-1}\alpha_n\left(\frac{n}{n+1}\right)+\alpha_n\alpha_{n+1}\left(\frac{n+1}{n+2}\right)\right]\\
&\quad\times\left[\frac{\tau_i\gamma_i}{\gamma_i'^2}\frac{\partial^2 f_n^{(i)}(\varepsilon)}{\partial\varepsilon^2}+\left(\frac{\tau_i'\gamma_i}{\gamma_i'^2}-\frac{\tau_i\gamma_i\gamma_i''}{\gamma_i'^3}+\frac{3}{2}\frac{\tau_i}{\gamma_i'}\right)\frac{\partial f_n^{(i)}(\varepsilon)}{\partial\varepsilon}\right]\\
&\quad+\alpha_{n-1}\alpha_n\left(\frac{n}{2}\right)\left(\frac{\tau_i'}{\gamma_i'}-\frac{n}{2}\frac{\tau_i}{\gamma_i}\right)f_n^{(i)}(\varepsilon)-\alpha_n\alpha_{n+1}\left(\frac{n+1}{n+2}\right)\frac{n}{2}\left(\frac{\tau_i'}{\gamma_i'}+\frac{n+1}{2}\frac{\tau_i}{\gamma_i}\right)f_n^{(i)}(\varepsilon)\\
&\quad+\alpha_{n+1}\alpha_{n+2}\left(\frac{n+1}{n+3}\right)\left[\frac{\tau_i\gamma_i}{\gamma_i'^2}\frac{\partial^2 f_{n+2}^{(i)}(\varepsilon)}{\partial\varepsilon^2}+\left(\frac{\tau_i'\gamma_i}{\gamma_i'^2}-\frac{\tau_i\gamma_i\gamma_i''}{\gamma_i'^3}+(n+3)\frac{\tau_i}{\gamma_i'}\right)\frac{\partial f_{n+2}^{(i)}(\varepsilon)}{\partial\varepsilon}\right.\\
&\quad\left.\left.+\frac{n+3}{2}\left(\frac{\tau_i'}{\gamma_i'}+\frac{n+1}{2}\frac{\tau_i}{\gamma_i}\right)f_{n+2}^{(i)}(\varepsilon)\right]\right\}\\
&=\begin{cases}\left[\frac{\partial f_0^{(i)}(\varepsilon)}{\partial t}\right]_{coll} & for\ \ n=0\\ -\frac{f_n^{(i)}(\varepsilon)}{\tau_i(\varepsilon)} & for\ \ n=2,4,6,8,\ldots,\infty\end{cases}
\end{aligned}
\tag{6}
$$

where $\alpha_n=\frac{n+1}{2n+1}$; γ_i' and γ_i'' are the 1st and 2nd derivatives of γ_i with respect to energy; $\tau_i(\varepsilon)$ is the total mean free time between collisions; and τ_i' is the 1st derivative of τ_i with respect to energy.

A similar system of equations is obtained for each band. To solve the system, we first truncate the expansion. Then, the entire system is discretized and then solved directly using Gaussian elimination for sparse matrices.

III. RESULTS:

In Fig. 1 we show the distribution function we obtained from solving the multi-band BTE. The solid line shows calculated values for the symmetrical part of the distribution function when 20 LP terms were used, while the dotted line shows the results of using only 2 LP terms; the diamonds give the results of Monte Carlo calculations which used the same Brunetti transport model. Solving the BTE to 20'th LP order required approximately 28 CPU seconds on a SUN4 workstation, while comparable Monte Carlo calculations took considerably longer.

Using the above method, we were able to obtain values for ionization coefficients with the use of only one unknown adjustable parameter, the ionization coupling factor A. The value of A has

been determined to be 45000 when an expansion of 20 LP terms was employed; and 55000 when only 2 LP terms were used. Fig. 2 demonstrates the agreement between the values obtained from our calculations and the experimental data of Lee *et. al.* [8] for impact ionization coefficients.

IV. CONCLUSION:

We have developed an efficient and physics-based method in obtaining impact ionization coefficients for the multi-band model in silicon. The impact ionization rate is derived from basic principles with a suitable random-k approximation for the multi-band model considered. A generalized Legendre polynomial expansion method is combined with numerical techniques to solve the system of four coupled multi-band Boltzmann equations. Excellent agreement has been obtained between our calculations and the results of both Monte Carlo method and experiment. The efficiency of this method makes it well suited for studying impact ionization and hot-electron phenomena in semiconductor devices.

V. ACKNOWLEDGMENT:

The authors are grateful to the Semiconductor Research Corporation and the National Science Foundation for supporting this project.

VI. REFERENCES:

[1] L. Keldysh, *Soviet Physics JETP*, Vol. 21, No. 6, pp. 1135-1144, 1965.
[2] Y.-Z. Chen and T.-W. Tang, *J. of Applied Physics*, Vol. 65, No. 11, pp. 4279-4286, 1989.
[3] N. Goldsman, Y. Wu and J. Frey, *J. of Applied Physics*, Vol. 68, No. 3, pp. 1075-1081, 1990.
[4] E. Kane, *Physical Review*, Vol. 159, No. 3, pp. 624-631, 1967.
[5] R. Brunetti and C. Jacoboni, *Solid-State Electronics*, Vol. 32, No. 12, pp. 1663-1667, 1989.
[6] A. Gnudi *et. al.*, *IEEE Trans. Comput.-Aided Design*, Vol. 12, No. 11, pp. 1706-1713, 1993.
[7] K. Hennacy and N. Goldsman, *Solid-State Electronics*, Vol. 36, No. 6, pp. 869-877, 1993.
[8] C. Lee *et. al.*, *Physical Review*, Vol. 134, No. 3A, pp. A761-A773, 1964.

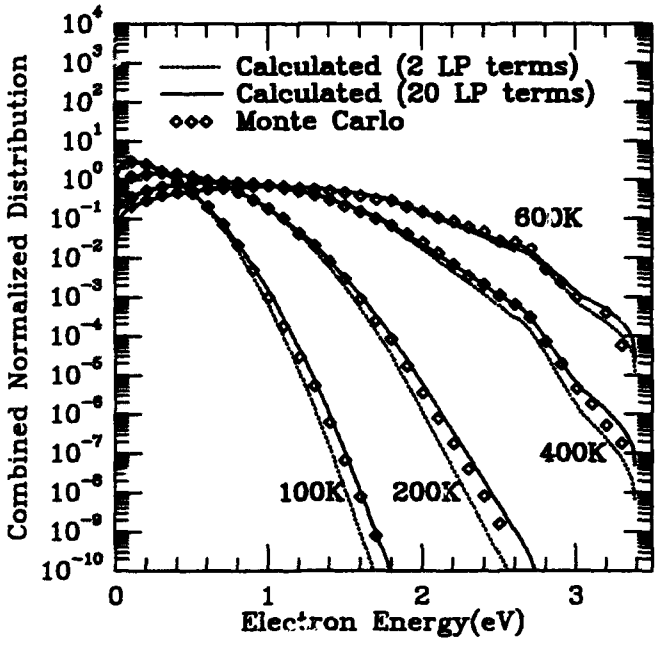


Fig. 1. The isotropic distribution function f_0 plotted for different electric fields are compared with results of MC calculations.

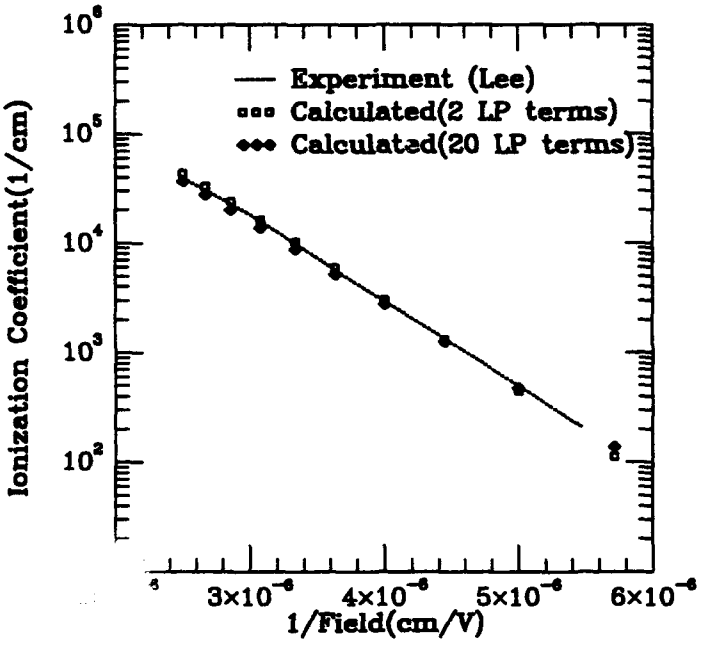


Fig. 2. Impact ionization coefficient is plotted versus the reciprocal electric field and compared with experimental values.

INVESTIGATION OF THE DIFFUSION INTO Si OF Au VIA FRANK-TURNBULL MECHANISM BY MONTE CARLO SIMULATION*

Q. S. Zhang, J. A. Van Vechten, and T. K. Monson
Center for Advanced Materials Research
Department of Electrical and Computer Engineering
Oregon State University, Corvallis, OR 97331

ABSTRACT

VIDSIM[1] (Vacancy and Interstitial Diffusion Simulator), a program for Monte Carlo simulation of point defect diffusion and interaction in diamond and zinc-blend crystal structures has been developed and used to carry out some fundamental studies in Si and AlGaAs. We report here the latest result of a computer simulation designed to examine the diffusion into Si, in the presence of vacancy (V) clusters, of Au or similar transition metals that migrate mainly as interstitials but reside as substitutional impurities. Frank-Turnbull mechanism is assumed in the diffusion process. Also, we investigate the effect of vacancy (V) clusters distributed in the sample on the profile. Au distribution of both interstitials (Au_i) and substitutionals (Au_{Si}) are presented.

I. INTRODUCTION

People have been showing great interest in doing research on diffusion in semiconductors because of its importance in industry as well as in science. A variety of models for point defect migration, interaction and reaction have been proposed and a great amount of theoretical work has been done to account for, in one way or another, experimental results or observations. Yet, there is much controversy over the validity of the models or mechanisms, such as kick-out (KO)[2] and Frank-Turnbull (FT)[3] mechanism, because we believe, there have been no rigorous connections between atom level assumptions and macroscopic consequences prior to our simulation program. For most differential equations which describe the diffusion processes, there exist no analytical solutions.

VIDSIM has been developed to resolve the relation between atom level assumptions and macroscopic consequences. It assumes no differential equation. It uses an atom level Monte Carlo algorithm and involves an enormous amount of computation to obtain statistically significant results. The program models the evolution of a user-defined initial set of defects in space and time for any mechanisms the user chooses to simulate. The complex relationship between assumptions and consequences is thus obtained without any of the approximating passage to limits that have confused the previous literature. We have previously published direct simulations of the KO mechanism, for the set of assumptions advocated by those who suggest it accounts for Au diffusion into Si, which showed that the true result is very different from that expected by those advocates[4,5].

We notice[5,6] that much needs to be understood in the diffusion process involving vacancies as well as clusters of them. We report here our latest work on a computer simulation designed to examine the diffusion into Si of Au via Frank-Turnbull mechanism in the presence of vacancy clusters. Results are applicable to Cu or Pt or similar transition metals, which, like Au, migrate mainly as interstitials but reside primarily as substitutional impurities.

---

* Supported in part by Komatsu Silicon USA, Inc. and by SEH America, Inc. and previously by the Air Force Office of Scientific Research and by Wacker Siltronic.

II. SIMULATION SET-UP

This particular simulation assumes that no Si self interstitials, Si_i's are initially present and none can be produced during the diffusion process, that there are vacancies initially present as clusters in the bulk and that single vacancies can be created and annihilated at the surfaces normal to the x direction. It also assumes that a source of Au_i is on the normal surface of the positive x boundary where Au_i will be annihilated if it returns there, and a reflexive boundary condition for Au_i is imposed on the far side, i.e., the negative x boundary. While vacancies are generated at a rate of $1.84x10^{-9}$ per surface site per attempt period (a.p.), which, in Si, is estimated to be 74 fs, the zone boundary phonon period, the Au_i source injects Au_i on the surface at a rate of $1.84x10^{-9}$ per attempt period per surface site or 184 injections (on average) per diffusion step $r \equiv 5x10^6$ a.p.. The sample simulated is 2000 a thick in x direction and 100 a both in the y and z directions, where a is the lattice constant for the material of interest. Periodic boundary conditions are imposed on the y and z directions. Initially the sample is perfect except for 19 clusters of 5 V's (V_5) each spaced at 100 a intervals in the x direction through the bulk of the sample, which corresponds to a concentration $[V] = 3.0x10^{16}$ cm^{-3}, a conservative estimated of [V] at the melting temperature based on the positron lifetime results reported by Dannefaer et al[7].

We assert that the above conditions are appropriate for the case that Au-Si molten alloy has formed on one side and that an amorphous thermal oxide layer, with no molten alloy, is present on the far side.

We collect an ensemble of 20 of the sample runs of simulation, each of which extends to 240 steps of $5x10^6$ a.p. i.e., $12x10^9$ a.p.. There were a total of about $6.25x10^{10}$ atomic hops in the simulation. Of the $8.75x10^5$ Au atoms injected, 1289 are retained as Au_{Si} and 224 are retained as Au_i at the end. Of $1.75x10^6$ vacancies injected, 3075 are retained. Note that less than 0.2% of the defects that are introduced survive at the end of this simulation.

III. RESULTS OF COMPUTER SIMULATION AND DISCUSSIONS

We show in Table 1 the rates at which vacancy clusters of various sizes emit V's and diffuse for a particular set of atom level assumptions[1].

We observe that, for the simulation of clusters of type V_5, most of the original clusters break up into single vacancies after 4 r simulation time; they do not migrate a significant fraction of the 100 a initial minimum spacing between them before they totally disperse. Those scattered V's from the original clusters join newly diffusing-in V's from the surfaces and can form new clusters of various sizes. Simulation shows that there are clusters of size of 6, 7, 8 and 9 that form spontaneously from single vacancies during the process. We notice that there are "magic numbers" for which small clusters are particularly stable, e.g., n = 6. Some of these also diffuse very slowly. We suppose that small clusters of such size account for the 450 ps signal observed in many, but not all positron annihilation spectra[8]. We also study the sample with a 17 V (V_{17}) cluster initially present and find that the simulation requires a great deal of commuter effort because there is so much atomic diffusion around the surface of the cavity. Having simulated a total time of 20 r, we have observed no emission of any V from such a cluster. Thus, we conclude that the mean time to emit a V exceeds 5 r. These clusters also diffuse very slowly. We suggest that such large

Table 1. Diffusivity and mean time to emit one or more vacancy from clusters of n vacancies. Time unit $r \equiv 5x10^6$ a.p.. These values result from the parameters assumed[1].

n=	1	2	3	4	5	6	7	8	17
$D_n(a^2/r)$	1177	594	559	419	236	4.0	47	171	1.5
E_n (r)	--	0.66	0.39	0.41	0.35	11.3	3.25	2.25	> 5

clusters correspond to the 1.5 ns signal in all positron annihilation spectra and for the D defects that are so difficult to anneal out[9].

We observe that more vacancy clusters of various sizes larger than divacancies are found to be on the far side of the sample, which might result from the asymmetric Au injection that gives rise to more break-up of clusters by Au_i near the source side. For example, referring the vacancy distribution at step 160, as shown in Fig.1, we see a peak at the location somewhere around 1650 a, i.e., about 350 a from the far surface. Investigation shows that the vacancy clusters, esp. of size 6, (V_6) which is surprisingly stable, are responsible for the sharp rise in the profile. Distribution of V_6 at several steps are shown in Fig. 2.

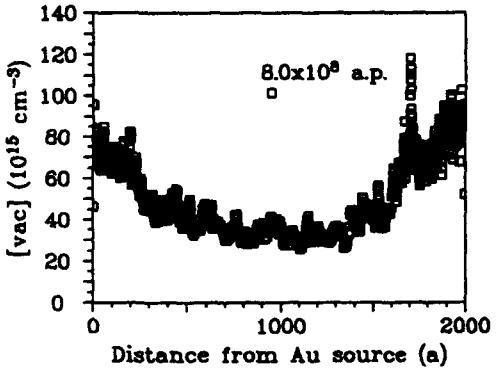


Fig. 1. The profile of vacancy concentration in the sample at step 160.

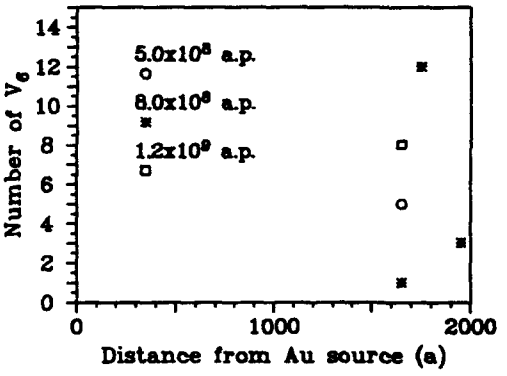


Fig. 2. Vacancy clusters of type V_6 observed at step 100, 160 and 240.

In Fig. 3 we show the profile of divacancies, $[V_2]$, which demonstrates a U shape. Since the [V](x) for single vacancies is not uniform, as shown in Fig. 1, some variation in $[V_2]$ as predicted from the Law of Mass Action if "local equilibrium" is assumed. This would imply $[V_2]$ should vary as $([V])^2$. Also, it is obvious that $[V_2]$ profile should dip at the sample surfaces because annihilating boundary conditions for vacancies are imposed. Fig. 4 shows a profile of Au_{Si} for t = $1x10^8$, $2x10^8$, $4x10^8$, and $1.2x10^9$ a.p., and Fig. 5 demonstrates the distribution of Au_i concentration through the sample at the same time period as in Fig. 4.

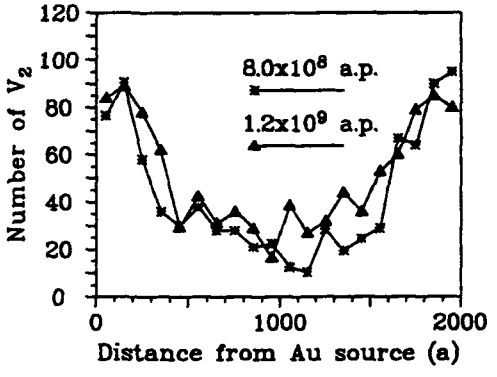


Fig. 3. The profile of divacancies, V_2 in the sample, observed at step 100 and 240.

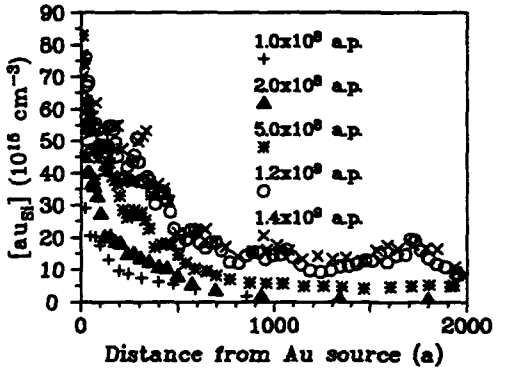


Fig. 4. The profile of Au substitutionals in the sample, observed at step 20, 40, 80 and 240.

From the Fig. 4 and Fig. 5, we see the distribution of Au_i is similar to that of Au_{Si} and not much more uniform despite the fact that the Au_i have a diffusivity of 22,894 $a^2 r^{-1}$, 19 times the diffusivity of the V's and dramatically larger than the effective diffusivity of Au_{Si}'s. The ratios of concentrations on both surfaces for both Au_{Si} and Au_i are observed to be about the same, which is similar to the result we reported when we simulated Au diffusion into Si assuming KO model[3,4]. In simulating KO model, we allowed no V's and made the Si_i's diffuse less rapidly than the Au_i's. However, there is no rise of $[Au_{Si}]$

near the far surface. The results shown in Fig. 4 are contrary to the assertions or assumptions of many previous investigators, particularly those who advocate the KO mechanism. In contrast to the monotonic distribution of Au_i shown in Fig. 4, they assert that since the Au_i diffuses so rapidly, a uniform distribution of Au_i will be established from the very beginning of the diffusion. We also note that the $[Au_i]$ rises with $[Au_{Si}]$ at the far side and the profile of $[Au_i]$ tends to be flat at the far side as simulation time elapses. The flatness of $[Au_i](x)$ near the far end is a result consistent with the reflexive boundary conditions used in our simulation.

In Fig. 6 we show the variation of $[Au_{Si}]$ in the center versus the square root of time. The trend shown is not exactly that as reported in previous literature. Two distinctions are found. First, the best fit of linear regression shows that the best straight line does not pass through the origin, as the previous authors'

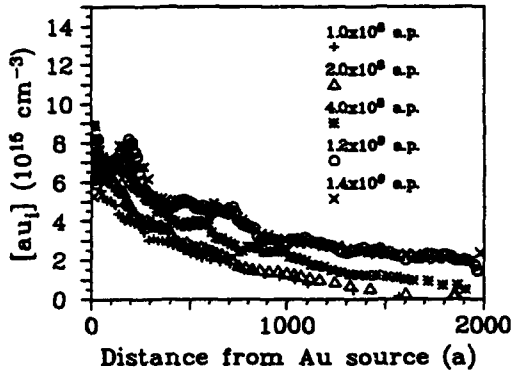


Fig. 5. The profile of Au interstitials in the sample, observed at step 20, 40, 80 and 240.

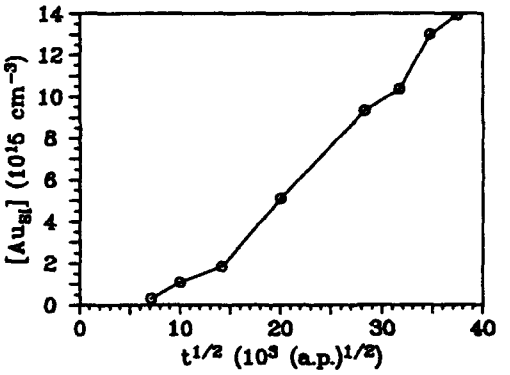


Fig. 6. The variation of Au substitutionals in the center of the sample versus the square root of diffusion time.

claim, but reaches $[Au_{Si}] = 0$ at finite positive time, which is necessary for Au_i to reach the center of the sample. Secondly, although the longer time data and the shorter time period can be fitted to a line, there is an inflection point that was not reported previously. To some extend, the simulation result is consistent with the result of our laboratory experiments[10].

REFERENCES

1. U. Schmid, N. C. Myers, and J. A. Van Vechten, Comp. Phys. Commun. 58, 329(1990).
2. A. Seeger, and W. Frank, J. Electron. Mater. 14a, 159 (1985).
3. F. C. Frank and D. Turnbull, Phys. Rev. 104, 617 (1956).
4. U. Schmid, J. A. Van Vechten, N. C. Myers, and U. Koch., Mat. Res. Soc. Symp. Proc. 163, 609 (1990).
5. J. A. Van Vechten, U. schmid, and Zhang Q. S., J. Electron. Mater 20, 431, (1991).
6. T. K. Monson, J. A. Van Vechten and Q. S. Zhang, Phys. Rev. B49, 2972(1994).
7. S. Dannefaer, T. Bretagnon, K. Abdurahman, D. Kerr, an D. Hahn, Mat. Res. Soc. Symp. Proc., 262, (1992).
8. P. J. Roksnoer and M. M. B. Van den Boom, J. Cryst. Growth 53, 563 (1981).
9. S. Dannefaer, P. Mascher, and D. Kerr, Phys. Rev. Lett. 56, 2195 (1986)
10. R. K. Graupner, J. A. Van Vechten, P. Harwood, and T. K. Monson, Mat. Res. Soc. Symp. Proc. 262, (1992).

ADVANCES IN MULTI-DIMENSIONAL TCAD

W. Fichtner
Integrated Systems Laboratory, Swiss and Federal Institute of Technology
and
ISE integrated Systems Engineering AG
Gloriastrasse 35
CH-8092 Zürich
Switzerland

Over the past few years, numerical process and device simulation have become increasingly popular in academic and industrial environments. Analogous to other fields such as biochemistry, petroleum and aerospace engineering, and weather forecasting, the use of software tools offers many advantages over the classical experimental approach.

Modern computing environments such as the latest generation of workstations permit simulations that were thought to be impossible before. Based on these impressive advances in computer hardware, TCAD (Technology Computer Aided Design) is one of the premier candidates for a virtual reality approach towards understanding, optimizing and predicting new process technologies, devices and circuits.

The push towards higher device density, smaller active feature sizes and more advanced fabrication steps puts stringent requirements on the quality, robustness and accuracy of TCAD software tools. The enormous complexity of todays micro- and opto-electronic devices has generated a strong need for better modeling support, especially in two and three dimensions.

In this presentation, a survey of the present multi-dimensional process and device simulation tools will be given. Particular emphasis will be laid upon structure and grid generation, the impact of numerical methods and computer architectures, and the need for parallel processing to tackle the problems of tomorrow.

Three-Dimensional Hydrodynamic Simulation of Submicron MOSFET's

Rajiv Madabhushi, Qi Lin, Gwo-Chung Tai, Neil Goldsman and Isaak D. Mayergoyz
Department of Electrical Engineering, University of Maryland, College Park, MD, 20742

Abstract

This paper presents a method for solving the 3-D hydrodynamic (HD) model in submicron semiconductor devices. The main features of this method are the fairly low memory and CPU time requirements, and excellent convergent property. Simulation results of a 3-D submicron MOSFET are provided.

I. Introduction

As device dimensions continue to shrink, 3-D-related phenomena, including the MOS bird's beak and narrow channels effects, can significantly impact device characteristics. As a result, demand for 3-D simulations has appeared[1,2]. As far as we know, very few robust 3-D device simulators exist, and we know of none which are based on the hydrodynamic model. In general, the HD model is difficult to solve because it is a highly nonlinear, singularly perturbed, highly coupled system of partial differential equations. While these difficulties are readily evident in 2-D, they are compounded when 3-D simulations are attempted. In addition to the standard obstacles of traditional 2-D simulations, the increased difficulties for 3-D simulation can be summarized as follows: (1) Memory: 3-D simulation of electrons and holes requires the solution of between 10^5 and 10^6 simultaneous discrete equations. Use of the standard Newton's method requires storing extremely large matrices to accomodate this large number of discrete equations. (2) Algorithm: Solving such large matrices is very difficult and readily leads to round-off errors and instabilities. (3) Boundary conditions in 3-D are more complicated than their 2-D counterparts, which leads to changes in the matrix structure and more difficulties when nonplanar surfaces are encountered. (4) Convergence and stability: More complicated coupling relations between mesh points and between equations significantly increase the nonlinearities already associated with the HD model.

We have overcome numerical problems associated with 3-D HD modeling, and developed an efficient 3-D HD simulator for predicting deep-submicron MOSFET performance. Instead of using solution techniques, such as Newton's method or the conjugate gradient approach, we adapt our solution approach specifically for 3-D HD device simulation. This required extending our highly stable and routinely convergent method for 2-D HD modeling to 3-D[3,4],

To tailor our approach specifically for the 3-D problem, we use a fixed point iterative approach which totally avoids solving large matrices. By defining new Slotboom-like variables for the HD model, we transform the original HD equations to self-adjoint form which guarantees that each discretized HD equation will correspond to a diagonally dominant matrix. A Scharfetter-Gummel-like (S-G) discretization is then performed on the self-adjoint forms of the current-continuity and energy balance equations. The S-G-like schemes resolve rapid variations of unknown variables, not only due to the mesh refinement, but due to the special design of the finite difference schemes as well, thereby helping to reduce the number of mesh-points, equations, and time to convergence. The coupled system is solved using SOR-type methods where the equation for each mesh point is updated explicitly. The explicit method obviates the need

to solve large matrices, and, due to the diagonal dominance, each HD equation is guaranteed to converge[5,6]. The explicit method is independent of matrix structure so implementing additional 3-D related boundary conditions do not noticeably affect the solution. Furthermore, memory requirements of the fixed-point iteration method are fairly low, thereby facilitating solution in 3-D. Finally, the explicit method is intrinsically parallel.

II. Solution of 3-D HD Equations

The Self-Adjoint form of the HD model

To solve the HD model in 3-D, we begin with the standard HD equations[7]. By using the Slotboom-like variables variables u, v, g_n, g_p for the standard HD variables n, p, T_n, T_p, the electron current density and electron energy flux can be expressed in the compact form:

$$\vec{J_n} = D_n n_i \exp\left(-\frac{T_n - T_L}{T_L} + \frac{q\phi}{k_B T_L}\right) \nabla u, \tag{1}$$

$$\vec{S_n} = -\kappa_{cn} \exp(\psi_n/a_T) \nabla g_n + \frac{\vec{J_n}}{-q}(\frac{1}{2} m_n^* v_{dn}^2). \tag{2}$$

$$u = \exp\left(-\frac{q\psi_n}{k_B T_L}\right), \quad T_n = g_n \exp(\psi_n/a_T), \quad a_T = \frac{\kappa_{cn}}{\frac{5}{2}\mu_n n k_B} = \frac{4}{5}\frac{k_B T_L}{q} = 0.0207V \tag{3}$$

By substituting above expressions into the original HD equations, one can transform the HD model into self-adjoint form:

$$\nabla^2 \phi = \frac{qn_i}{\epsilon_s}\left(u\exp(-\frac{T_n - T_L}{T_L} + \frac{q\phi}{k_B T_L}) - v\exp(-\frac{T_p - T_L}{T_L} - \frac{q\phi}{k_B T_L})\right) - \frac{qD}{\epsilon_s} \tag{4}$$

$$\nabla \cdot \left(D_n n_i \exp(-\frac{T_n - T_L}{T_L} + \frac{q\phi}{k_B T_L}) \nabla u\right) = R(\phi, u, v) \tag{5}$$

$$\nabla \cdot (\kappa_{cn} \exp(\psi_n/a_T) \nabla g_n) = n\frac{\frac{3}{2}k_B g_n \exp(\psi_n/a_T) + \frac{1}{2}m_n^* v_{dn}^2 - \omega_.}{\tau_{n\omega}} - \vec{J_n} \cdot \vec{E} - \nabla \cdot \left(\frac{\vec{J_n}}{q}\frac{1}{2}m_n^* v_{dn}^2\right) \tag{6}$$

It is clear from the above expressions that the Poisson, the current-continuity, and the energy-balance equations are each self-adjoint differential equations with respect to the variables ϕ, u and g_n.

Iterative Method for HD Equations

We use an S-G-type method to discretize the current-continuity and energy-balance equations. This S-G approach helps to analytically resolve the rapid variations in n and T_n, thereby reducing the number of mesh-points which can become quite large for 3-D applications. With the S-G-like discretization, we assume that S_n and J_n are constant between mesh-points. Integration of the above self-adjoint forms between the mesh-points is then readily performed. The overall variation of S_n and J_n is then accounted for discretely on the mesh-points. This discretization yields the following 3-D general expression for the current-continuity and energy-balance equations:

$$c_{i+1,j,k}H_{i+1,j,k} + c_{i-1,j,k}H_{i-1,j,k} + c_{i,j+1,k}H_{i,j+1,k}$$
$$+c_{i,j-1,k}H_{i,j-1,k} + c_{i,j,k+1}H_{i,j,k+1} + c_{i,j,k-1}H_{i,j,k-1}$$
$$-(c_{i+1,j,k} + c_{i-1,j,k} + c_{i,j+1,k} + c_{i,j-1,k} + c_{i,j,k+1} + c_{i,j,k-1} + L_{i,j,k})H_{i,j,k} = -\gamma_{i,j,k} \quad (7)$$

In the above equation $H_{i,j,k}$ represents the discrete form of the Slotboom variable either u or g_n. The $c's$ are the discretization coefficients corresponding to each HD equation. $L_{i,j,k}$ represents the diagonal term arising from the RHS of the self-adjoint equations, while $\gamma_{i,j,k}$ represents the 'constant' term on the RHS of eqns.(5) and (6). It is interesting to note that the discretization of the self-adjoint forms yields coefficient matrices which are diagonally dominant. This can be observed since all the coefficients $c_{i,j,k}$, as well as $L_{i,j,k}$, have the following property:

$$c_{i+1,j,k} \geq 0, \quad c_{i-1,j,k} \geq 0, \quad c_{i,j+1,k} \geq 0, \quad c_{i,j-1,k} \geq 0, \quad c_{i,j,k+1} \geq 0, \quad c_{i,j,k-1} \geq 0, \; L_{i,j,k} \geq 0. \quad (8)$$

Eqn.(7) represents a system of $N = (N_i \times N_j \times N_k)$, where N_i, N_j, N_k represent the number of meshpoints in each dimension, respectively. Ordinarily, such a system would be solved implicitly using a Newton-type approach. However, such an approach would yield extremely large matrix equations which have extensive memory requirements and are susceptable to round-off error.

Our approach, which has been specially tailored for hydrodynamic 3-D applications, allows for eqn.(7) to be solved explicitly, thereby avoiding large matrices entirely. We solve eqn.(7) for $H_{i,j,k}$, a fixed-point method, such as Jacobi iteration technique, is then applied to update the HD-Slotboom variable at each mesh-point using the following equation:

$$H_{i,j,k}^{(n+1)} = \frac{A_{i,j,k}^{(n)} + \gamma_{i,j,k}}{B_{i,j,k} + L_{i,j,k}} \quad (9)$$

where $A_{i,j,k}^{(n)}$ represents the sum of the off-diagonal 'cH' terms in eqn.(7), and $B_{i,j,k}$ represents the sum of the 'c' coefficients of the diagonal term $H_{i,j,k}$.

This iteration scheme minimizes memory allocation, requiring only a few vectors of length N. Furthermore, due to the property of diagonal dominance, the convergence for the solution of each HD equation is guaranteed [4,5], while the convergence of the overall system is obtained with a modified Gummel method. Finally, by observing the decoupled algorthm of eqn.(9), it is clear that the method is readily parallelized.

III. Numerical Results

To examine the new method, we simulated a semi-recessed, 3-D submicron MOSFET, as shown in Fig. 1. The MOSFET has $0.5\mu m$ of channel length and $0.5\mu m$ of channel width. The electron temperature distribution at $V_{ds} = 1.0V$ and $V_{gs} = 3.5V$ is shown in Fig.2. Narrow channel effects can be observed.

[1] K-C. Wu, G. R. Chin and R. W. Dutton, *IEEE Trans. CAD*,vol.10,no.9,1132, 1991
[2] J-H. Chern, J. T. Maeda, L. A. Arledge, Jr., and P. Yang, *IEEE Trans. CAD*,vol.8,no.5,516, 1989
[3] Q.Lin,N.Goldsman and G.C.Tai,*Solid-State Electron*, vol.36,no.3,411, 1993.
[4] Q.Lin,N.Goldsman and G.C.Tai,*Solid-State Electron*, vol.37,no.2,359, 1994.
[5] I.D.Mayergoyz, *Journal of Appl. Phys.*,59(1),195,1986.
[6] C. Korman and I.D. Mayergoyz, *Journal of Appl. Phys.*, 68 (3), pp.1324, 1990.
[7] M. Lundstrom,*Fundamentals of Carrier Transport*,Modular Series in Solid State Devices, 1990.

*This work was supported by the Semiconductor Research Corporation.

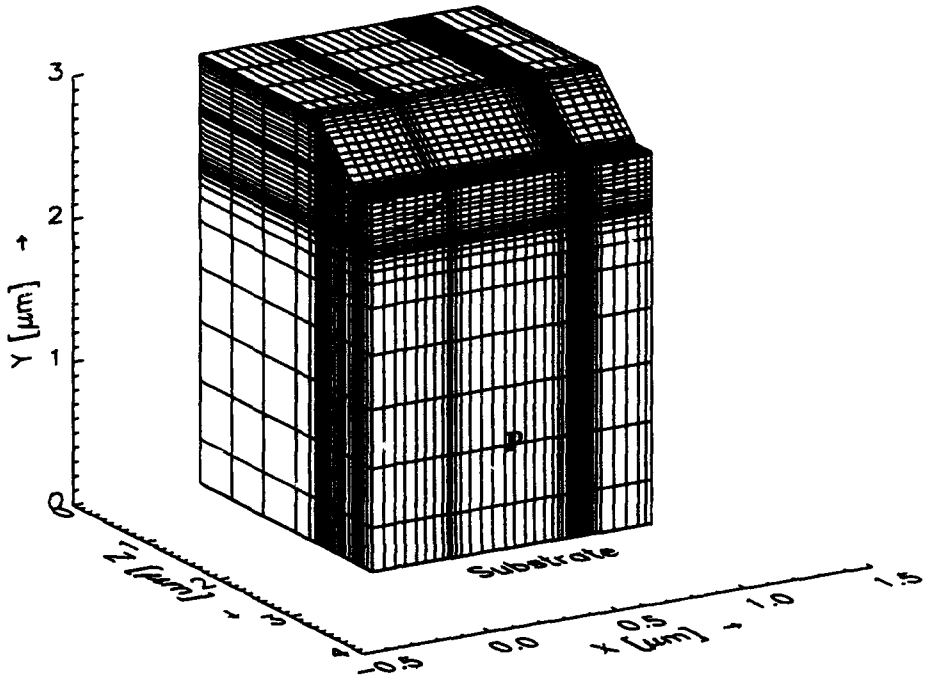


Fig.1. Semi-recessed MOSFET device with $0.5\mu m$ effective channel length and $0.5\mu m$ channel width.

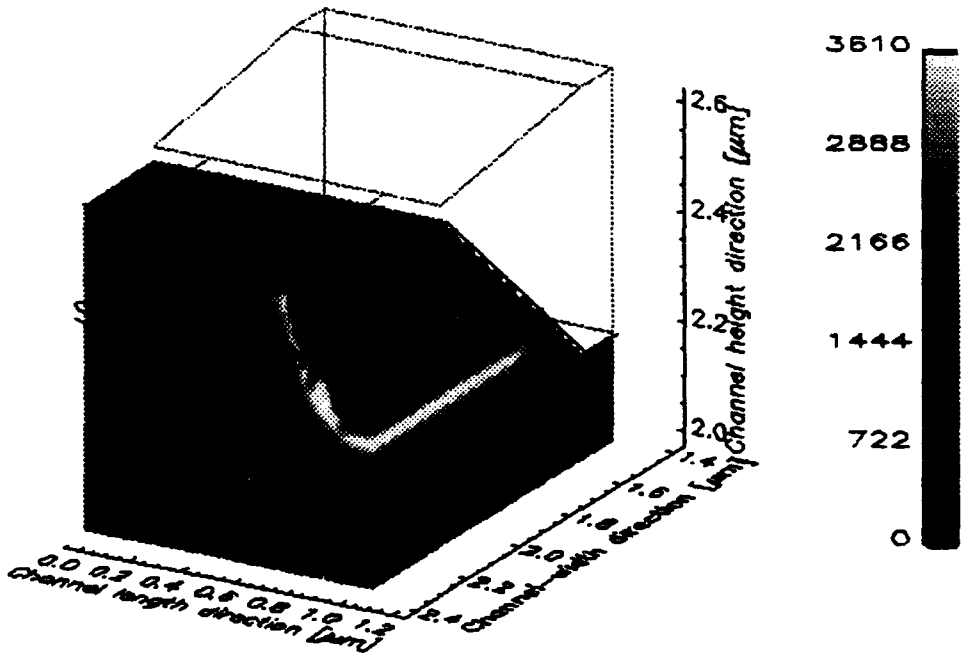


Fig.2. Electron temperature for the device with $V_{drain} = 1.0V$ and $V_{gate} = 3.5V$. (We have enlarged the channel region to show the 3-D effect).

THREE-DIMENSIONAL SIMULATION OF THE EFFECT OF RANDOM IMPURITY DISTRIBUTIONS ON CONDUCTANCE FOR DEEP SUBMICRON DEVICES

J-R. Zhou and D. K. Ferry
Department of Electrical Engineering and
Center for Solid State Electronics Research
Arizona State University, Tempe, AZ 85287-6206

Abstract

We present a 3-D simulation of small semiconductor devices, investigating the random impurity fluctuation and distribution effect on the conductance. Instead of using a uniform background charge for the impurities, discrete charges are assigned atomistically to computing cells by checking the assigned random numbers generated by computer, so that the scheme satisfies that the requirement that the mean value of the total discrete charge equals that of the uniform doping. The simulation is performed for MESFET structures. The random impurity distribution effect on devices with gate lengths less than 0.1 μm and narrow width will cause non-negligible conductance variations and be a major source of device variability within a single chip.

I. INTRODUCTION

In general, semiconductor device operation depends on the use of electrical potential barriers (such as gate depletion) in control of the carrier (electrons and holes) transport through the devices, in order to achieve signal switching and signal modulation. Although a successful device design is quite complicated and involves many aspects, the device engineering is mostly to devise a "best" device design by defining optimal device structures and manipulating impurity profiles to obtain optimal control of the carrier flow though the device. This becomes increasingly difficult as the device scale becomes smaller and smaller. New problems keep hindering the high performance requirement. Well-known problems include hot carrier effects, short-channel effects, etc. We discuss a potential problem caused by impurity fluctuation [1] which can not be perfectly controlled by the device engineer as devices become too small, and intend to provide better understanding of its effect on device design requirement for small devices.

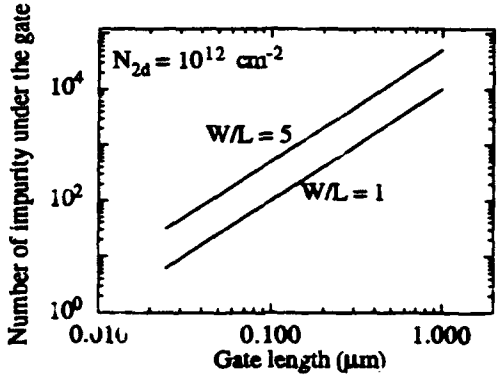


Fig. 1 Number of electrons under the gate versus gate length (L) for two gate width (W) to gate length ratio.

Impurities in semiconductor devices are randomly distributed as a result of the nature of processing, such as ion implantation. Although electron transport in the devices always experiences

*Work supported by the Army Research Office.

the effect of the random distribution of the impurities, the statistical contribution of these effects to the electronic performance of devices with a large operation domain (in space dimension, such as volume in which electrons flow) is negligible, and a simplified uniform background impurity distribution (the average of the impurity charges in space) is adequate in describing the effect of the fixed charges in the devices. Only devices with a small active domain which is susceptible to a large percentage fluctuation of the impurities, will exhibit noticeable conductance variations. For device scaling to the deep submicron regime, especially for device feature size less than 0.1 μm, as shown in Fig. 1, the number of impurity under the gate will approach the hundreds level and several tens of dopant number level. When the active device region contains so few dopant atoms that the statistical fluctuation of the dopant is comparable to the dopant number itself, the dopant fluctuation, either in total number and/or in spatial distribution, in the device will cause non-negligible effects on device performance. The anticipated effects include the classical statistical effects such as the device current level shift and threshold voltage shift due to the total dopant number fluctuation and/or distribution. Only a few attempts have been devoted to study the effect of random atomistic impurity on device performance, with the most recently reported research [3] using a drift-diffusion model to simulate the random impurity effect on sub-0.1 μm MOSFET devices. We investigate the classical effects from the simulation of a 3-dimensional device structures for MESFET by using hydrodynamic equations, with the discrete 3-dimensional random impurity distribution and fluctuation included.

II. DEVICE SIMULATION

The device structure considered is a 3-D volume as depicted in Fig. 2, which shows the contact definitions and a possible impurity distribution. The discrete impurity region is defined only in the highly doped layer away from the simulation domain boundary (smooth uniform doping is used at the contact ends) in order to use the existing simulation program and avoid dealing with very complicated rough-boundary conditions for the time being. This treatment should not affect the simulation results since most of the active device region (the channel) is well inside the discrete impurity distribution region and device operation is dominated by the electron transport through this region. The charge in a discrete cell is set to be either one or zero following a distribution scheme. The corresponding uniform doping in the highly doped layer is 1.5×10^{18} cm^{-3}. In sampling of 5000 devices, we plot the frequency as a function of dopant number under the gate in Fig. 3, which has a mean value at 36 (approximately one tenth of the mean value for the total discrete impurities in the device, since the volume of the gate region is one tenth of the volume of the total discrete region), with a standard deviation of 5.99.

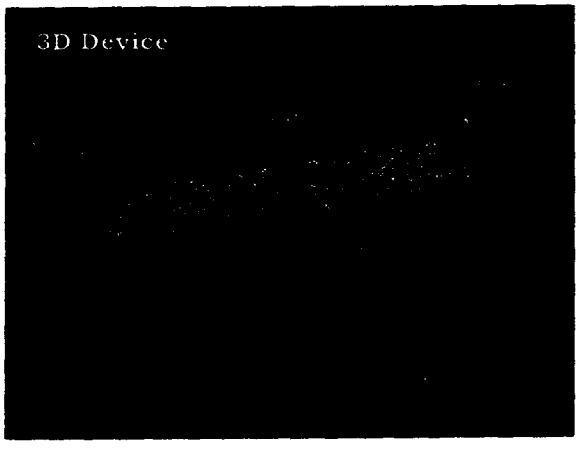


Fig. 2 A 3D device model and computer simulated discrete impurity distribution.

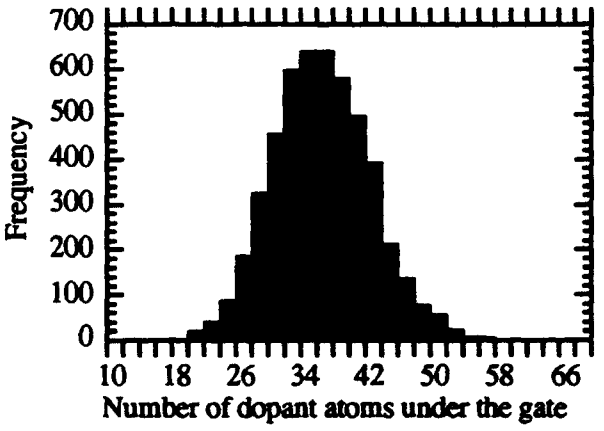


Fig. 3 Statistical frequency as a function of the number of impurity under the gate for 5000 devices.

The simulation domain is discretized into uniform small volume cells of 3 nm × 3 nm × 3 nm, and various physical quantities are computed for these small cells. The device simulation method is the same as that used in [4], which numerically solves a set of hydrodynamic equations that describe the conservation of particle, momentum and energy of electrons, in conjunction with Poisson's equation, using a finite-difference algorithm.

III. SIMULATION RESULTS

The simulated device structure (Fig. 2) is a domain of 0.36 μm (L) × 0.1 μm (H) × W, with W in the range of 0.042 μm to 0.162 μm. The thickness of the highly-doped layer is 40 nm, and results for two different gate lengths are discussed here. In Fig. 4, we plot the drain current versus gate voltage for three different MESFETs, where the drain potential is 0.5 V in all cases. One may notice that for the 24 nm gate devices, the current increase is not linearly proportional to the device width, as one would expect for the uniform doping case. Since there is no narrow width effect included in this simulation, and the number of impurity under the gate is 24 and 56 for W = 42 and W = 60 (more than doubled for the later case), respectively, the deviation from the expected behavior is apparently due to the impurity fluctuation and distribution in the devices. It is worth while to point out that the total impurity for the entire discrete region is not doubled: 358 for W = 42 and 610 for W = 60, respectively.

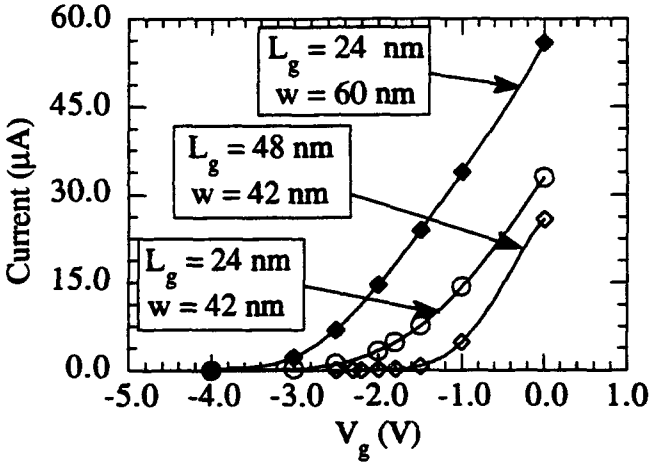


Fig. 4 I-V_g characteristics of MESFETs for three different gate cases, for V_d = 0.5 V.

The current fluctuation, caused by the variation in random impurity distribution and fluctuation, is pronounced for the device geometry simulated here. In Fig. 5 and Fig. 6, we plot the drain current as a function of gate voltage and drain voltage, respectively, for different number of impurities under the gate. Two characteristics are obvious. First, the total dopant fluctuation under the gate causes significant current variations. Secondly, it's not necessarily that more impurity under the gate will definitely cause more current flow, one can see the numbers are not ordered in both figures. This means that not only the number of impurities under the gate is important, but also the actual positions of the impurities. In Fig. 5, the current difference becomes wider as the gate bias becomes more negative. This is expected since the number of impurities in the opening channel becomes smaller as the depletion under the gate becomes wider, thus the fluctuation increases until the channel is completely depleted. For comparison, in Fig. 6, the results of simulation with uniform background doping is also included. Obviously, the current levels with discrete impurities fluctuates the current level of the uniform doping. The simulation shows that the variations can be as large as 30 per cent for the particular MESFET device structure.

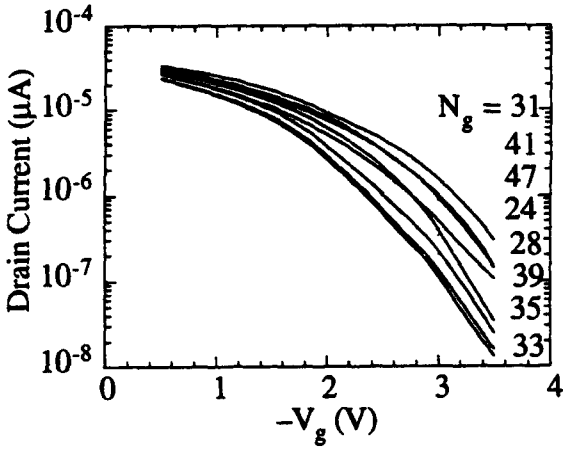


Fig. 5 I-V characteristics of gate voltage for a 24 nm gate MESFET, showing the current fluctuation as the impurity fluctuate under the gate.

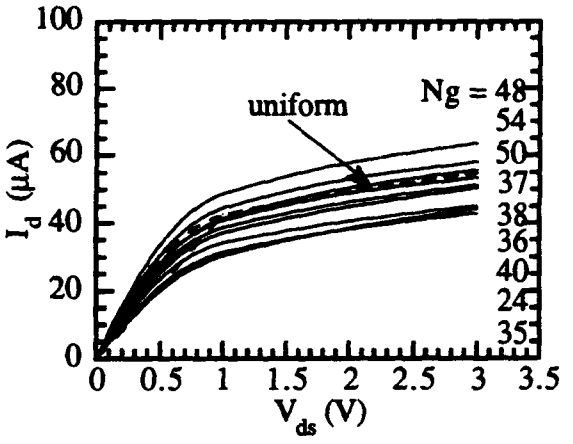


Fig. 6 I-V characteristics of drain voltage for a 24 nm gate MESFET, showing the current fluctuation as the impurity fluctuate under the gate.

We emphasize here that the current fluctuation mainly depends on the fluctuation and the distribution of the impurities under the gate, by inspecting the relation of the current fluctuation and the total dopant in the whole discrete region, we find that the current variation doesn't follow the total dopant fluctuation in the discrete dopant region clearly. This means that the impurity fluctuation in the whole device region is much less critical to the current flow through the channel, compared to the influence of the impurities under the gate. This is as expected, since the critical region that determines the current flow is the region under the gate, and also the impurity fluctuation and distribution under the gate are essentially independent of that of the whole device region.

IV. CONCLUSION

We investigated the effect of random impurity fluctuation and distribution on small device operations. For the device structure simulated here, the results suggest that the effects of random impurity fluctuation and distribution can cause large current variation for small devices if the total gate area is very small. As we expected, from our simulation, the random impurity effect on the device performance will be reduced with increase of the gate length or gate width. But the effect could be effective, even for a device with gate length close to 0.1 μm. A wider gate length device will be helpful in suppression of the effect of random impurity fluctuation. A full 3-D simulation of the device including the random impurity effect for engineering application may not be practical. We think that a combination of limited 3-D simulation plus statistical method might be needed in providing an applicable method to estimate the random impurity effect on device performance for device down scaling to 0.1 μm range.

References

[1] R. Keyes, "Physical limits in Semiconductor Electronics," in ***Science*** 195, p. 1230, 1972.

[2] H-S. Wong, and Y. Taur, " Three-Dimensional "Atomistic" Simulation of Discrete Random Dopant Distribution Effect in Sub-0.1μm MOSFET's," ***IEDM***, p. 705, 1993.

[3] J-R. Zhou, and D. K. Ferry, "Simulation of ultra-small GaAs MESFET using quantum moment equations," ***IEEE Trans. Electron Devices***, vol. 39, No. 3, pp. 473-478, 1992.

FAST AND ACCURATE HFET MODELLING FOR MICROWAVE CAD APPLICATIONS

Dr. R. Drury *(Mem IEEE)* and Professor C.M. Snowden *(Snr mem IEEE)*
Microwave and Terahertz Technology Group
Department of Electronic and Electrical Engineering
The University of Leeds, Leeds, UK

ABSTRACT

A new *quasi-two-dimensional* HFET model has been developed that solves the physical device equations in a more rigorous fashion than previously reported. The model incorporates a quantum mechanical description of the free electron concentration, self-consistently solving Schrödinger's and Poisson's equations, making it applicable to devices with scale lengths smaller than 20nm. The conventional one-dimensional charge control simulation is shown to be inadequate at the drain edge of the gate and is replaced by a quasi-two-dimensional version that more accurately describes the channel under drain-source bias. This modification produces much improved pinch-off characteristics which are essential for digital and low-noise characterisation.

I. INTRODUCTION

Quasi-two-dimensional (Q2D) FET models are based upon the fact that in the active region of the device the equipotential lines are essentially parallel, normal to the free surface, allowing the full two-dimensional device equations to be separated into their x and y components [1,2]. Conventional HFET models then proceed to solve these equations in terms of a charge-control law, taken perpendicular to the heterojunction(s), and a 'channel simulation' involving the carrier dynamics. The charge-control element typically consists of a self-consistent solution of Poisson's equation with the charge density (1) but with the first term, $\partial \mathcal{E}_x/\partial x$, set to zero.

$$\frac{\partial(\varepsilon_r \mathcal{E}_x)}{\partial x} + \frac{\partial(\varepsilon_r \mathcal{E}_y)}{\partial y} - \frac{q}{\varepsilon_0}.\left(N_D^+ - n\right) = 0 \qquad (1)$$

The channel simulation then solves the full Poisson equation together with the current-continuity and the energy and momentum conservation equations. Here the values for $\partial \mathcal{E}_y/\partial y$ and N_D^+ obtained from the charge-control law are used, and the equations solved in a 'current-driven' form. However, the omission of the $\partial \mathcal{E}_x/\partial x$ term in the charge-control equation leads to important errors. When the $\partial \mathcal{E}_x/\partial x$ term tends to $-\infty$ the electron density increases indefinitely to compensate, but when the opposite extreme is approached, $\partial \mathcal{E}_x/\partial x$ tending to $+\infty$, the electron density can only be reduced to zero. The equations described above do not operate in this fashion since as the $\partial \mathcal{E}_x/\partial x$ term in the channel simulation increases to $+\infty$ they predict the sheet electron density tends to $-\infty$, which is clearly wrong. This leads to poor simulation of the pinch-off characteristics where the most extreme fields are produced.

II. THE QUASI-TWO-DIMENSIONAL MODEL

The new scheme described in this paper modifies the charge-control law, making the $\partial \mathcal{E}_y/\partial y$ and N_D^+ terms functions of $\partial \mathcal{E}_x/\partial x$. This is performed by supplementing the one-dimensional Poisson equation used in the charge-control law with a constant term representing $\partial \mathcal{E}_x/\partial x$ which is applied to the channel and substrate penetration regions, Figure 1.

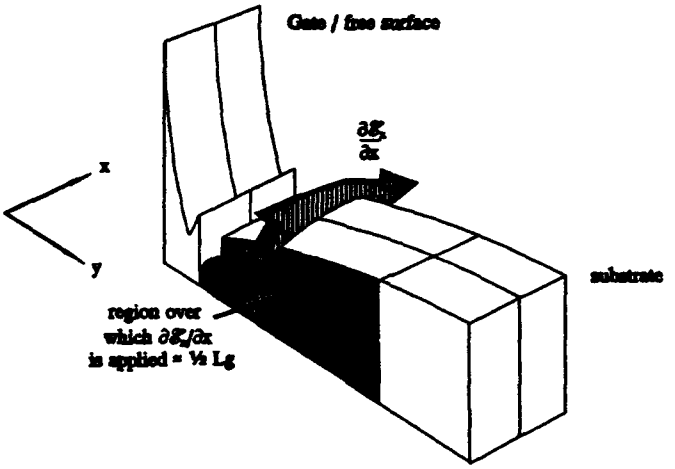


Figure 1 Slice of the conduction band edge illustrating the region over which the constant term, $\partial \mathcal{E}_x/\partial x$, is applied.

This produces typical conduction band-edge diagrams illustrated in Figure 2. Here the $\partial \mathcal{E}_x/\partial x$ term is varied illustrating the effect at extreme biases.

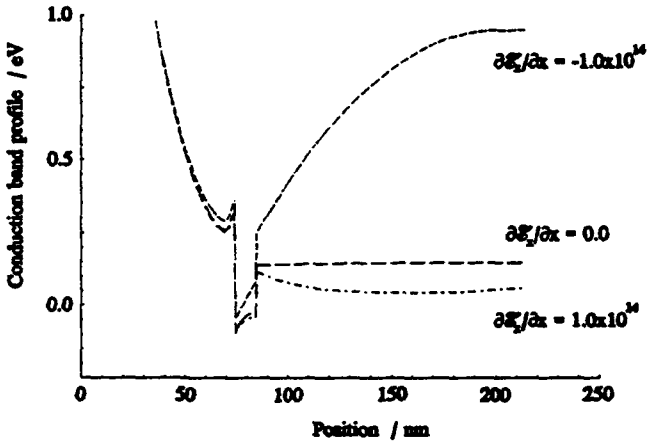


Figure 2 Conduction band profiles for various values of $\partial \mathcal{E}_x/\partial x$

To accommodate the small scale lengths associated with heterostructure devices quantum mechanics is introduced by including Schrödinger's equation in the charge-control law. The self-consistent solution of Poisson's and Schrödinger's equations is performed within the framework of a modified Newton-Raphson iterative scheme that rapidly, accurately and robustly solves the equations over

all appropriate biases. This method requires the partial derivatives of λ and Ψ with respect to electrostatic potential [3]. Applying perturbation theory to Schrödinger's equation, these parameters turn out to be -1 and 0 respectively (2).

$$\frac{\partial \lambda}{\partial V} = -1 \quad , \quad \frac{\partial \Psi}{\partial V} = 0 \tag{2}$$

The equations are then solved for a range of biases and the results stored in a look-up table. This produces the two-dimensional surface illustrated in Figure 3, where the two-dimensional nature of this term is apparent. Here the 'effective sheet electron density' listed on the z axis represents the integral of $\partial \mathcal{E}_y/\partial y + N_D^+$ over the whole of a vertical slice. The two-dimensional look-up table is then used in the channel simulator which extracts the value of this combined term using a two-dimensional cubic spline routine.

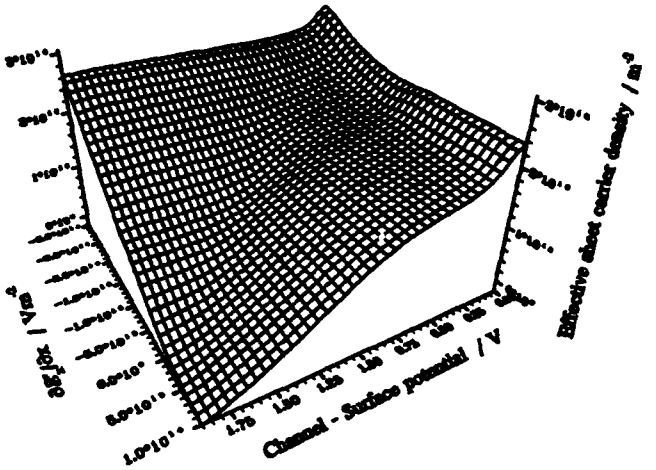


Figure 3 Two-dimensional variation of the effective sheet electron density ($\int \partial \mathcal{E}_y/\partial y + N_D^+ dy$)

Figure 4 illustrates the effect the two-dimensional charge control law has on the pinch-off characteristics of a simulated device. Here the poor pinch-off of the conventional scheme is illustrated and compared with the improved results of the new model. It is worth noting that accurate simulation of pinch-off is essential for digital device modelling and also that most low-noise devices are operated in this region.

III. Microwave simulation

The S-parameter calculation requires the microwave gate current to be calculated together with the access capacitances. The first term is proportional to the capacitance associated with the charge-control law and turns out to be relatively independent of the $\partial \mathcal{E}_x/\partial x$ term. Hence a one-dimensional look-up table is sufficient. This method provides an accurate estimate of the gate current evident in the good agreement between measured and simulated S-parameters illustrated in Figure 5.

This project is funded by M/A-COM, Corporate R&D, Boston, USA

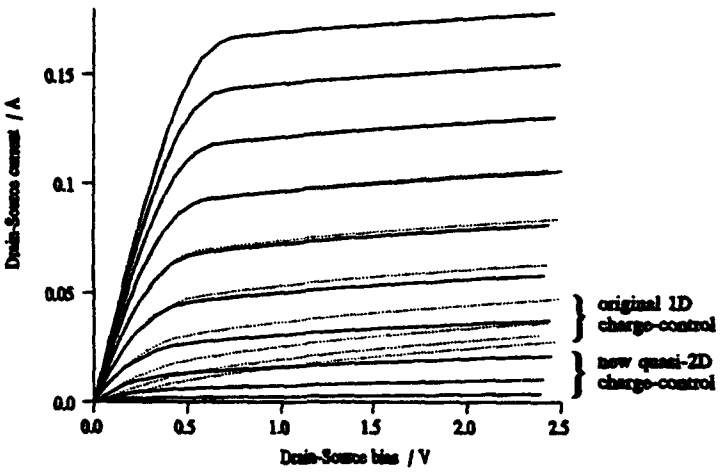


Figure 4 I_{DS}-V_{DS} curves for old and new charge-control schemes

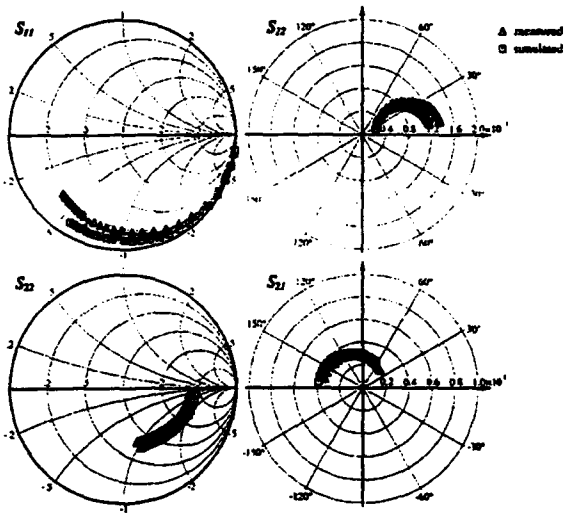


Figure 5 Comparison of measured and simulated S-parameters for a British Telecom AlGaAs/GaAs HEMT

IV. REFERENCES

[1] C.M. Snowden and R.R. Pantoja, "Quasi-Two-Dimensional MESFET Simulations for CAD", *IEEE Trans Electron Devices*, Vol. ED-36, No. 9, pp. 1564-1574, September 1989

[2] R.R Pantoja, M.J. Howes, J.R. Richardson and C.M. Snowden, "A Large-Signal Physical MESFET Model for Computer-Aided Design and Its Applications", *IEEE Trans Microwave Theory and Techniques*, Vol MTT-37, No. 12, pp. 2039-2045, December 1989

[3] R. Drury, C.M. Snowden and R.E. Miles, "A New Fast Full 2D Quantum Model for Heterojunction Field Effect Transistors", *International Workshop on Computational Electronics*, Leeds, England, August 1993

DUAL ENERGY TRANSPORT MODEL FOR ADVANCED DEVICE SIMULATION

Zhiping Yu, Lydia So*, Edwin C. Kan, and Robert W. Dutton
Center for Integrated Systems, Stanford University, CA 94305

Abstract

Dual energy transport (DUET) model in semiconductor devices including heterostructures has been developed to simulate the distribution of carrier and lattice temperatures in addition to profiles of the electrostatic potential and carrier concentrations. The modeling approach is in consistency with the conventional drift-diffusion (DD) model, making it easy to implement in the existing code. Carrier energy dependent mobility and impact ionization models have been examined and are used for simulation of various velocity overshoot and hot electron effects. Two simulation examples, one for the submicron MOSFET and another for the deep-submicron SOI, are presented through comparison with measurement data to demonstrate the improvement of the new model over DD model in predicting the device characteristics for modern (submicron) structures.

I INTRODUCTION

As the feature size of semiconductor devices shrinks to the quarter-micron regime, nonlocal effects such as hot electrons and velocity overshoot become important in determining the device characteristics. The conventional drift-diffusion transport model has been and continuously been used in industry and academia for design and analysis of IC devices largely because its auxiliary physical models such as the field-dependent mobility model and impact ionization model are well calibrated. But it fails to predict those device characteristics which becomes critical in sub- and deep sub-micron devices. A notable example is the substrate current in MOSFET. Neither can DD model provide such vital information as to the average kinetic energy of carriers in the device. On the other hand, Monte Carlo (MC) method can provide very detailed information about the carrier distribution in real and momentum spaces. But in addition to the excessive CPU time requirement and complexity of model parameters, most present MC codes can only simulate one-carrier device behavior, thus are not yet suitable for the design of practical devices. Through tracing back to the origin of DD model from Boltzmann Transport Equation (BTE) and by relieving the constraints of constant effective mass and temperature, we were able to develop a more complete transport model in semiconductors, which reveals not only the carrier concentration and current density (essentially a measure of carrier average velocity) but also the carrier energy density. Assuming Fermi-Dirac (FD) statistics as the basis for the distribution function and applying the perturbation theory, the average kinetic energy can be linked to the temperature parameter used in FD statistics in the same formulation as for an ideal gas in classical thermodynamics, thus correctly identifying the

*also with Los Alamos National Lab.

concept of carrier temperature. Moreover, the lattice temperature can easily be incorporated in the model by considering the energy exchange among the carrier subsystems and lattice. We thus proposed a dual (carrier and lattice) energy transport model (DUET) for semiconductor devices, and have successfully implemented the model in Stanford's device simulation code, PISCES, as version 2ET. Since the code was up to work we have conducted several tests and results obtained from the simulation are encouraging. Recently, we also extend the capabilities of the code to cover the electrical simulation of heterostructures such as light emitter diodes (LEDs) and surface emitting diodes (SELs).

In the following we first introduce the complete formulation of the DUET model, and describe the energy dependent physical models as implemented in PISCES-2ET code. Finally, examples are given and future work is discussed.

II DUET MODEL FORMULATION

DUET model uses six state variables – potential (ψ), electron and hole concentrations and temperatures (n, p, T_n, and T_p), and lattice temperature (T_L) to describe the status of a semiconductor device. All governing equations for these variables are derived from the conservation or balance principle for matter and energy. Except of the Poisson's equation for ψ, this conservation principle can be expressed in the following equation using Fick's second law:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{F} + g - r \tag{1}$$

where u represents the concerned physical quantity, $\mathbf{F}$ is the flux of u, and g and r are generation and recombination rates of u, respectively. The key issue in the modeling is thus to find expressions for $\mathbf{F}$, g, and r in terms of u. It is well known that in DD model, the carrier flux consists of the diffusion and drift terms, or by introducing the quasi-Fermi level E_F, is proportional to the gradient of E_F. The fundamentals behind the DD model lie actually in BTE, which itself is a representation of the continuity principle,

$$\frac{d}{dt} f(\mathbf{r}, \mathbf{k}, t) = \left.\frac{\partial f}{\partial t}\right|_{coll} \tag{2}$$

where $\mathbf{k}$ is the wavenumber vector, and the way of constructing the distribution function of f. Following Stratton's approach [1], in the existence of the electric field by applying the perturbation theory and relaxation time approximation (RTA) the distribution function at any instant can be expressed as

$$f(\mathbf{r}, \mathbf{k}) = f_0(\mathbf{r}, E) - \tau(\mathbf{r}, \epsilon)\frac{\hbar}{m^*}\left(\mathbf{k} \cdot \nabla_{\mathbf{r}} f_0 - q\frac{\partial f_0}{\partial \epsilon}\mathcal{E} \cdot \mathbf{k}\right) \tag{3}$$

where E is the carrier energy and ϵ is the kinetic part of E, f_0 is the even part of f in $\mathbf{k}$–space and is dependent on the carrier kinetic energy only. The relaxation time τ is assumed to depend on ϵ only too. $\mathcal{E}$ is the electric field. All other symbols have conventional meanings. If f_0 is taken as the Fermi-Dirac distribution function, one can readily obtain the expression for the carrier density and flux by definition as follows:

$$n = N_C F_{1/2}\left(\frac{E_{Fn} - E_C}{k_B T_n}\right) \tag{4}$$

$$\mathbf{j}_n = n\mu_n \nabla E_{Fn} + qn\mu_n Q_n \nabla T_n \tag{5}$$

where N_C is the effective density for the conduction band and E_C is the energy level for the conduction band edge, $F_{1/2}$ is the Fermi integral of order one half. Coefficients μ_n and Q_n in

Eq. (5) are electron mobility and thermopower, respectively, and are related to each other. It is obvious that when T_n is constant the above expression for the current is reduced to the conventional DD model.

To find the governing equation for carrier temperature, we start from the balance equation for the kinetic energy. For electrons, we have

$$\frac{\partial w_n}{\partial t} = -\nabla \cdot \mathbf{s}_n + \mathbf{j}_n \cdot \mathcal{E}_n - u_{wn} \tag{6}$$

where w_n is the electron kinetic energy density, $\mathbf{s}_n$ is the energy flux, and u_{wn} is the net energy loss rate. The Joule heat term of $\mathbf{j}_n \cdot \mathcal{E}_n$ represents the conversion from the potential to kinetic energy and the subscript n in $\mathcal{E}_n$ indicates the fact that for heterostructures, the electric field might be different for electrons and holes. Again using Eq. (3) and by definition w_n and $\mathbf{s}_n$ are computed as follows:

$$w_n = \frac{3}{2} n k_B T_n \gamma_n \tag{7}$$

$$\mathbf{s}_n = -P_n T_n \mathbf{j}_n - \kappa_n \nabla T_n \tag{8}$$

where γ_n is the degeneracy factor which equals unity when Boltzmann statistics is used, P_n and κ_n are thermoelectric power and thermal conductivity for electrons, respectively. From Eq. (7) and taking $\gamma_n = 1$, we can identify that T_n indeed has the meaning of temperature for a classical electron gas.

The remaining task in completing the model formulation is to find the carrier and its energy exchange among sub-systems. For carrier exchange, i.e., recombination and generation, we include the Shockley-Read-Hall (SRH), Auger, and radiative recombinations, and impact ionization. All these carrier exchanges are accompanied by the energy exchange. In addition, we also include the energy exchange between carriers and lattice through phonon scattering modeled using energy relaxation times, τ_{wn} and τ_{wp}.

We thus arrive at the following set of equations:

Poisson's equation

$$\nabla \cdot (-\epsilon \nabla \psi) = q(p - n + N_D^+ - N_A^-) \tag{9}$$

Carrier continuity equations:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \mathbf{j}_n - u \tag{10}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \cdot \mathbf{j}_p - u \tag{11}$$

Energy balance equations for carriers:

$$\frac{\partial w_n}{\partial t} = -\nabla \cdot \mathbf{s}_n + \mathbf{j}_n \cdot \mathcal{E}_n - u_{wn}$$

$$\frac{\partial w_p}{\partial t} = -\nabla \cdot \mathbf{s}_p + \mathbf{j}_p \cdot \mathcal{E}_p - u_{wp}$$

where

$$u_{wn} = (u_{srh} + u_{rad}) \frac{3}{2} k_B T_n - (u_{n,Auger} - g_{n,imp}) \left[E_g(T_L) + \frac{3}{2} k_B T_p \right] - g_{p,imp} \frac{3}{2} k_B T_n + \frac{w_n(T_n) - w_n(T_L)}{\tau_{wn}} \tag{12}$$

$$u_{wp} = (u_{srh} + u_{rad})\frac{3}{2}k_B T_p - (u_{p,Auger} - g_{p,imp})\left[E_g(T_L) + \frac{3}{2}k_B T_n\right] - g_{n,imp}\frac{3}{2}k_B T_p + \frac{w_p(T_p) - w_p(T_L)}{\tau_{wp}} \quad (13)$$

Thermal diffusion equation for lattice:

$$c_L\frac{\partial T_L}{\partial t} = \nabla\cdot(\kappa_L\nabla T_L) + u_{srh}\left[\frac{3}{2}k_B T_n + E_g(T_L) + \frac{3}{2}k_B T_p\right] + \frac{w_n(T_n) - w_n(T_L)}{\tau_{wn}} + \frac{w_p(T_p) - w_p(T_L)}{\tau_{wp}} \quad (14)$$

In the transport expressions for the current density and energy flux, there are four coefficients, μ, Q, P, and κ, and they are all related to each other through the relaxation time, τ. This is one of the advantages of the DUET model. Once one of them is known, the others can be deduced from the known parameter. In reality, however, these coefficients can be treated as empirical parameters or obtained from experiment. For example, knowing the dependence of μ on the carrier temperature, thermopower Q can be obtained for electrons:

$$Q_n = \frac{k_B}{qn}\left[1 + T_n\frac{\partial}{\partial T_n}\ln\mu_n(\mathbf{r}, T_n)\right] \quad (15)$$

Finally, we list the energy dependent mobility and impact ionization models as used in PISCES-2ET:

$$\mu(N, T_L, E_\perp, T_c) = \frac{\mu_0(N, T_L, E_\perp)}{1 + \gamma(N, T_L, E_\perp)[w(T_c) - w(T_L)]} \quad (16)$$

where the subscript c for n or p, and $\gamma(N, T_L, E_\perp) = \mu_0(N, T_L, E_\perp)/[q\tau_w v_{sat}^2(T_L)]$ [2]. And the impact ionization rate, α,

$$\alpha = A\exp[-(b/\mathcal{E}_{eff})^m] \quad \text{with} \quad \mathcal{E}_{eff} = \frac{3}{2}\frac{k_B}{q}\frac{T_c - T_L}{\tau_w v_{sat}} \quad (17)$$

III SIMULATION EXAMPLES

We present two examples which show that the DD model is no longer accurate in predicting $I-V$ characteristics for submicron devices when the non-stationary phenomena such as the velocity overshoot and nonlocal field dependence of physical parameters such as the impact ionization rate become important. While both DD and DUET models provide good simulation results compared to the measurement for devices with relatively long channel length, DD model starts to break for output characteristics of SOI at $L_{eff} = 0.12\,\mu$m (Fig. 1) and substrate current of MOSFET at $L_{eff} = 0.8\,\mu$m (Fig. 2). On the other hand DUET can consistently model the device characteristics well even when the device size is scaled down to the deep submicron range.

IV CONCLUSION

DUET model follows the same moment approach as DD model does but has two obvious advantages. Firstly, it closes the system in a more consistent way and resolves the conflict intrinsic to DD model, i.e., the field dependent mobility model as commonly used in device simulators employing DD model vs. the constant temperature assumption leading to the DD current expression. Secondly, it is able

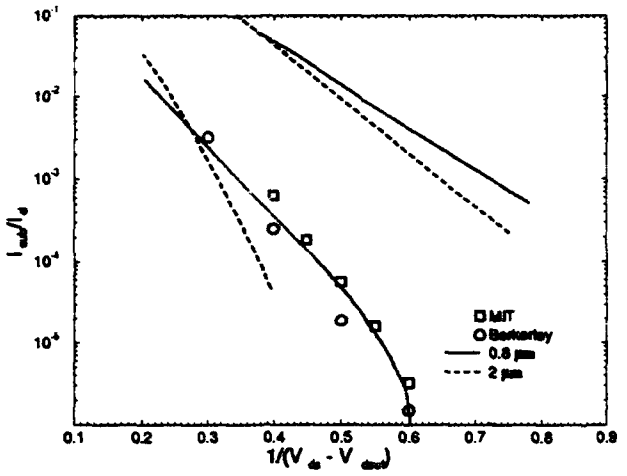


Figure 1: Simulation results for the substrate current in MOSFET with two different channel length (2 and 0.8 μm) and the comparison is made for 0.8 μm case between the ET-simulated and measured results (from MIT and UC Berkeley, respectively). The upper curves are simulated using DD model while the lower curves are obtained from ET simulation.

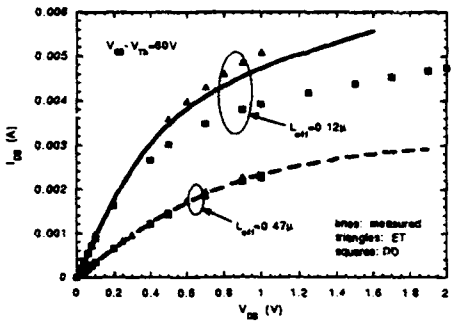


Figure 2: Simulated and measured data for SOI structure with different channel length.

to provide information regarding the carrier kinetic energy. However, there is still a need to carefully calibrate the transport coefficients and to develop more reliable energy-dependent physical models. Especially for the impact ionization and breakdown simulation, since these phenomena are basically determined by the high energy tail of the carrier distribution, the dependency on the average energy has to be elaborated and verified through experiments.

V Acknowledgement

This work has continuously been supported by SRC through contract SRC 93-SJ-116, and the authors are grateful to the encouragement and input from industrial sponsors associated with SRC.

References

[1] R.Stratton: IEEE Trans. Elect. Dev., **ED-19**, 1288 (1972)

[2] W.Haensch, M.Miura-Mattausch: J. Appl. Phys. **60**, 650 (1986)

GALERKIN METHODS FOR THE BOLTZMANN EQUATION USING VARIABLE COORDINATE SYSTEMS

Christian Ringhofer *)
Department of Mathematics, Arizona State University, Tempe, AZ85287, USA

1. INTRODUCTION

In this paper we present a solution method for the Boltzmann transport equation based on a series expansion in a variable coordinate system. In an appropriately scaled form, the Boltzmann transport equation is given by

$$\epsilon^2 \partial_t f + \epsilon div_x(v(k)f) + \epsilon div_k(Ff) = Qf, \tag{1}$$

Equation (1) is already given in a scaled and dimensionless form where the dimensionless parameter ϵ stands for the scaled mean free path. The term Qf on the right hand side of equation (1) denotes the collision operator. For the purpose of this paper we restrict ourselves to the collision operator resulting from the relaxation time approximation. This means, that the operator Q in (1) is given by

$$Qf(x,k,t) = a(\rho,u,t)M(x,k) - b(\rho,u,T)f, \tag{2}$$

where $M(x,k)$ is a given equilibrium distribution and a and b are functions of the density ρ , the group velocity u, and the scaled temperature T , given by

$$\rho(x,t) = \int_{R_k^3} f(x,k,t)dk, \quad \rho u(x,t) = \int_{R_k^3} v(k)f(x,k,t)dk, \quad \rho(3T+|u|^2) = \int_{R_k^3} |v|^2 f(x,k,t)dk. \tag{3}$$

The hydrodynamic model equations are usually obtained by by integrating equation (1) against the functions $1, v(k)$ and $|v(k)|^2$, assuming either a parabolic band structure or making an effective mass approximation, and assuming that the density function f has the shape of a drifted Maxwellian in the wave vector direction.

$$f(x,k,t) = \frac{\rho(x,t)}{\pi(2T)^{\frac{3}{2}}} exp[\frac{-|v-u(x,t)|^2}{2T}]. \tag{4}$$

For elastic scattering terms, which preserve momentum and energy, the hydrodynamic model can be justified via asymptotic analysis for small values of the scaled mean free path, the quotient of the mean free path and the length scale under consideration [BAR]. For inelastic collision processes, which are usually present for electron transport in solids, a similar asymptotic analysis exists, leading to slightly different model equations [POU]. However, the hydrodynamic model is frequently used in regimes where the scaled mean free path is actually quite large. An alternative approach is to expand the solution of the Boltzmann transport equation in a set of basis functions in velocity space (c.f. spherical harmonics) with coefficients which are dependent on position and time. After using a Galerkin approximation a set of macroscopic equations is obtained for these coefficients [VEN], [LIN]. However, the resulting expansion converges quite slowly in the presence of high velocities and strong forces. In this paper we present a series expansion method, also based on a Galerkin procedure which is better suited for high field regimes and a relatively large mean free path. The basic idea is to use a coordinate transformation in phase space which adapts locally to the shape of the distribution function f. This transformation is chosen such that the resulting expansion reduces to the hydrodynamic model if only a few terms are used.

2. GALERKIN APPROXIMATION

The method presented in this paper is a Galerkin approximation of the Boltzmann transport equation (1) in a weighted L^2 space, using a variable transformation in the velocity variable which is dependent on position and time. We employ the coordinate transformation $k \rightarrow w$, given in general by

$$w = G(x,k,t) \tag{5}$$

and expand the solution of the Boltzmann transport equation as

$$f(x,k,t) \approx \sum_{n \epsilon N} \hat{f}(x,n,t)\psi_n(x,k,t), \tag{6}$$

where the basis functions ψ_n are given by some suitable set of basis functions ϕ_n under the variable transformation G, which will be chosen such as to adapt locally to the shape of the distribution function f.

$$\psi_n(x,k,t) = \phi_n(w) \quad , \quad w = G(x,k,t) \tag{7}$$

holds. N in (6) denotes some suitable, finite index set. We introduce the weighted L^2 inner product for two scalar functions $f(x,k,t)$ and $g(x,k,t)$ and two vector valued functions $\mathbf{f}(x,k,t)$ and $\mathbf{g}(x,k,t)$ by

$$< f,g > (x,t) = \int_{R_k^3} \mu(G(x,k,t)) f(x,k,t) g(x,k,t) dk, \quad < \mathbf{f},\mathbf{g} > (x,t) = \int_{R_k^3} \mu \mathbf{f}^T \mathbf{g} dk. \tag{8}$$

The variable transformation G is assumed to be affine and the basis functions ϕ_n are assumed to be normalized. So

$$\int_{R_w^3} \mu(w)\phi_n(w)\phi_m(w)dw = \delta_{m,n} \tag{9}$$

holds. Taking the inner product $< .,. >$ of the Boltzmann transport equation with the test function $\psi_m, m \epsilon N$ yields the macroscopic equations for the coefficients $\hat{f}(x,n,t)$. Inserting the expansion (6) into the Boltzmann transport equation (1), taking the scalar product with the basis function $'\psi_m$, gives

$$\partial_t \hat{f}_m + \sum_{j=1}^{3} \sum_{n \epsilon N} \partial_{x_j} [A_{j,m,n} \hat{f}_n] - \sum_{n \epsilon N} B_{m,n} \hat{f}_n = C_m, \tag{10}$$

where the tensors A, B and C are given by

$$(a) \quad A_{j,m,n}(x,t) =< \psi_m, v_j \psi_n >, \quad (b) \quad C_m(x,t) = a < \psi_m, M > -b\hat{f}_m \tag{11}$$

$$B_{m,n}(x,t) =< \frac{1}{\mu}\psi_n, \partial_t(\mu\psi_m) + v \bullet \nabla_x(\mu\psi_m) + F \bullet \nabla_v(\mu\psi_m) > .$$

To this point the equations (10) represent a Galerkin approximation of the Boltzmann transport equation in a variable coordinate system, which will be convergent, provided the equations (10) are stable and the derivatives of the density function f stay bounded in the weighted L^2 space with the weight μ. We make the following choices for the basis functions ψ_m and Φ_m and the weight function μ, which lead to a generalization of the hydrodynamic model equations. We set

$$(a) \quad \mu(w) = exp(|w|^2), \quad (b) \quad \Phi_m(w) = exp(-|w|^2)P_m(w), \quad c) \quad \psi_m(x,k,t) = \Phi_m(G(x,k,t), \quad m \epsilon N \tag{12}$$

where the P_m are chosen as suitably orthogonalized polynomial functions. Furthermore, we choose the affine transformation G such that

$$|G(x,k,t)| = \frac{|v - \mathbf{p}(x,t)|}{\alpha(x,t)} \tag{13}$$

holds for some vector $\mathbf{p}$ and some scalar α. The basic idea behind these choices is that, because of (12), taking the scalar product of a given function with the basis function ψ_m corresponds to computing a certain linear combination of the moments of that function, and, because of (13), the zero order basis functions will have the shape of the drifted Maxwellian (4). This results in the hydrodynamic model being a special case of the above approximation procedure.

To this point we have not specified the choice of the functions $\alpha(x,t)$ and $p(x,t)$, which make up the variable transformation in velocity space. Indeed, any choice of $\alpha(x,t)$ and $p(x,t)$ would yield a convergent

Galerkin method for the Boltzmann transport equation, provided the density function $f(x,k,t)$ is sufficiently smooth. We now choose p and α dependent on the solution itself in order to minimize the number of terms needed in the expansion procedure. We choose $p(x,t)$ as the mean velocity u given by the solution of the Galerkin approximation. So we set

$$p(x,t) = \frac{1}{\rho(x,t)} \int_{R_k^3} v f(x,k,t) dk \quad . \tag{14}$$

The function α is chosen such that it matches the macroscopic temperature T in (3). So we set

$$\alpha(x,t) = \sqrt{\frac{2}{3}} \sqrt{\frac{1}{\rho(x,t)} \int_{R_k^3} |v|^2 f(x,k,t) dk - |p|^2} \quad with \quad \rho(x,t) = \int_{R_k^3} f(x,k,t) dk. \tag{15}$$

Of course, this choice of α and p transforms the linear problem (10) into a nonlinear one.

The hyperbolic system (10) can be discretized by any method suitable for systems of hyperbolic conservation laws. For the purpose of this paper the method of choice is the Lax - Wendroff scheme.

3. A NUMERICAL TEST EXAMPLE

To demonstrate the effect of the corrections to the hydrodynamic model resulting from the above approximation procedure, we present a one dimensional example of the reflection of an electron wave at a potential barrier. Assuming a parabolic band structure, the density function $f(x,k,t)$ remains cylindrically symmetric around the k_1- direction in the one dimensional case. So

$$f(x,k,t) = f(x_1, k_1, k_2^2 + k_3^2, t) \tag{16}$$

holds. The affine variable transformation G in (5) is given by

$$G(x,k,t) = \frac{k - p(x,t)\mathbf{e}}{\alpha(x,t)}, \quad \mathbf{e} = (1,0,0)^T, \tag{17}$$

and the polynomial basis functions P_m in (12) are given by the standard Laguerre polynomials. Calculations were performed for the scaled mean free path ε in (1) equal to 0.2, which corresponds to the regime where the hydrodynamic model is usually applied. Figure 1 shows the force F corresponding to the potential barrier. Figure 2 shows the currents at a given point in time resulting from three different calculations. The solid line has been obtained by using 10 terms in the expansion, 5 in the k_1- direction times 2 in the orthogonal $k_2^2 + k_3^2-$ direction. The dotted line gives the current resulting from the corresponding hydrodynamic model, using only three terms in the expansion. Finally, the the dashed line represents the 'numerically exact' solution, obtained by using 32 times 8 modes. As observed in the past, the hydrodynamic solution exhibits artificial velocity overshoot phenomena. These phenomena are not present when 10 terms are used, and the current is essentially compute correctly. Figure 3 shows the L^2 norm of various modes as a function of time. It can be seen that the higher order modes produce a significant correction. In particular, a simple algebraic calculation shows that the (1,0) term is responsible for non - scalar temperatures. Finally, Figure 4 shows a snapshot of the distribution function f at the midpoint $x = 0.5$ and a certain point in time as a function of k_1 and $k_2^2 + k_3^2$. If the assumptions underlying the hydrodynamic model were correct, the distribution function would have to be a Maxwellian, shifted along the k_1- direction. The additional peak in the exact distribution function is responsible for the artificial velocity overshoot in the hydrodynamic model.

*) supported by ARPA Grant No. F49620-93-1-0062

REFERENCES:

[BAR] C.Bardos, F. Golse, D. Levermore, in 'Multi - Dimensional Hyperbolic Problems and Computations' J.Glimm, A. Majda eds. pp. 1-13, Springer (1991).
[LIN] H. Lin, N. Goldsman, I. Mayergorz: Solid State Electronics, vol. 35, p33 (1992).
[POU] F. Poupaud: Appl. Math. Letters, vol.9, pp65-69 (1993).
[VEN] D. Ventura, A.Gnudi, G. Baccarani: Appl. Math. Letters, vol.5, pp45-57 (1992).

FIGURE 1
Electric Field

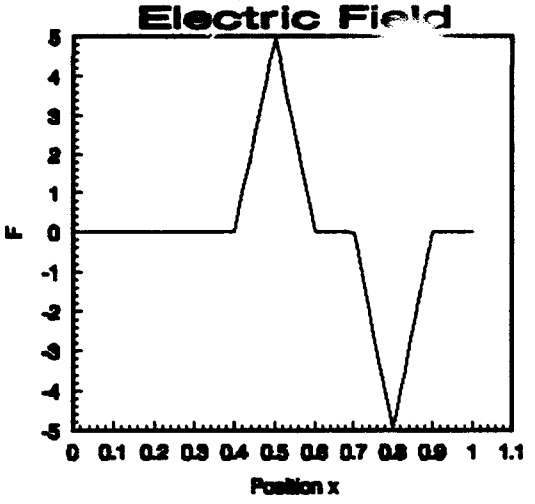


FIGURE 2
Currents at t=0.02

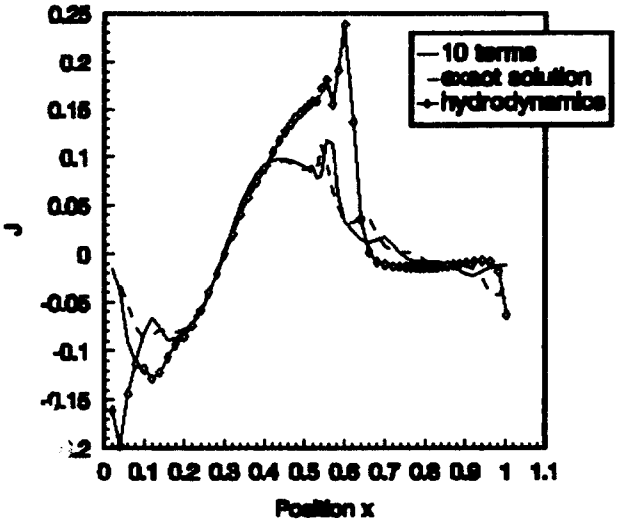


FIGURE 3
Expansion Coefficients

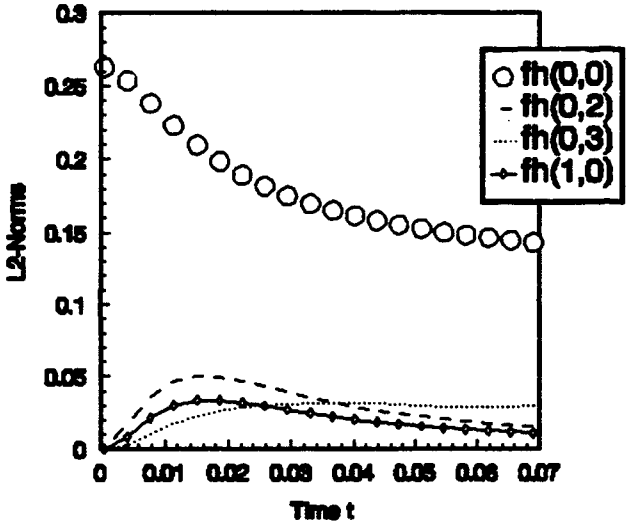
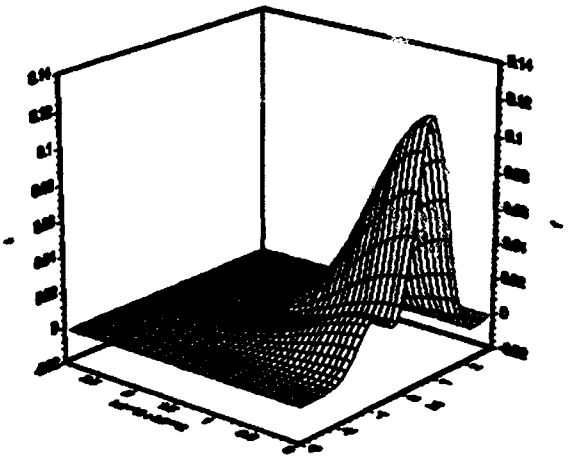


FIGURE 4
Distribution f for x=0.5 and t=0.02



APPLICATION OF HIERARCHICAL TRANSPORT MODELS FOR THE STUDY OF DEEP SUBMICRON SILICON N-MOSFETS

S. Jallepalli, C.-F. Yeap, S. Krishnamurthy, X. L. Wang†,
C. M. Maziar and A. F. Tasch Jr.

Microelectronics Research Center
The University of Texas at Austin, Austin, TX 78712-1100
†now with Intel Corporation, Santa Clara, California

Abstract

In this paper, we present an integrated tool set with a hierarchy of transport models ranging from the drift-diffusion (DD), through various hydrodynamic(HD) to Monte Carlo (MC) models. Good agreement is achieved between experimental long-channel n-MOSFET drain current data and simulations using the DD, HD, and MC models. The MC simulator is also applied to the study of transport in deep submicron, silicon n-MOSFETs with special attention given to issues related to power supply scaling.

I. INTRODUCTION

Moment-based simulators based on drift-diffusion (DD) and hydrodynamic (HD) formalisms provide tools for device design and research. It is widely recognized that these formalisms do not account for all phenomena of importance to deep submicron device performance. A more complete and physical, but expensive, description of the device behavior can be obtained from Monte Carlo (MC) simulations. While carefully tuned DD simulators do a remarkably good job of predicting device terminal characteristics, especially for longer channel MOSFETs, HD and MC simulators offer deeper insights useful for the design of deep submicron MOSFETs.

In this paper, we analyze the effects of scaling channel length and power supply on device reliability and drive using our HD and MC simulators. Two nMOS transistors with effective channel lengths of 0.13 and 0.08 μm and realistic doping profiles are considered.

II. MOMENT-BASED SIMULATORS

The two-dimensional MOSFET simulator includes a hierarchy of transport models ranging from the traditional DD, a parabolic and a non-parabolic HD [1], an energy transport model and the HD model proposed by Stettler *et al.* [2]. The hierarchical implementation is embedded within the device simulator MINIMOS 5.2 [3]. A unified,, robust and efficient discretization method was used to discretize the hierarchical HD models. We have also extended the field-dependent mobility model of Shin *et al.* [4] to HD applications by replacing the local longitudinal electric field with an "effective" field obtained from the full band MC energy versus field relation. Each of the HD models is cast into a generalized form allowing a uniform numerical discretization for all models. A specific HD model is selected by choosing the values of H, η, R_1, and R_2 [1]:

Current Flow:

$$J = \mu n E + \mu H T \nabla n + \mu n(1+\eta)\nabla(HT) \quad (1)$$

Energy Flow:

$$S = -\tfrac{5}{2} R_1 J T - \tfrac{5}{2} R_2 (n\mu T)\nabla T \quad (2)$$

III. MONTE CARLO SIMULATOR

The Monte Carlo simulator is based on SLAPSHOT [5], a tool that uses analytic fits to the pseudopotential bandstructure of silicon. Advanced features of SLAPSHOT include a scattering rate computation based on the pseudopotential bandstructure and a detailed calculation of the impact ionization rate based on an anisotropic energy threshold. Ionized impurity scattering, acoustic intra- and inter-valley phonon scattering and optical phonon scattering are included. The impact ionization strength was tuned to give good agreement with the experimental ionization rates in bulk silicon [5]. Surface scattering is included via surface roughness and interface fixed charge scattering [6]. Additionally, the transport kernel in the MC device simulator has two windows for repetition in real space

This work was supported in part by the Semiconductor Research Corporation (SRC- 94-SJ-099), Texas Advanced Technology Program (TATP), NSF-PYI, ECS9057633, Advanced Micro Devices (AMD) and Motorola.

and one in the energy domain. Convergence to a self-consistent solution is obtained through iterations with a non-linear Poisson solver.

IV MODEL DEVELOPMENT AND VALIDATION

An important application of our MC simulator has been in the development of our HD model. Quantities such as the relaxation time, field dependence of the average energy and velocity, amount of heat flux have all been determined using SLAPSHOT. As pointed out by Ramaswamy *et al.* [7], MC plays a significant role in verifying the contributions of the various quantities in the HD equations.

The different models in the hierarchy of device simulators will have to demonstrate reasonable agreement in the device characteristics with experimental data for long channel MOSFETs (with smoothly varying lateral fields) before they can be used to study deep submicron MOSFETs. Drain current calculations were performed with all three simulators for a range of drain and gate biases for two different sources of devices. The first device is an LDD MOSFET with an effective channel length of 0.48 microns. As seen from Figure 1, surface scattering plays a significant role in decreasing the drain current in the linear region. However, its role is quite small in the saturation region. The second device is a single drain MOSFET with an effective channel length of 0.32 microns. The agreement with the experimental data is very good (Figs. 1 and 2) and suggests cautious optimism in the use of the HD and MC simulators for shorter channel lengths. This agreement is partly due to the work that has gone into describing surface scattering in the HD and MC codes. Also, recent work by Ramaswamy *et al.* [7] suggests, once again, the utility of our nonparabolic HD model. In the remainder of this work, "HD model" refers to our nonparabolic HD model.

V. CHANNEL LENGTH AND POWER SUPPLY SCALING

HD and MC simulations were performed on two test structures with effective channel lengths of 0.13 and 0.08 microns. These are representative channel lengths for future MOSFET technologies. The devices are single drain MOSFETs with a junction depth of 40 nm, oxide thickness of 5 nm and a step-like channel doping. We looked at single drain MOSFETs to ascertain if they offer reasonably good device performance. We were also interested in device performance at these dimensions under worst case conditions. The doping at the surface (mid to high 10^{17} cm^{-3}) was adjusted to obtain good turn-off characteristics (threshold voltage of 0.5 V in saturation for both the channel lengths) and a deep implant (low 10^{18} cm^{-3}) was included to minimize punchthrough.

For constant lateral fields, shorter channel lengths have shown higher reliability (lower substrate current) [8]. However, operation at peak substrate current can determine the lifetime of a device and is hence of great interest. For this reason gate voltages that maximized substrate current at a given drain bias (power supply) and channel length were chosen.

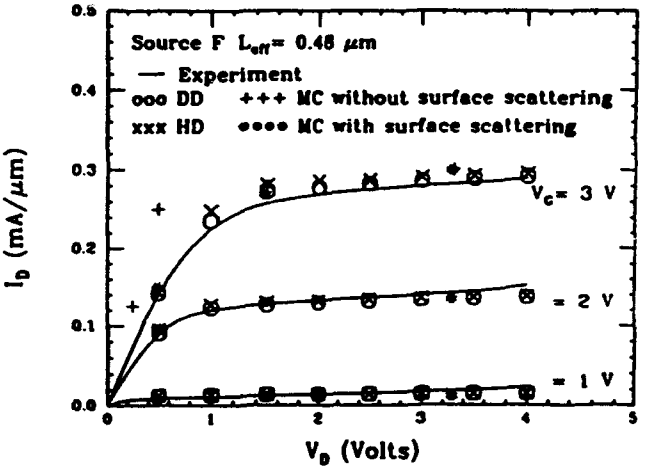


Figure 1: Comparison of the I_D vs V_D characteristics as simulated by the DD, HD and MC tools with experimental data for the device with L_{eff}=0.48 microns.

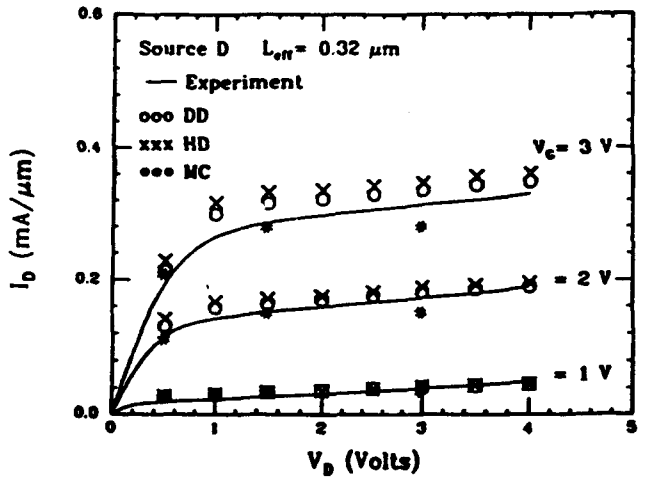


Figure 2: Comparison of the I_D vs V_D characteristics as simulated by the DD, HD and MC tools with experimental data for the device with L_{eff} = 0.32 microns.

VI. RELIABILITY

Substrate current is an index of the amount of impact ionization in a device. Impact ionization degrades threshold voltage, transconductance and affects oxide integrity through electron trapping in the oxide and at the interface. The energy distributions of the carriers provide information on the amount of hot carriers in the device. The average electron distributions are shown in Figures 3 and 4 respectively for the two drain voltages. Even though the two channel lengths exhibit similar distributions in energy, the distribution at about 1.4 eV (where impact ionization takes place) is much higher for the 2.5 V case (and higher for the shorterchannel length). Also, please refer to Figure 5 for a plot showing the average electron energies (from both HD and MC simulations) in the channel. This suggests a much higher substrate current with 2.5 V on the drain.

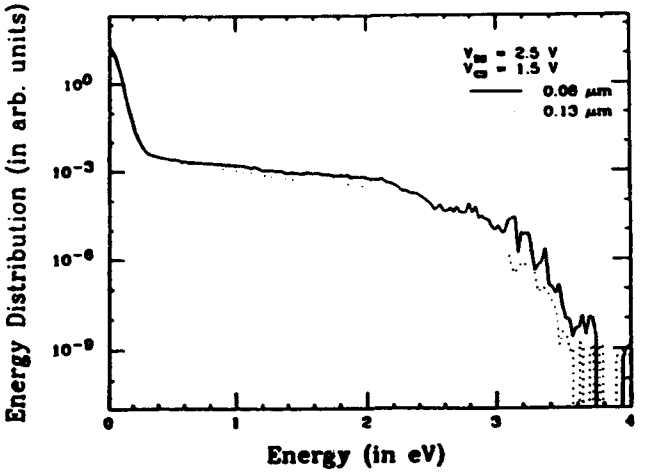


Figure 3: The electron distributions with 2.5 V on the drain and 1.5 V on the gate for channel lengths of 0.08 microns (solid line) and 0.13 microns (dots).

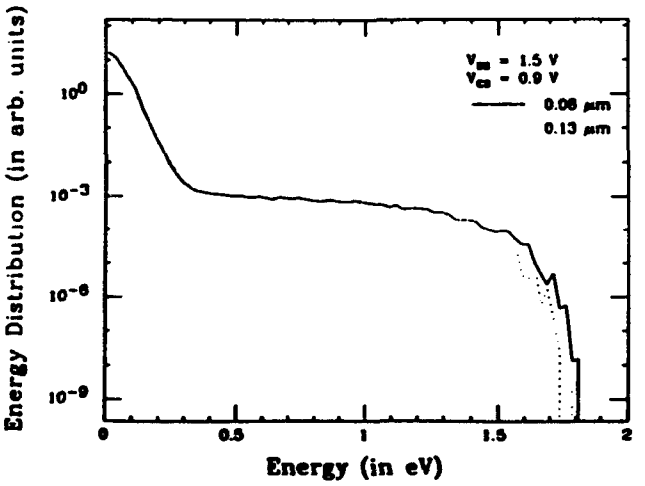


Figure 4: The electron distributions with 1.5 V on the drain and 0.9 V on the gate for channel lengths of 0.08 microns (solid line) and 0.13 microns(dots).

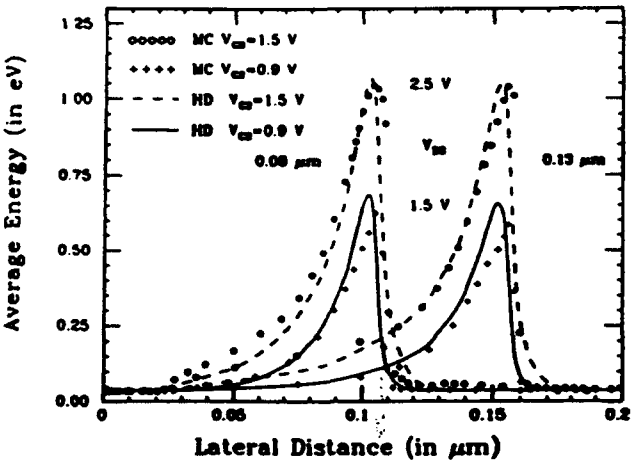


Figure 5: Average electron energies with HD and MC models: 1.5 V on the drain and 0.9 V on the gate for channel lengths of 0.08 microns (solid line) and 0.13 microns (dashes); 2.5 V on the drain and 1.5 V on the gate for channel lengths of 0.08 microns (dots) and 0.13 microns (dots and dashes). Symbols are used for the MC data.

Contrary to the observed dependencies of substrate current on channel lengths for constant field cases, we observe that substrate current increases with a decrease in the channel length when operation at peak substrate current is considered. This supports previous studies [9] which predicted continued degradation of device lifetime with channel length scaling at sub 0.2micron channel lengths. However, as expected, the dependence of substrate current on the drain bias (power supply) is much greater than on the channel length (more so as we get close to the threshold for impact ionization).

Despite the increasing role of interconnects on the overall circuit delay, device speed is still an important issue. The larger the drive current the faster the charge transfer and hence the circuit speed. The drain currents in the two devices were obtained to be 0.23 and 0.3 mA/micron respectively from MC simulations for the 2.5 V case, while for the 1.5 V case, the drain currents in the two devices were 0.07 and 0.09 mA/micron respectively. The average lateral velocities obtained from MC simulations are shown in Fig. 6. Velocity overshoot is comparable over the two channel lengths and drain biases, but the drain currents themselves are

different. This is due to the differences in the velocities before the overshoot (closer to the source). It is interesting to note that the overshoot is actually smaller with 2.5 V on the drain. This is probably due to the fact that average energies are larger with 2.5 V on the drain (as compared to the 1.5 V case) (resulting in increased scattering).

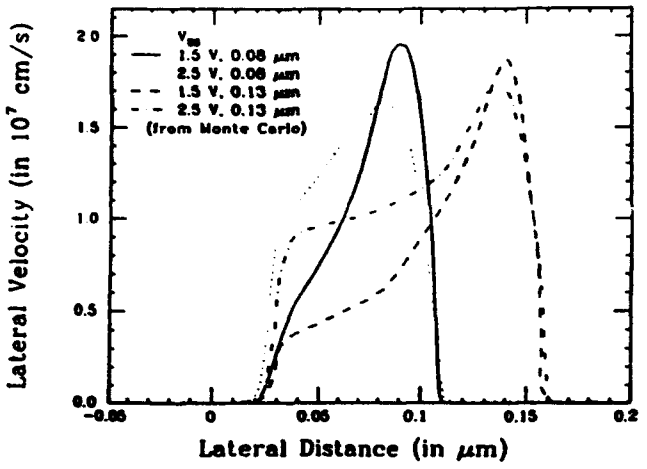


Figure 6: Average lateral velocities with: 1.5 V on the drain and 0.9 V on the gate for channel lengths of 0.08 microns (solid line) and 0.13 microns (dashes); 2.5 V on the drain and 1.5 V on the gate for channel lengths of 0.08 microns (dots) and 0.13 microns (dots and dashes).

It is evident that scaling to sub 0.1 micron channel lengths should be accompanied by a decrease in the power supply. Also, further scaling at sub 0.1 micron channel lengths leads to limited gains in the drive current and negligible increase in the hot carrier degradation. The biggest drawback with scaling the power supply is the decrease in the drive current. A decrease in power dissipation is an added benefit of decreasing the power supply. As long as the device speed is a significant factor determining the overall circuit speed, serious attention must be paid to maximizing the drive current as well. Different structures and/or doping profiles can be investigated for this purpose. The above discussion suggests that a trade-off between drive and substrate currents can be achieved by choosing a suitable power supply with or without additional device design.

VII. SUMMARY

In this paper, we presented an application of a hierarchy of transport models ranging from the drift-diffusion to the Monte Carlo. We demonstrated good agreement between the experimental drain current data and our simulations for devices from two different sources with channel lengths of 0.32 and 0.48 microns. Using this agreement as a first test of model validity, we applied the MC tool to test devices with channel lengths of 0.13 and 0.08 microns. Some issues related to power supply voltage selection such as its effect on device reliability and drive current were highlighted.

References

1. T. J. Bordelon, X. L. Wang, C. M. Maziar and A. F. Tasch Jr., *Solid State Electronics*, 131 (1992).
2. M. A. Stettler, M. A. Alam and M. S. Lundstrom, *IEEE Trans. on Electron Devices*, Vol. 40, No. 4, (1993).
3. MINIMOS 5, *Technische Universitat Wien*, (1991).
4. H. Shin, G. M. Yeric, A. F. Tasch Jr. and C. M. Maziar, *Solid State Electronics*, Vol. 34, No. 6, 545 (1991).
5. X. L. Wang, V. Chandramouli, C. M. Maziar and A. F. Tasch Jr., *J. Appl. Phys.*, Vol. 73, No. 7, 3339 (1992).
6. S. Jallepalli, X. L. Wang, C. -F. Yeap. C. M.Maziar and A. F. Tasch Jr., *TECHCON*, 361 (1993).
7. S. Ramaswamy and T. Tang, *IEEE Trans. Electron Devices*, ED-41, 76 (1994).
8. L. Henrickson, Z. Peng, J. Frey and N. Goldsman, *Solid State Electronics*, 1275 (1990).
9. M. Dutoit, J. -P. Mieville, Z. M. Shi, N. Revil and S. Cristoloveanu, *VLSI Symposium*, 361(1993).

IMPEDANCE FIELD IN SUBMICRON n^+nn^+ InP DIODES

V.Mitin[1], V. Gružinskis[2], E. Starikov[2], P. Shiktorov[2], L. Reggiani[3], L. Varani [4]
[1] *Department of ECE Wayne State University Detroit, MI 48202, USA*
[2] *Semiconductor Physics Institute A. Goštauto 11, 2600 Vilnius, Lithuania*
[3] *Dipartimento di Fisica Università di Modena, Via Campi 213/A, 41100 Modena, Italy*
[4] *Centre d'Electronique de Montpellier, Université des Sciences et Techniques du Languedoc, 34095 Montpellier Cedex 5, France*

Abstract

The physical processes responsible for microwave power generation in submicron n^+nn^+ InP diodes are analyzed through the spatial profiles of the impedance-field spectrum calculated by a closed hydrodynamic approach. The usual subdivision of the n-region into a dead and active zone is carried out. The dead zone is found to manifest itself as a purely real resistance which is practically independent of the frequency. One or more spatial zones which are responsible for the generation are shown to be formed in the active region of the diode. By reducing the n-region length the additivity of the contributions from each part of the device into the generation spectrum is proven.

I. INTRODUCTION

Near-micron n^+nn^+ InP diodes are widely used in modern electronics as generators in the millimeter region of the electromagnetic spectrum [1,2]. To improve the high-frequency performance of these generators various doping profiles [2,3] and a reduction of the n-region length [4,5] are usually suggested. For a proper choice of these parameters and to clarify some problems related to their design, an appropriate physical modeling of these devices is mandatory. To provide the detailed description of the physical processes responsible for the diode performance a quantitative analysis of the parameters which allow for a spatial analysis of various physical quantities has to be preferred. Indeed, by allowing one to construct a spatial map of the device properties of interest, the designing of the device is significantly facilitated. The main aim of this paper is to demonstrate that the impedance field can be successfully used for this sake.

II. THE IMPEDANCE FIELD APPROACH

When considering n^+nn^+ diodes as microwave power generators, the characteristics which describe the capabilities of the device to amplify small perturbations are of great importance. Under current-driven operation, these capabilities can be rigorously described through the local impedance-field, which is given by the ratio of the Fourier components of the local electric-field $\delta E_\omega(x)$ and the total current $\delta j_\omega(x)$ at circular frequency $\omega = 2\pi f$ in point x as: $\nabla Z(\omega, x) = \delta E_\omega(x)/\delta j_\omega(x)$. Since for the one-dimensional structure considered here, the total current is constant in space, the impedance field reflects spatial behavior of possible perturbations of the local electric field caused by harmonic perturbation of j. Integration of the impedance field throughout the structure gives the small-signal impedance of the whole diode, $Z(\omega) = \int_0^l \nabla Z(\omega, x)dx$. Since the impedance field describes the additive contributions which every point of the diode gives to the small-signal impedance, $\nabla Z(\omega, x)$ can be used for a detailed spatial analysis of the diode performance. To simulate the carrier transport in submicron n^+nn^+ InP diodes the full hydrodynamic model [4-10] based on the carrier concentration, drift velocity and mean energy conservation equations coupled with the Poisson equation for the self-consistent electric field is used. This model was demonstrated to provide an excellent agreement with the Monte Carlo calculations for both bulk semiconductors [7-9] and short n^+nn^+ structures [4-6,10]. In the present paper we apply it to study the contributions of the various parts of the diode to $Z(\omega)$ and its dependence on reducing the diode length. For this sake we use an impulsive procedure [10] which enables us to obtain

under the current-driven operation simultaneously the spectra of both $\nabla Z(\omega, x)$ and $Z(\omega)$ in the frequency range of interest.

III. RESULTS AND DISCUSSION

We consider a n^+nn^+ structure at $T = 300\ K$ with parameters which are typical of to-date diode generators [1,2]: the n-region length $l_n = 1\ \mu m$, $n = 2 \times 10^{16}\ cm^{-3}$, $n^+ = 10^{18}\ cm^{-3}$. Abrupt homojunctions between n and n^+ regions are assumed. The cathode and anode n^+-region lengths are taken to be 0.1 and $0.3 \mu m$, respectively.

Under the current driven operation the electron heating in the diode can be considered (to somewhat extent) as a local property. This is illustrated in Fig. 1, which presents a stationary profiles of the drift velocity in the structures with different lengths of the n region calculated for the same total current j_0. Reduction of the n region length does not change the velocity profile in the common region and it looks as a cut of the corresponding part of n-region which is on the anode side. For the case of $l_n = 1\ \mu m$ ($U_d = 8\ V$) the velocity overshoot results in the spatial negative differential-conductivity (SNDC) in the space region $0.3 < x < 1.1\ \mu m$, where the drift velocity decreases with increasing the spatial coordinate. The real part of the small-signal impedance calculated for the considered n^+nn^+ structure is reported for different lengths of the n-region in Fig. 2. For the case of $l_n = 1\ \mu m$ (see curve 1), the amplification condition, $Re[Z(f)] < 0$, is fulfilled inside the two bands: $f = 70 \div 200$ and $250 \div 340\ GHz$ where microwave power generation is possible. We remark that, by shortening the n-region, the condition for amplification shifts to high frequencies, as expected. Figures 3 (a) and (b) report the spatial profile respectively of the real and imaginary parts of the impedance field for the structure with $l_n = 1\ \mu m$. Curves 1,2, and 3 correspond, respectively, to the frequencies $f_1 = 55\ GHz$, $f_2 = 125\ GHz$, and $f_3 = 290\ GHz$. As it follows from Fig. 3, $\nabla Z(f, x)$ is practically independent from frequency in the near-cathode area of the n-region ($x = 0.10 \div 0.30\ \mu m$).

From the comparison with curve 1 of Fig. 1, we find that this is the region where the drift velocity exhibits a sharp increase up to its maximum value. This space region is usually called as the dead zone of short diodes. Comparing Figs. 1 and 3, one can conclude that the dead zone manifest itself as a near cathode region with a pure real and positive resistance which is independent from the frequency up to the plasma range, and the end of the dead zone coincides with the maximum value of velocity overshoot. The remaining part of the n-region, where SNDC takes place, can be considered as the active region of the diode. By increasing the frequency, the active region with negative values of $Re[\nabla Z(f)]$ appears at first close to the anode and then widens and shifts towards the cathode. There, at sufficiently high frequencies several spatial regions with $Re[\nabla Z(f)] < 0$ can appear. In general, the maximum number of active regions which shows up in the spatial dependence of $Re[\nabla Z(f)]$ is equal to the number of generation bands in the frequency dependence of $Re[Z(f)]$. It due to the fact that the curves in Fig. 3 correspond to the growing space-waves of the local electric field starting at the beginning of the active zone and vanishing at the anode contact.

Figure 4 reports the effect of a reduction of the n region length on $Re[\nabla Z(f)]$ calculated at $f_3 = 290\ GHz$. Curve 1 corresponds to $l_n = 1.0\ \mu m$ and curve 2 to $l_n = 0.9\ \mu m$ when the anode n^+-region is shifted to the left up to the first nearest point in which $Re[\nabla Z(f)]$ vanishes. We observe that, in doing so, the second generation band disappears, the new profile practically coincides with curve 1 in the common region, and only one active zones followed by a zone with positive values of $Re[\nabla Z(f)]$ remains. Moreover, $Re[Z(f)]$ of the whole diode becomes positive at $f_3 = 290\ GHz$ (see curve 2 in Fig. 2). To make $Re[Z(f)]$ at this frequency negative again, it is necessary to shift the n^+-anode contact to the second point where $Re[\nabla Z(f)]$ crosses the zero axis. This case is illustrated by curves 3 in Fig. 4 (analogously as in Figs. 1 and 2) which is calculated for $l_n = 0.72\ \mu m$. In this way one removes the near-anode region with $Re[\nabla Z(f)] > 0$ and, as a consequence, the diode can again generate at frequency $f_3 = 290\ GHz$ since its $Re[Z(f_3)]$ becomes again negative (see Fig. 4, curve 3). The generation band of the shorted diode is so extended to the higher frequency range which fully covers the second generation band of the initial diode.

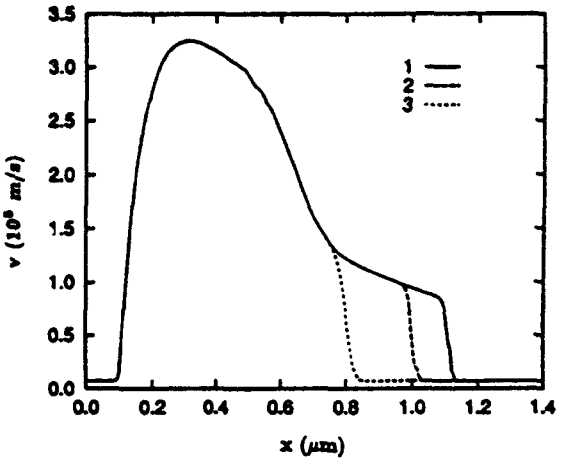


Fig. 1 - Spatial profiles of the drift velocity calculated for n^+nn^+ InP diodes with different n-region length l_n : $1 - 1.0\ \mu m$, $2 - 0.9\ \mu m$, $3 - 0.72\ \mu m$ (curves 1 to 3, respectively).

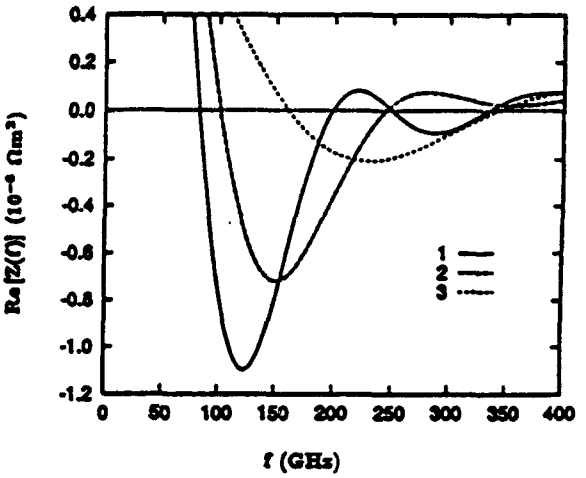
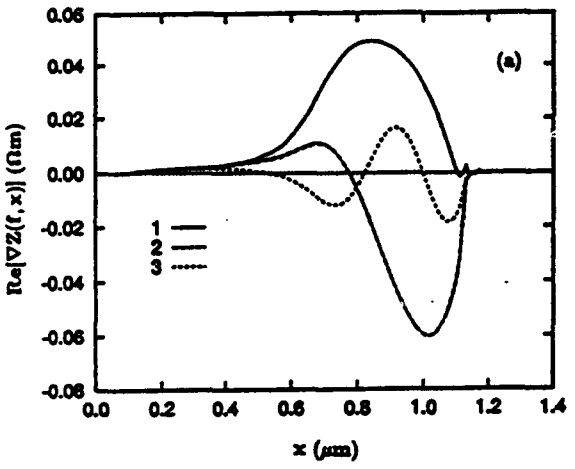


Fig. 2 - Frequency dependence of the real part of the small-signal impedance. The notation is the same of Fig. 1.



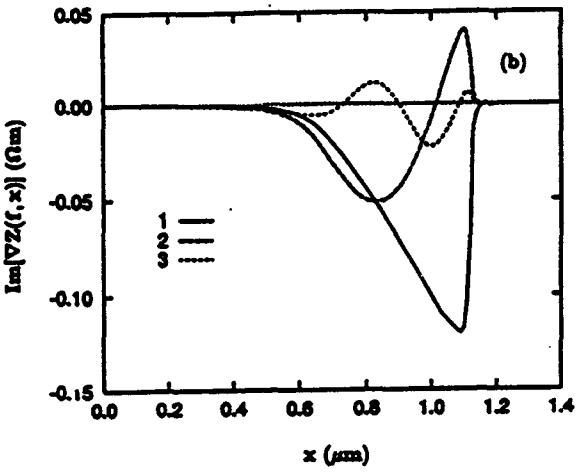


Fig. 3 - Spatial profiles of (a) the real and (b) imaginary part of the impedance field calculated with the HD approach for the n^+nn^+ InP diode of Fig. 1 with $l_n = 1.0\ \mu m$ at three different values of the frequency: $1 - 55\ GHz$, $2 - 125\ GHz$, $3 - 290\ GHz$. $l_n = 1.0\ \mu m$. $U_d = 8\ V$.

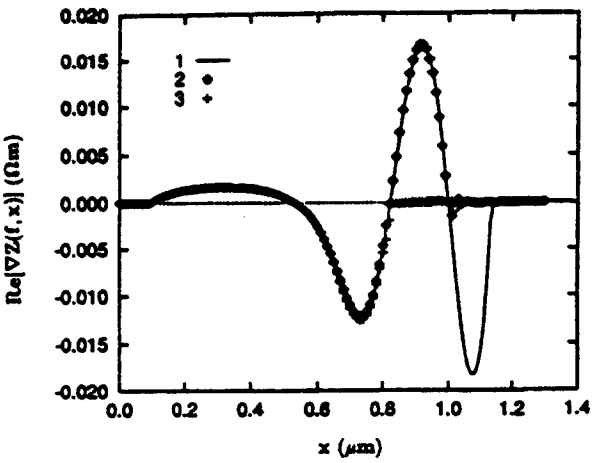


Fig. 4 - Spatial profiles of the real part of the impedance field calculated at the frequency $f = 290\ GHz$. The notation is the same of Fig. 1.

IV. CONCLUSIONS

The spatial dependence of the impedance field we have obtained in the whole frequency range of interest constructs a map which reflects the main physical processes occurring in the different regions of the device and can be used for several purposes such as: to give a comprehensive analysis of the device performance, to provide a proper choice of the device design, etc. Under current operation mode, the diode can be considered as a sequence of seriesly connected zones which give additive contributions to the amplification (and generation) spectrum. Each contribution can be described by a local impedance-field. Moreover, since the carrier flux starts at the source and ends at the drain, the local characteristics depend on the pre-history of carrier motion from the source only and contain no information about a further motion of carriers towards the drain.

References

[1] A. Rydberg, Electron Dev. Lett., **11**, 439 (1990).

[2] R. Kamoua, H. Eisele, G.I. Haddad, Solid-State Electron., **36**, 1547 (1993).

[3] H. Tian, et al., J. Appl. Phys., **72**, 5695 (1992).

[4] V. Gruzhinskis, et al., Appl. Phys. Lett., **61**, 1456 (1992).

[5] V. Mitin, V. Gruzhinskis, E. Starikov, P. Shiktorov, J. Appl. Phys., **75**, 935 (1994).

[6] V. Gruzhinskis, et al., SISDEP **5**, 333 (1993).

[7] V. Gruzhinskis, E. Starikov, P. Shiktorov, Solid-State Electron., **36**, 1055 (1993).

[8] V. Gruzhinskis, E. Starikov, P. Shiktorov, Solid-State Electron., **36**, 1067 (1993).

[9] V. Gruzhinskis, et al., Semicond. Sci. Technol., **8**, 1283 (1993).

[10] V. Gruzhinskis, et al., Noise in physical systems and 1/f fluctuations, P. H. Handel and A. L. Chung eds., AIP Press, New York, 1993, p. 312.

AN ALTERNATIVE METHOD FOR COMPACT MODEL CONSTRUCTION AND PARAMETER EXTRACTION

Edwin C. Kan and Robert W. Dutton
Center for Integrated Systems, Stanford University, Stanford CA 94305

ABSTRACT

The conventional method to extract circuit parameters from device simulators is through I-V and C-V curve-fitting on a presumed device model, whose basic form is often analytically derived using drift-diffusion equations with space-charge-region approximation. The resulting device model is usually either too simple to reflect detailed device behaviors or too complex that most of its internal parameters are rule-of-thumb fitting factors. In consideration of the device simulation level, the profile information on dopings and physical quantities is mostly lost after fitting of only the terminal characteristics. In this paper, we implement an alternative methodology for linking the device and circuit simulators based on the lumped-element (LE) model [1] and a new compiler-based circuit simulation environment [2]. Since the device model is not hard-wired, we show that a more flexible tradeoff between accuracy, predictivity and efficiency may be obtained.

I. INTRODUCTION

Compact device models used in circuit simulation remain as the base of the electronic CAD design system, since it is the first abstraction level beyond the spatial coordinates. Yet compact models were less useful in consideration of predictive designs during technology evolution owing to many nonphysical fittings for the purpose of accuracy. In recent years as the computational resources become more powerful and accessible, technology CAD tools and environment have grown mature on their usage and calibration in the device level. However, on a typical module level with 500 to 10,000 transistors), TCAD tools, even though they can be presumed as very accurate and predictive (for a summary of TCAD limitations and recent developments, see [3]), are still too expensive even in any scale. Hence, the abstraction step toward compact models must be applied. Traditionally, parameter extraction for compact models is based on optimal fit of IV or CV data obtained from lab measurements or simulated TCAD terminal characteristics (see [4-6] for examples). The compact model usually has different levels of complexity, but within each level the circuit linkage is hard-wired. For the most complex level, the physical meaning of each element is often sacrificed for fitting accuracy. This procedure will also lose most of the insight and sensitivity of the detailed profile information provided by the TCAD simulations. Therefore, we choose to implement another method for compact model construction and parameter extraction, which can have flexible configuration inside the model and utilize the profile information from TCAD to construct its parameters. This method can be conveniently implemented in compiler-based circuit simulation environment [2]. We will demonstrate the procedure by a transient study of a 1-D $n+-p$ junction diode [7].

II. The LUMP-ELEMENT METHOD

The lumped-element (LE) method (sometime called the equivalent-circuit method) [1] has been proposed more than three decades ago as an alternative representation of the drift-diffusion (DD) and generation-recombination (G-R) mechanisms inside a device besides the partial-differential

equations (PDE) representation. In LE, the device is partitioned to charge elements, and within the element, the fluxes (J_n the electron, J_p the hole and J_D the displacement currents) are connected with charge storage elements and current sources representing the time derivative of carrier concentrations and G-R, while the continuity equation becomes Kirchhoff's Current Law (KCL) in the newly constructed equivalent circut. If every spatial node in the PDE scheme corresponds to a dual charge element, these two representations can be regarded as equivalent. Nearly all of today's device simulators [3], however, employ the PDE scheme for discretization owing to its better and more convenient numerical properties (such as the Scharfetter-Gummel scheme and tight relations to computational geometry). Nevertheless, for construction of compact models, LE offers very important and direct physical insight. It has been shown that the one-lump model of the bipolar transistor is analogous to the Ebers-Moll model [1]. Moreover, it has been demonstrated that by including the geometrically distributed effects, the compact model can be more flexible and accurate [8]. Yet, the number of lump levels to account for distributed effects can be kept very small [9].

Based on these observations, we implement the LE scheme *only at* the parameter extraction step. Given the solution profiles from device simulators, a more flexible and physically transparent compact model can be constructed accordingly. For a simple example of an $n^+ - p$ junction diode, the compact model can be extracted to optimally fit the IV and CV data (the conventional method, see Fig. 1) or can be constructed so that each element corresponds to some profile variation (the LE method, see Fig. 2). The resistors are identified with variations in potential while the carrier concentrations are almost constant and net charge trivial. The junction capacitor is identified with a net immobile charge dipole, and the diffusion capacitor is identified with a surplus of minority carriers. The values of these circuit elements are bias (or state) dependent (nonlinear resistors and capacitors, current sources and charge-storage elements), and advanced circuit simulators which provide two-terminal table-lookup and nonlinear parameter calculations can take these inputs with little extra efforts [2, 7, 10].

Below we will demonstrate the LE scheme for transient analyis step by step on the 1-D $n+-p$ junction diode. The material parameters and device structure is chosen arbitarily to minimize computational efforts and confusion (such as the device is much longer than the minority carrier diffusion length, simplified doping profiles and physical models in device simulation, etc.). The diode is first simulated with steady-state device simulators. The device states at reverse bias, close to and at equilibrium, at subthreshold and at heavy injection are recorded and analyzed. The charge profiles and corresponding model elements for reverse bias and heavy injection states are shown in Figs. 3-5, respectively. When the circuit element is not necessary at certain state (for instance, the diffusion capacitance at the reverse bias), it is simply given a trivial value in the nonlinear table lookup entry. For detailed analyses on junction operations that can be predicted by device simulators, see [11]. Fig. 6 shows the transient simulation results using the device simulator, the best-fit scheme and the present LE scheme.

III. ANALYSIS

Our approach has the following advantages over the conventional method of parameter extraction using optimal fitting:

1. the linkage and the number of circuit elements is not hard-wired. Since the contours of equal potentials or concentrations become circuit nodes in the new compact model and the tradeoff between accuracy and efficiency can be readily made. Exact accuracy of the device simulator can be obtained by constructing the full node-to-element LE scheme. However, the compu-

tational efforts to achieve acceptable accuracy should be much smaller than full LE (most of the important electronic mechanisms happen at a small portion of the entire device), and much smaller than the PDE scheme using smart-grid adaptation where Steiner points from grid construction can be surplus due to the geometrical constraints.

2. the error in device simulation parameters (such as mobility) can be directly reflected on the circuit element parameters (such as resistance). The errors from the abstraction steps and from the input parameters of device simulation can be easily separated. In comparison with parameters extraction based on the optimal fit of IV and CV data, statistical analysis and worst-case estimation on process variations can be performed more efficiently. Also, since the abstraction over the spatial coordinates has been solely performed in one step, the concept of statistical metrology [12] for manufacturing becomes clear.
3. since the optimal fitting of all element parameters on lumped terminal IV and CV data is never invoked, advanced (and hence time-consuming) numerical algorithms like numerical annealing to find multiple local minimum are not necessary.
4. the circuit elements are physically transparent. The improvement of compact models can be done in an automatic, evolutionary way, based on the the progress of the device simulation.

Acknowledgement: This work is supported by National Center for Computational Electronics (NCCE) through NSF ELS-9200560-A1.

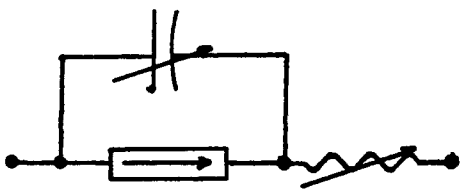


Fig. 1. Compact model of a p-n diode for the optimal-fit method.

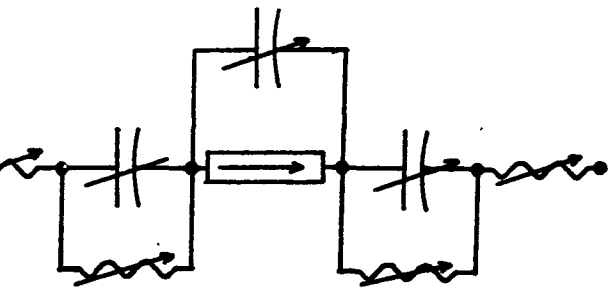


Fig. 2. Compact model of a p-n diode for the equivalent-circuit method. ρ is the net charge.

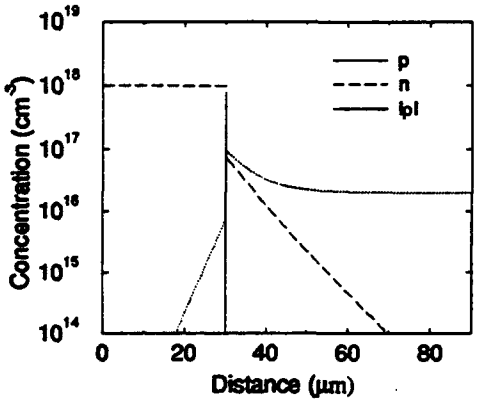


Fig. 3. The concentration profiles at a high-injection state of an $n^+ - p$ diode.

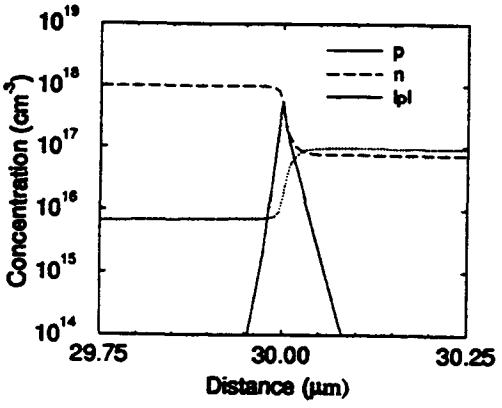


Fig. 4. The junction magnification of Fig. 3. The abrupt junction is at 30 μm.

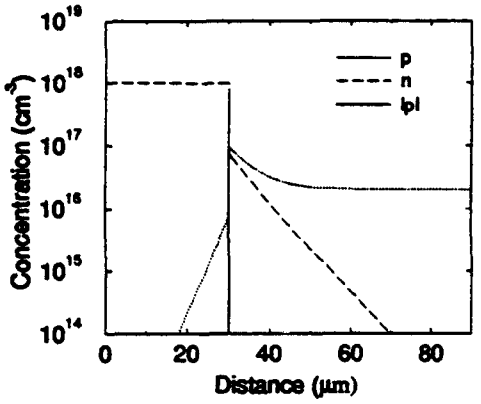


Fig. 5. The concentration profiles at a reverse-bias state.

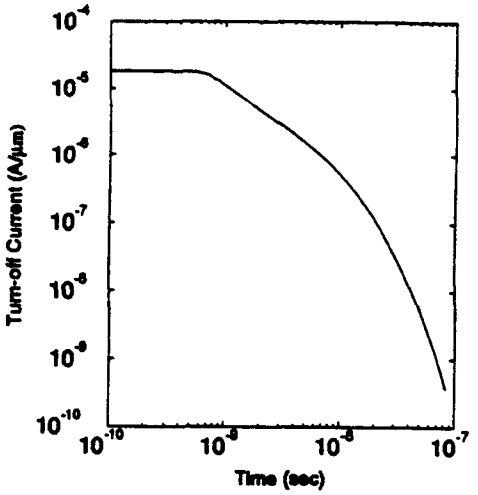


Fig. 6. Transient analysis of switching from the high-injection state to the reverse-bias state.

References

[1] J. G. Linvill, "Lumped models of transistors and diodes, " *Proc. IRE*, pp. 1141-1152, 1958; C. T. Sah, "The equivalent circuit model in solid-state electronics - Part I: the single energy level defect centers, " *Proc. IEEE*, vol. 55, no. 5, pp. 654-671, 1967; C. T. Sah, "Equivalent circuit models in semiconductor transport for thermal, optical, auger-impact, and tunnelling recombination-generation-trapping processes, " *Phys. Stat. Sol*, vol. 7, pp. 541-559, 1971.

[2] A. T. Yang, *MISIM 2.3 User's Manual*, University of Washington, 1993.

[3] F. Fasching, S. Halama and S. Selberherr, *Technology CAD Systems*, Wien: Springer-Verlag, 1993.

[4] P. Yang and P. B. Chatterjee, "An optimal parameter extraction program for MOSFET models, " *IEEE Trans. ED*, vol. ED-30, no. 9, pp. 1214-1219, Sept. 1983.

[5] S. Wang, J. Lee and C. Chang, "An efficient and reliable approach for semiconductor device parameter extraction, " *IEEE Trans. CAD*, vol. CAD-5, no. 1, pp. 170-179, Jan. 1986.

[6] G. J. L. Ouwerling, "Physical parameter extraction by inverse device modelling: application to one and two-dimensional doping profiling, " *Solid-State Electron.*, vol. 33, no. 6, pp. 757-771, 1990.

[7] A. T. Yang, Y. Liu and J. T. Yao, "An efficient nonquasi-static diode model for circuit simulation, " *IEEE Trans. CAD*, vol. 13, no. 2, pp. 231-239, Feb. 1994.

[8] R. W. Knepper, "Modeling advanced bipolar devices for high performance applications, " in *IEDM Tech. Dig.*, 1990, p. 177; R. W. Knepper, S. P. Gaur, F. Y. Chang and G. R. Srinivasan, "Advanced bipolar transistor modeling: process and device simulation tools for today's technology, " *IBM J. Res. Develop.*, vol. 29, no.3, pp. 218-228, May 1985.

[9] N. N. Chan and R. W. Dutton, "Lump partitioning of IC bipolar transistor models for high-frequency applications, " *IEEE Trans. CAD*, vol. CAD-4, no. 2, pp. 143-149, April 1985.

[10] R. R. Daniels, A. T. Yang and J. P. Harrang, "A charge-base transistor model, " *IEEE Trans. ED*, vol. 40, no. 10, pp. 1723-1731, Oct. 1993.

[11] R. W. Dutton and Z. Yu, *Technology CAD: Computer Simulation of IC Processes and Devices*, Boston: Kluwer Academic, 1993.

[12] D. Batelink, "Statistical metrology - at the root of manufacturing control, " presented at Stanford University, Nov. 15, 1993.

TRANSPORT IN TWO-DIMENSIONAL QUANTUM WELL HEMTS

John P. Kreskovsky* and H.L. Grubin*
Scientific Research Associates, Inc.
50 Nye Road, P.O. Box 1058
Glastonbury, CT 06033-6058

Abstract

A parametric study of quantum well HEMT structure is performed through numerical simulations based on a set of quantum hydrodynamics equations. From a reference structure, the effects of variations in planar layer doping, gate recess depth and channel depth on device performance are investigated. The role of the quantum potential in establishing the 2-DEG in the channel is also examined.

I. Introduction

Advances in modeling and computational techniques have now made it possible to simulate and study relatively complex devices on a highly sophisticated level. In this paper we apply a set of quantum hydrodynamic equations to the simulation of an advanced quantum well HEMT to further gain an understanding of the detailed transport within such structures and to demonstrate how such a simulation can be used to obtain quantitative information on how variations of relevant design parameters affect device performance. We begin with the reference structure shown in Fig. 1, an InP-based HEMT with an 800 angstrom InGaAs channel on an AlInAs buffer layer. A 200 angstrom AlInAs Schottky enhancement layer separates the channel from the N^+ InGaAs cap layer, and a 30 angstrom spacer layer separates the channel from the Si doped planar layer. The gate is recessed 100 angstroms into the enhancement layer. Such structures have been the focus of much recent attention; e.g., Refs. [1-3]. The role of quantum mechanics in the analysis of such structures is readily apparent in that the channel is a quantum well in which a 2-DEG gas forms at the interface between the channel and the spacer layer. This 2-DEG is seeded by the planar doped layer.

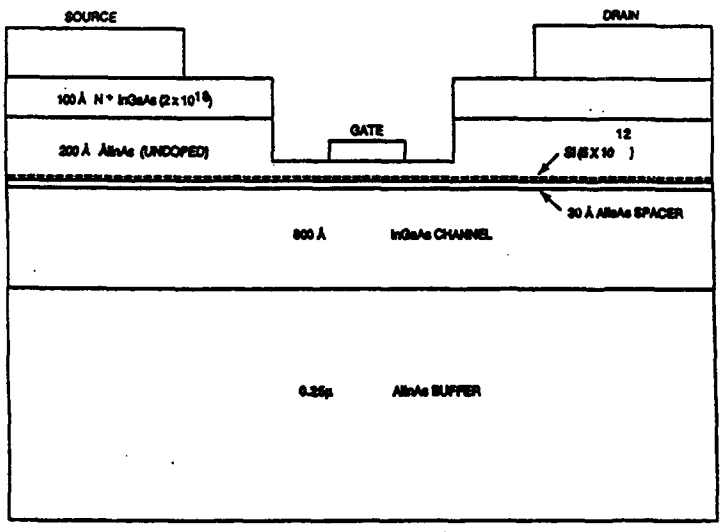


Figure 1. Quantum well HEMT structure used in one and two dimensional simulations.

The equations implemented in our simulation procedure are the hydrodynamics transport equations with corrections for quantum mechanical effects. Various formulations of these equations have appeared in the literature [4].

*Supported in part by ARO and ONR

II. Simulated Results

To begin our simulations we compute the equilibrium solution at zero bias to establish the 2-DEG in the channel. We compare the distribution of electrons in the channel with the classical result in Fig. 2.

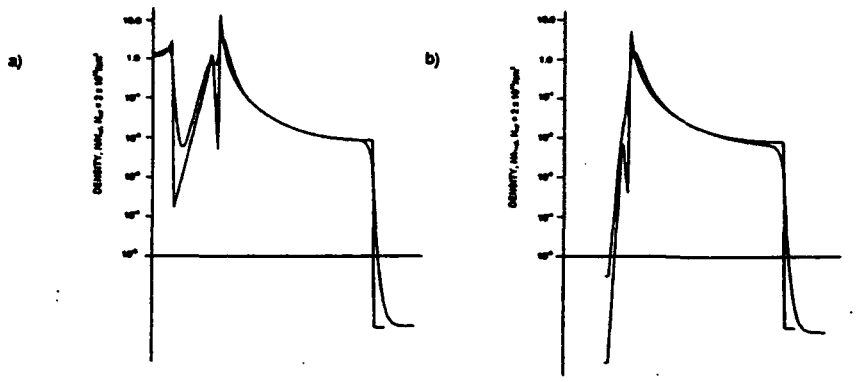


Figure 2. Comparison of classical and quantum corrected density distributions in a a plane normal to the device surface, a) under cap layer, b) under gate.

Figure 2a shows the result under the cap layer and 2b shows the result under the gate. We note that the quantum effects reduce the peak and raise the minimum densities at the heterojunction and yield a continuous density variation. It is also of interest to examine the quantum potential in the region surrounding the gate. Figure 3 shows a blow-up of the quantum potential under equilibrium conditions. The region extends only partly into the channel. The most significant feature here is that the quantum potential shows a nearly one-dimensional structure in the direction normal to the heterojunction interfaces, even at the edge of the gate recess. Some two-dimensionality is observed at the edge of the gate, at the gate surface, but in the channel the structure is still primarily one-dimensional in spite of the depletion of the 2-DEG. This is because the gradients of the density normal to the heterojunction are much greater than those associated with the gate depletion region. The quantum potential thus plays its major role in establishing the structure of the 2-DEG profile normal to the interfaces.

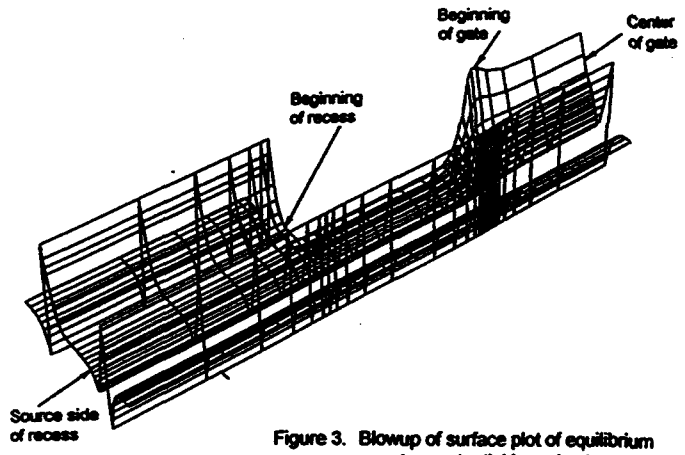


Figure 3. Blowup of surface plot of equilibrium quantum potential in region between the center of the gate contact and to the source side of the gate recess.

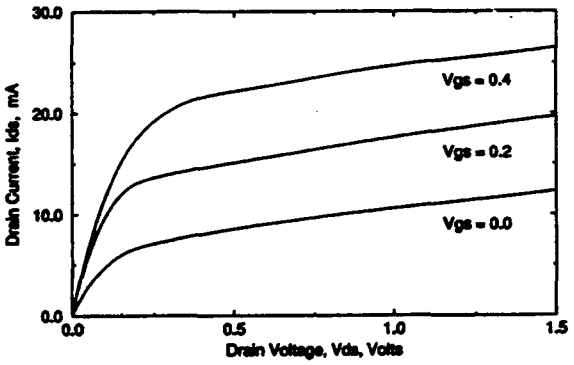


Figure 4. Current-voltage characteristics for reference device structure.

Figure 4 shows the predicted current voltage characteristics for the reference structure and Figure 5 shows surface plots of density, potential, velocity and temperature at a bias of $V_{ds} = 0.5$ volts and $V_{gs} = 0.4$ volts. The high concentration of electrons under the source and drain contacts was introduced to mimic the metalization of the contacts. We also not that the density, potential and velocity in the 2-DEG are almost constant except directly

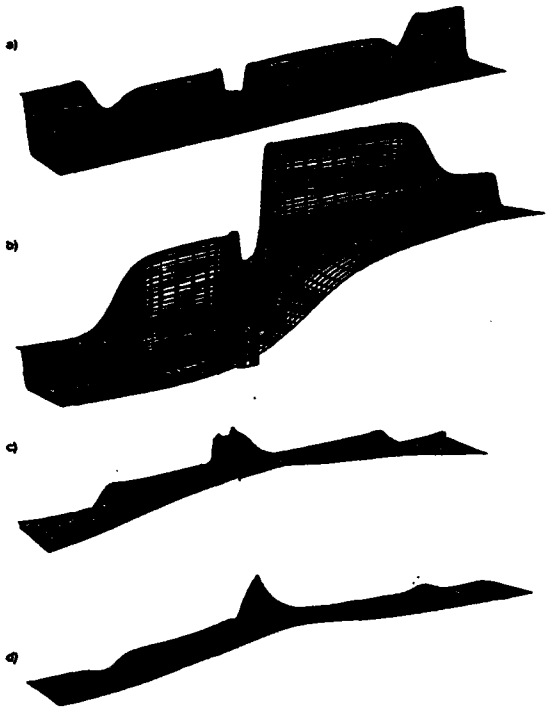


Figure 5. Surface plots of a) density, b) potential, c) velocity and d) temperature at $V_{ds} = 0.5$, $V_{gs} = 0.4$ volts.

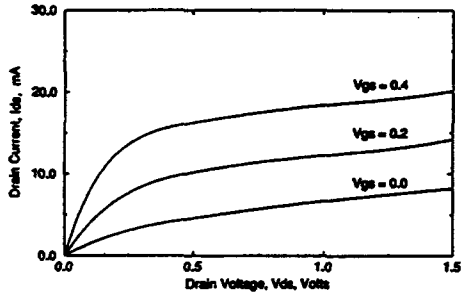


Figure 6. Current-voltage characteristics for a 20% reduction in planar layer doping.

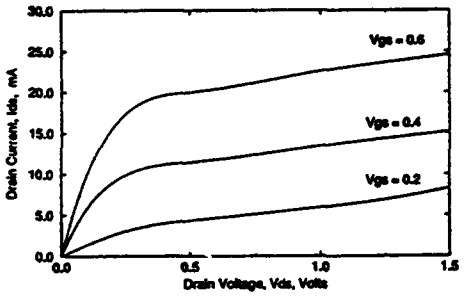


Figure 7. Current-voltage characteristics for a gate recess increased from 100 to 150 angstroms.

under the gate. Under the gate the velocity approaches 1.35×10^7 cm/sec. There is also some heating of the electrons; however, this is surprisingly small. The peak temperature at this bias reaches only 340°K. For this structure we obtained a transconductance of 705 ms/mm, a capacitance of 0.0387 pf and a cutoff frequency of 144.7 GHz at V_{ds} = 1.0 volts and V_{gs} = 0.2 volts. The transconductance was nearly constant over the gate bias investigated while f_τ varied from 156 GHz at $V_{gs} = 0.0$ to 135 GHz at $V_{gs} = 0.4$.

Having established the performance and characteristics of the reference structure we then began our parametric study by reducing the doping of the planar layer. The I-V characteristics for this structure are shown in Fig. 6. The results show a reduction in the current level due to the reduced conductivity of the channel. The saturation characteristics of this device are also slightly harder indicating higher output resistance. The transconductance, capacitance and cutoff frequency for this structure were found to be 587 ms/mm, 0.035 pf and 135.5 GHz, respectively at V_{ds} = 1.0 volts and V_{gs} = 0.2 volts. The transconductance varied from 530 ms/mm to 640 ms/mm, while f_τ ranged from 138 GHz to 131 GHz.

We next increased the gate recess depth to 150 Å. This would be expected to reduce the current levels since the closer proximity of the gate to the channel would result in greater depletion at a given bias level. We would also expect an increase in the transconductance and capacitance. Figure 7 shows that this is indeed the case. The current levels are significantly lower than the reference structure. A bias of 0.6 volts on the gate is required to obtain current levels previously obtained at $V_{gs} = 0.4$ volts. At $V_{gs} = 0.2$ volts and $V_{ds} = 1.0$ volts the transconductance, capacitance and cutoff frequency are virtually the same as the original structure. However, when we compared the results based on the drain current level we found that this structure did exhibit higher tranconductance and capacitance, but the cutoff frequency remained in the range of 145 GHz. At $V_{gs} = 0.4$ volts g_m was approximately 850 ms/ms and the capacitance about 0.45 pf. This structure also showed significantly greater variation in transconductance and capacitance with gate bias than the other structure.

The final simulations were performed for a device in which the channel depth was reduced to 200 angstroms.

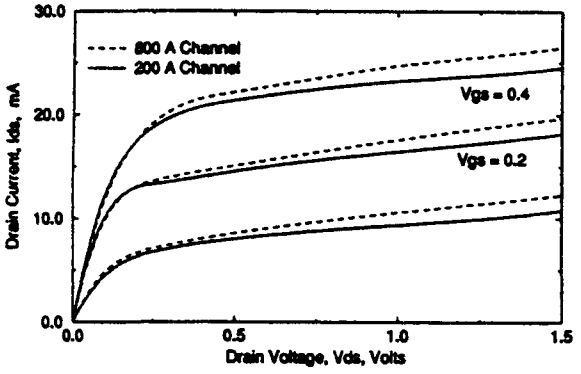


Figure 8. Current-voltage characteristics for a reduction in channel depth from 800 to 200 angstroms.

The current-voltage characteristics for this structure are shown in Fig. 8, where they are compared to the reference structure. The reduced channel depth results in greater confinement and harder saturation. This difference in the output conductance is further reflected in the transconductance, capacitance and cutoff frequency. At $V_{gs} = 0.2$ and $V_{ds} = 1.0$ these quantities were 680 ms/mm, 0.037 pf and 147 GHz. Another interesting result for this structure was that the transconductance decreased with increasing gate bias, from 737 ms/mm at $V_{gs} = 0.0$ to 625 ms/mm at $V_{gs} = 0.4$. In all the other devices the transconductance increased with increasing gate bias. The result was such that at low gate bias the cutoff frequency exceeded 190 GHz but at high gate bias it dropped to a low of 111 GHz. Thus, this device exhibited the highest and lowest cutoff frequencies of any of the devices investigated, depending on the bias level.

III. Conclusions

We have applied a set of quantum corrected hydrodynamic equations to investigate transport in quantum-well HEMTs. The results of the study show the importance of the quantum potential in establishing the distribution of charge in the 2-DEG in the channel. Surprisingly the results also show that quantum effects predominantly influence the density distribution normal to the heterojunction interfaces. The role of the quantum potential in affecting transport in the 2-DEG along the channel, even under the gate, appears limited. This is because the gradients in the density normal to the interfaces are much greater than those along the channel, including channel-wise gradients at the edges of the depletion region. However, the quantum mechanical corrections must be included if the distribution of charge and the charge sheet density of the 2-DEG is to be accurately predicted.

We have also applied our simulation procedure to study the effect of various device design parameters on device performance. In this way we have demonstrated the usefulness of such a procedure in both initial device design and optimization of a device.

IV. References

1) V. Zhao, D.C. Tsui, and P.C. Chao, "Electron Transport in 0.15-μm Gate $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As$ HEMT," *IEEE Trans. Electron Devices*, vol. 40, no. 6, pp. 1067-1070, June 1993.

2) L.D. Nguyen, A. S. Brown, M.A. Thompson and L.M. Jelloian, "50-nm Self-Aligned-Gate Pseudomorphic AlInAs/GaInAs High Electron Mobility Transistors," *IEEE Trans. Electron Devices*, vol. 39, no. 9, pp. 2007-2014, Sept. 1992.

3) R. Plana, L. Escotte, O. Llopis, H. Amine, T. Parra, M. Gayral, and J. Graffeuil, "Noise in AlGaAs/InGaAs/GaAs Pseudomorphic HEMTs from 10 Hz to 18 GHz," *IEEE Trans. Electron Devices*, vol. 40, no. 5, pp. 852-858, May 1993.

4) H.L. Grubin and J.P. Kreskovsky, "Quantum Moment Balance Equations and Resonant Tunneling Structures," *Solid State Electronics*, vol. 32, no. 12, pp. 1071, 1075, 1989.

BREAKDOWN SIMULATION OF SEMICONDUCTOR DEVICES INCLUDING ENERGY BALANCE AND LATTICE HEATING

Y. Apanovich, R. Cottle, E. Lyumkis, B. Polsky, A. Shur, A. Tcherniaev and P. Blakey
Silvaco International, 4701 Patrick Henry Drive, Bldg. 3, Santa Clara, CA 95054, USA

Abstract

A self-consistent nonisothermal energy balance model has been incorporated into a general purpose device simulator, ATLAS. The breakdown characteristics of submicron BJT and SOI transistors have been investigated and compared with the results predicted by simpler models.

I. INTRODUCTION

Most semiconductor device simulation uses the drift-diffusion and isothermal (constant lattice temperature) approximations. These can lead to poor accuracy in predicting the electrical characteristics of modern semiconductor devices. 'Energy balance' models can account for non-local transport effects; and 'nonisothermal' models can account for lattice heating. Most advanced simulation has focused on isothermal energy balance models and on nonisothermal drift-diffusion models. However, models that include both energy balance and nonisothermal effects have started to appear [1-5].

The implementation and use of a nonisothermal energy balance model is described here. The breakdown characteristics of deep submicron BJT and SOI devices are calculated using four different models: isothermal drift-diffusion (DD), nonisothermal drift-diffusion (NDD), isothermal energy balance (EB), and nonisothermal energy balance (NEB). Direct comparisons are made, and interesting physical effects are identified.

II. PHYSICAL MODEL AND NUMERICAL TECHNIQUES

The NEB model is a set of six partial differential equations for electrostatic potential, electron and hole concentrations, electron and hole carrier temperatures, and lattice temperature. The dependencies of all transport parameters on both carrier temperature and lattice temperature are included. The NEB model is an extension of Stratton's energy balance model [6,7], and is similar to the models used in [1] and [5].

A general 2D implementation of the NEB model is now available in the ATLAS device simulator. Numerical solutions are obtained using box integration on a general triangular grid, and Sharfetter-Gummel type discretizations for carrier current and energy flux densities. The implementation of the NEB supports realistic heat-sinks and very general thermal boundary conditions [8]. The fully coupled Newton algorithm, and several decoupled block schemes [9,10], can be used to solve the discretized nonlinear algebraic systems.

III. BJT EXAMPLE

The BJT structure has doping concentrations in the emitter, base, n^- collector, and n^+ collector of 5 10^{19}, 5 10^{18}, 4 10^{17} and 10^{19} cm^{-3} respectively. The emitter, base, n^- collector, and n^+ collector region widths are 50, 50, 100, and 50 nm respectively. The heat flux is set equal to zero at boundaries, except at the bottom of the device where different values of a thermal resistor are connected to a 300 K source. The collector is connected to the collector supply voltage V_{cc} through an external resistance. V_{cc} is ramped with the emitter-base voltage held at -0.7V.

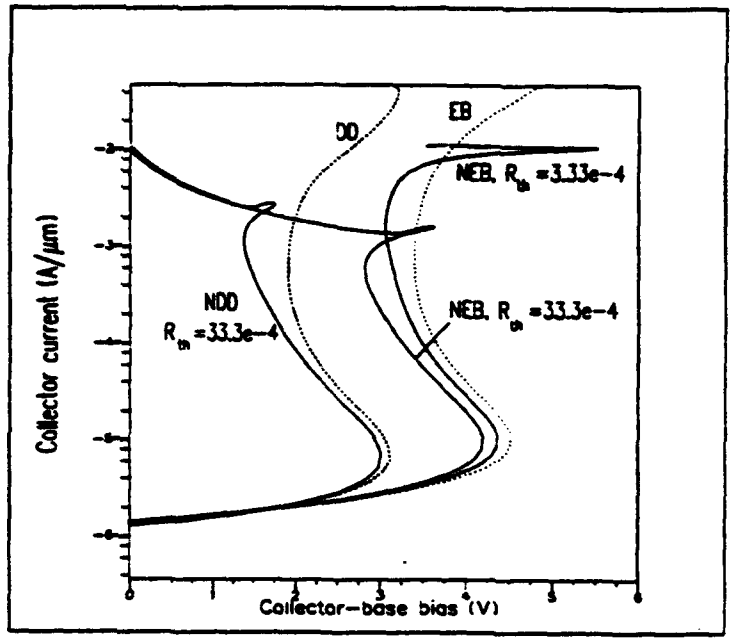


Figure 1. Log collector current versus collector-base voltage for the DD, NDD, EB and NEB models. V_{eb}=-0.7V.

Figure 1 shows the calculated collector current vs base-collector voltage as predicted by the DD, NDD, EB and NEB models. Figure 2 shows the maximum lattice temperature in the device vs collector voltage for the NEB model with Rth=3.33 10^{-4} Kcm^2/W and Rth=33.3 10^{-4} Kcm^2/W, and for the NDD model with Rth=33.3 10^{-4} Kcm^2/W. These results display several interesting features. In the limit of low collector voltage and low current the results are, as anticipated, very close. The shift of breakdown voltage predicted by the EB model, as compared to the DD model, is due to the well-known overestimation of impact ionization in the DD model. If the thermal resistance is set to zero the NEB and NDD produce virtually the same results as the EB and DD, respectively. This is because the active region of the device is very small, and specifying an isothermal boundary condition provides effective "cooling". The situation changes dramatically when a realistic thermal resistance is added to the bottom of the device. For Rth=3.33 10^{-4} Kcm^2/W the NEB predicts almost the same breakdown voltage (first snap back) as the EB. However, the predicted behaviour in the high current region is very different. The NEB predicts second (thermal) breakdown, which is not predicted by the EB. Increasing the value of Rth to a value of 33.3 10^{-4} Kcm^2/W leads to a decrease in the first and second breakdown voltages.

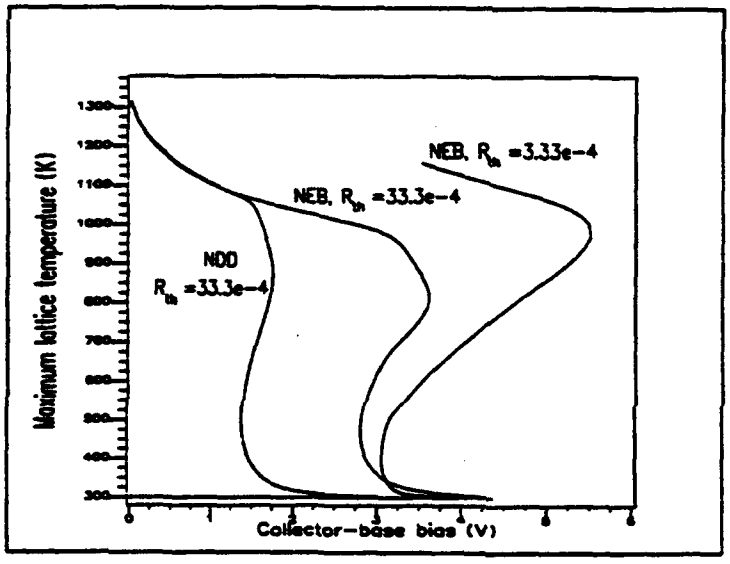


Figure 2. Maximum lattice temperature versus collector-base voltage for the NDD and NEB models. V_{eb}=-0.7V.

The NDD and NEB models predict behavior that is qualitatively similar. For currents around 10^{-2} A/μm the results of the NEB and NDD are quantitaively similar. This is an initially surprising result. The explanation is that carrier temperatures are close to the lattice temperature, and the drift diffusion approximation becomes reasonable for such conditions.

IV. SOI EXAMPLE

The SOI device has gate oxide, body, and substrate oxide thicknesses of 0.017, 0.16, and 0.5 μm respectively. The channel length is 0.5μm and the doping concentration in the channel is 10^{17} cm^{-3}. The lattice temperature is set equal to 300 K on the gate and along the bottom of the device, and the normal component of the heat flux is set equal to zero on the other part of the boundary.

Figure 3 shows the predicted drain current as a function of drain voltage, for a gate voltage of 1.5 V, calculated using the DD, EB, NDD and NEB models. Figure 4 shows the maximum lattice temperature in the device, as a function of drain voltage, calculated using the NDD and NEB models. The significant shift of the breakdown voltage between the DD and EB models is again observed. The NEB and the NDD models show a slight decrease in breakdown voltage compared with the EB and DD models. The predicted behavior in the high current/high temperature region is very different, even qualitatively, between the EB and NEB models. The large difference between the results predicted by the EB and NEB models is due to decreased impact ionization rates at higher lattice temperatures. This indicates that the NEB model is required for accurate simulation in the strong breakdown region.

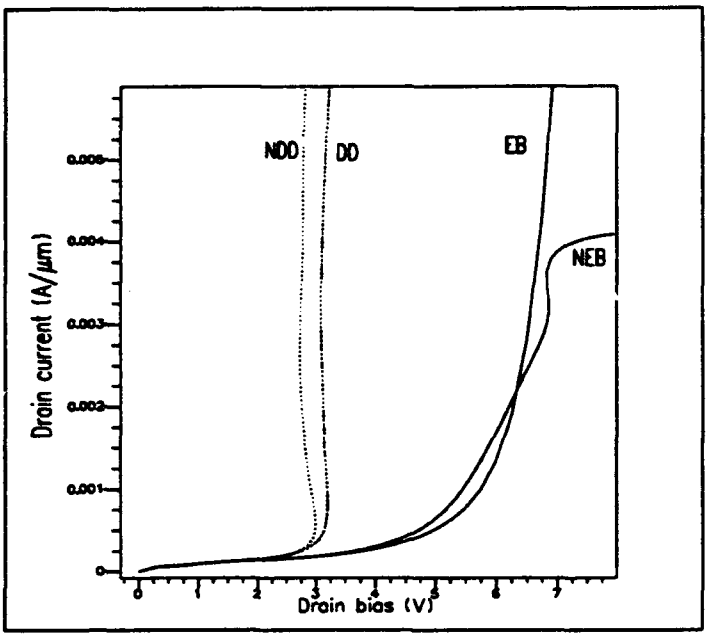


Figure 3. Drain current versus drain voltage for the DD, NDD, EB and NEB models. V_g=1.5V.

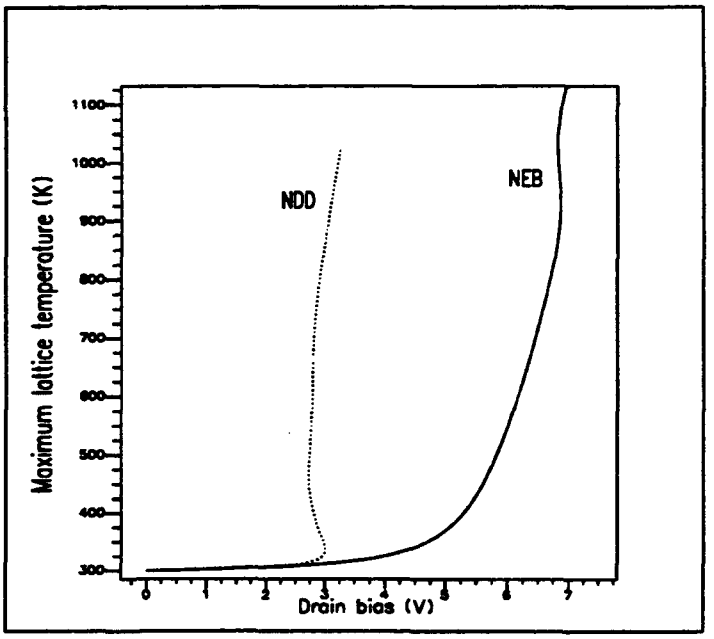


Figure 4. Maximum lattice temperature versus drain voltage for the NDD and NEB models. V_g=1.5V.

V. CONCLUSIONS

A self-consistent nonisothermal energy balance model has been incorporated into a general purpose 2-D device simulator. The breakdown characteristics of submicron BJT and SOI devices have been investigated for the first time with a model of this generality, and have been compared with the results predicted by simpler models. The results demonstrate clearly the influence of both nonisothermal and energy balance effects on the device characteristics in the strong breakdown region, and the magnitude and nature of the discrepancies associated with the use of less general models.

REFERENCES

1. S. Szeto and R. Reif, *Solid State Electron.*, vol. 32, pp. 307-315, 1989

2. A. Benvenuti, G. Ghione, M. Pinto, W. Coughran, and N. Shryer, *Proceedings of the 1992 IEDM*, pp.737-740, 1992

3. K. Katayama and T. Toyabe, *Proceedings of the 1989 IEDM*, pp. 135-138, 1989

4. A. Pierantoni, A. Liuzzo, P. Ciampolini, and G. Baccarani, *Proceedings of SISDEP V*, pp. 125-128, 1993

5. D. Chen, Z. Yu, K. C. Wu, R. Goossens, and R. Dutton, *Proceedings of SISDEP V*, pp. 157-160, 1993

6. R. Stratton, *Phys. Rev.*, vol. 126, pp. 2002 - 2013, 1962

7. R. Stratton, *IEEE Trans. Electon Dev.*, vol. ED-19, pp. 1288-1292, 1972

8. Y. Apanovich, B. Cottle, B. Freydin, E. Lyumkis, B. Polsky, and P. Blakey, *Proceedings of SISDEP V*, pp. 289-292, 1993

9. B. Meinerzhagen, K. H. Bach, I. Bork, and W. Engl, *NUPAD IV Abstracts*, Seattle, USA, pp. 91-96, 1992

10. Y. Apanovich, E. Lyumkis, B. Polsky, and P. Blakey, *Proceedings of SISDEP V*, pp. 233-236, 1993

A GENERAL HYDRODYNAMIC SOLVER FOR DEEP SUBMICRON SILICON DEVICES

Mei-Kei Ieong and Ting-wei Tang
Department of Electrical and Computer Engineering
University of Massachusetts, Amherst MA 01003, USA

Abstract

In order to study the effect of different hydrodynamic (HD) transport parameters/models on the simulation of deep submicron device characteristics, we have developed a general purpose 2-D HD solver which is capable of solving various HD models. The code is written so that it dose not depend on a specific form of the parameters/models which is introduced only at the final stage. There are other unique features of the code which make it versatile and efficient.

I. INTRODUCTION

In recent years, the hydrodynamic (HD) model has become a very popular device simulation tool because of its capability for describing nonstationary and nonlocal phenomena in semiconductor devices and yet requiring less computation time than the more rigorous Monte Carlo (MC) method. Besides Poisson's equation, the hydrodynamic equations (HDE) consist of particle, momentum and energy conservation equations. However, there exist many HD transport models which use different assumptions and approximations [1](e.g., ansatz on the momentum distribution function, simple energy band structure,the relaxation time approximation,...etc.). Many of these approximations become questionable in the simulation of semiconductor devices in the deep submicron regime. With this in mind, we have developed a general purpose 2-D HD solver which is capable of solving different HD models on virtually the same computer code and therefore making it easier to compare the effect of different HD models.

II. PHYSICAL TRANSPORT MODELS

The system of HD equations for electrons used in our solver consists of[1]:

$$\nabla \cdot (n\vec{V}) = 0 \tag{1}$$

$$qn\vec{V} = n\mu^* \left[\vec{F} - \hat{U} \cdot \nabla n - (1 - \lambda_p)\nabla \cdot \hat{U}\right] \tag{2}$$

$$\nabla \cdot (n\vec{S}) = n\left[\vec{V} \cdot \vec{F} - \frac{W - W_o}{\tau_e}\right] \tag{3}$$

$$\begin{aligned} n\vec{S} = {} & \frac{\mu_s^*}{\mu^*} n(W\hat{I} + \hat{U}) \cdot \vec{V} + \frac{\mu_s^*}{q}\left[(W\hat{I} + \hat{U}) \cdot \hat{U} - \hat{R}\right] \cdot \nabla n \\ & - \frac{\mu_s^*}{q} n\left[(1 - \lambda_{ep})\nabla \cdot \hat{R} - (1 - \lambda_p)(W\hat{I} + \hat{U}) \cdot (\nabla \cdot \hat{U})\right] \end{aligned} \tag{4}$$

where $\vec{V} = \langle \vec{v} \rangle$ is the average particle velocity, $\vec{F} = -q\vec{E}$ is the electric force on the electron, $\hat{U} = \left\langle \vec{v}\hbar\vec{k} \right\rangle$ is an energy tensor, $W = \langle \varepsilon \rangle$ is the average particle energy, $\vec{S} = \langle \varepsilon \vec{v} \rangle$ is the average energy flux, $\hat{I}$ is the unity tensor, μ^* is the bulk electron mobility, τ_ε is the energy relaxation time, μ_ε^* is the bulk energy mobility and $\hat{R} = \left\langle \vec{v}\varepsilon\hbar\vec{k} \right\rangle$ is a fourth-order moment. The parameters λ_p and $\lambda_{\varepsilon p}$ are dimensionless constants of order unity which represent the deviation of the collision moments $\vec{C_p}$ and $\vec{C_{\varepsilon p}}$ [1] from their corresponding homogeneous values. Eqs.(1)-(3) represent the conservation of particles, momentum and energy, respectively. Eq.(4) represents the conservation of the third-order moment $\left\langle \hbar\vec{k}\varepsilon \right\rangle$. It has been shown that this general HD model can represent most of existing HD models[3]. For instance, with the choice of $W = \frac{3}{2}k_B T_e$, $U = \frac{2}{3}W$, $R = \frac{10}{9}W^2$, $\mu_\varepsilon^* = \mu^*$ and $\lambda_p = \lambda_{\varepsilon p} = 0$, the above equations reduce to the simplified HD model, or the energy balance equation [2]. The Eqs.(1)-(4) are supplemented by the Poisson equation and the continuity equation for holes.

III. NUMERICAL IMPLEMENTATION

Most of the previous strategies [4] for the discretization of HDE are basically extension of Scharfetter-Gummel discretization scheme[5] which requires the solution of a first-order differential equation. Some difficulties arise when transport coefficients which appear in the coefficient of the differential equation are complicated function of energy. In our simulator, instead, the concept of artificial diffusivity [6] was introduced. Eqs.(2) and (4) can be rearranged and viewd as the convection-diffusion equations for the carrier density, n, and the average energy, W, respectively. It is well known that when the mesh Reynolds number is larger than 1, the numerical solution often shows *wiggles*. In order to overcome the numerical stability, we added the artificial (numerical) diffusivity [6] to the respective diffusion term in Eqs.(2) and (4). Depending on the mesh Reynolds number, α, one may choose artificial diffusivity as:

$$D_a = [(1-\delta)(\alpha Coth(\alpha) - 1) + |\alpha|\delta]\, D_n, \tag{5}$$

where, D_n represents the real (physical) diffusivity, $0 < \delta < 1$ and $\delta = 1$ corresponds to the upwind scheme, while $\delta = 0$ corresponds to the optimum scheme. One can also show that the optimum scheme is equivalent to the Scharfetter-Gummel scheme when the coefficient of the first-order convection-diffusion equation are constant. So far, the optimum scheme, $\delta = 0$, exhibits a good stability in all of the numerical experiments. Since our solver was intended for solving as many different HD models as possible, it was designed in such a way that the user does not have to calculate the derivatives in assembling the Jacobian matrix. We used a numerical finite-difference scheme to compute the derivatives from any user-supplied function. Thus, the final form of the models/parameters can be introduced by users in this stage. The resulting Jacobian matrix requires a further row scaling to ensure that the Jacobian matrix is well-conditioned. The conjugate gradient square (CGS) method with incomplete LU-decomposition (ILU) as preconditioner is used to solve the Jacobian matrix. Finally, a modified Gummel's decoupled method is applied to the solution of HDE's in which the energy balance equation is decoupled from the rest of the equations.

Another feature of our simulator is that it can be run under the parallel mode. The message passing architecture was used to distribute works onto different processors. A portable software system PVM (Parallel Virtual Machine [7]) was utilized to handle all of the processor communications. Various transport models and/or bias conditions can be processed simultaneously. So far, the performance of the parallel mode has been evaluated only on a network of DECstations. An average

speed-up of 8.9 on a network running of 10 DECstations-5000 has been achieved.

IV. NUMERICAL RESULTS

Using our newly developed HD solver, we have simulated several deep submicron devices. The first example is a Si double-gate thin-film SOI MOSFET structure (see Fig.1) The calculated electric field and average electron energy at the front interface with gate length of $L_g = 0.15, 0.1, 0.05\mu m$ are shown in Fig.2 and Fig.3, respectively. Inspite of decrease in the gate length for the same applied V_{ds} and V_{gs}, the peak electric field and the corresponding peak average energy remain relatively unchanged. These results agree qualitatively with the MC data published by[8]. Another expample compares the influence of transport parameters on the simulated characteristic of a thin-film fully depleted SOI MOSFET (not shown) with $t_{oxf} = 7nm, t_{oxb} = 80nm, t_{si} = 30nm$ and $L_g = 0.1\mu m$. The two-dimensional distribution of electrostatic potential is shown in Fig.4. Fig.5 shows the drain current as a function of the drain voltage obtained from (a) the DD model and the HD model with (b) $\lambda_p = 1, \lambda_{ep} = 0$; (c) $\lambda_p = \lambda_{ep} = 0.5$; (d) $\lambda_p = \lambda_{ep} = 0$. It is observed that the HD models (b),(c), and (d) produce a higher drain current than the DD model (a) due to the velocity overshoot effect. The calculated drain current in these models varies as much as 30 percent. As illustrated in Fig.6, the velocity profiles (averaged along the channel) heavily depend on the choice of λ_p and λ_{ep}. The DD model (a) fails to predict the velocity overshoot effect. The velocity in the channel predicted by the model (b) is the largest due to a higher energy in the channel. The velocity predicted by the model (c) does not show any spurious velocity overshoot near the drain junction and is qualitatively in agreement with the Monte Carlo result reported in the literature. The simplified HD model (d) overestimates the velocity near the drain junction while predicting a lower velocity (thus, a lower drain current) in the channel when compared to the models (b) and (c).

V. CONCLUSIONS

We have developed a general purpose 2-D HD solver which can be adopted to almost any existing HD models. The stability of the numerical solution of discretized HDE's is achieved by introducing a suitable amount of artificial diffusity. We have also discussed some unique features of the code which make it versatile and efficient. Finally, the effect of transport parameters, λ_p and λ_{ep}, on the $I_{ds} - V_{ds}$ characteristics of thin-film fully depleted SOI-MOSFET is demonstrated.

ACKNOWLEDGEMENT

This work was supported in part by NSF Grant ECS-9003518 and by a research contract from the IBM SRDC facility in East Fishkill.

REFERENCES

[1] T-w Tang et al., *IEEE Trans. Elec. Dev.*, **ED-40**, 1469 (1993).
[2] R. K. Cook, *IEEE Trans. Elec. Dev.*, **ED-30**, 1103 (1983).
[3] S. Ramaswamy and T-w Tang, *IEEE Trans. Elec. Dev.*, **ED-41**, 76 (1994).
[4] A. Forghieri et al., *IEEE Trans. CAD*, **CAD-7**, 231 (1988).
[5] D. L. Scharfetter and H. K. Gummel,*IEEE Trans. Elec. Dev.*, **ED-16**, 64 (1969).
[6] J. P. Kreskovsky, *IEEE Trans. Elec. Dev.* ,**ED-34**, 1128 (1987).
[7] Al Geist el al., *PVM 3 User's Guide*, Oak Ridge National Laboratory, (1993).
[8] C. Fiegna el al.,*The Proceeding of VPAD*, 102 (1993).

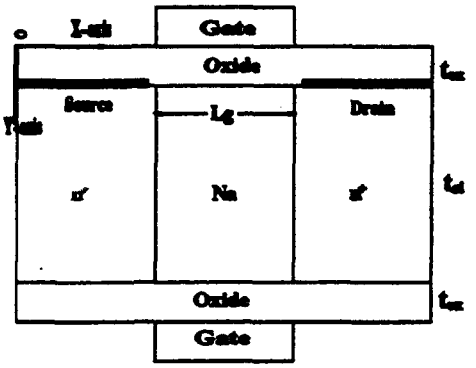


Figure 1: Double-gate SOI MOSFET, $t_{ox} = 3nm$, $t_{si} = 10nm$, $L_g = 0.15, 0.1, 0.05\mu m$,$V_{ds} = 2.5V, V_{gs} = 1.25V$.

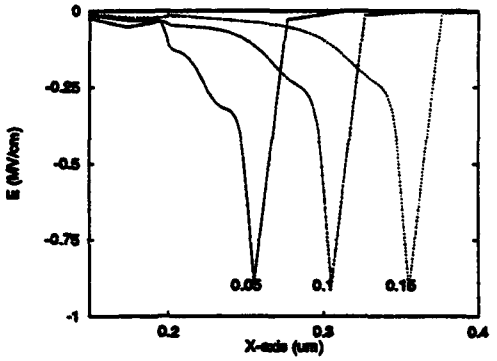


Figure 2: Electric field at the interface, $L_g = 0.15, 0.1, 0.05\mu m$.

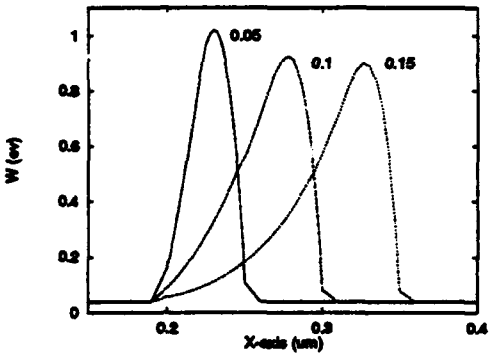


Figure 3: Average electron energy at the interface, $L_g = 0.15, 0.1, 0.05\mu m$.

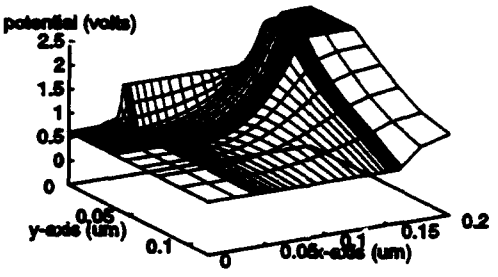


Figure 4: Electrostatic potential distribution for $\lambda_p = \lambda_{cp} = 0.5$, $V_{ds} = 1.8V, V_{gs} = 0.8V$.

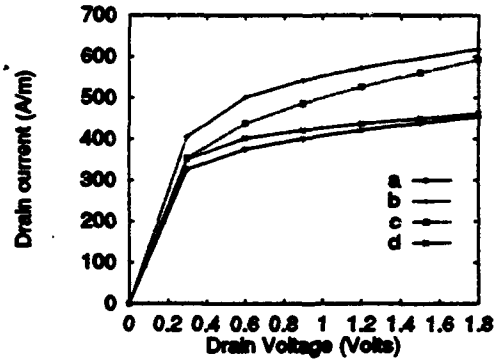


Figure 5: $I_{ds} - V_{ds}$ characteristics for $V_{gs} = 0.8V$.

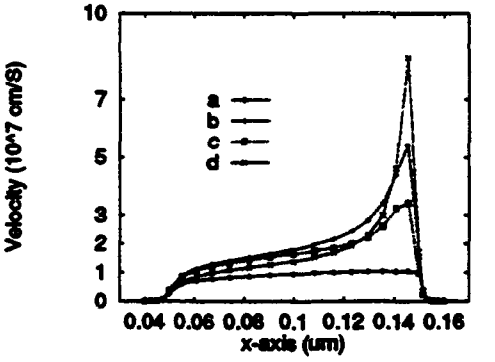


Figure 6: Average electron velocity along the channel.

INCLUSION OF VISCOUS EFFECTS IN THE HYDRODYNAMIC MODELING OF ULTRASMALL SILICON DEVICES

Ting-wei Tang and Joonwoo Nam
Department of Electrical and Computer Engineering
University of Massachusetts, Amherst MA 01003

Abstract

The effect of viscosity on the modeling of the second moment energy tensor $\hat{U}$ was investigated for the narrow base width Si BJT's. The overestimation of velocity overshoot occurring at the base-collector junction often predicted by the conventional hydrodynamic (HD) models can be reduced if a viscous term is included in the modeling of $\hat{U}$. However, the overall effect of including the viscous effect in the HD model on the device characteristics is yet to be investigated.

I. INTRODUCTION

In recent years, many advanced HD models have been developed for the simulation of deep submicron MOSFET's or ultranarrow base width Si BJT's. One of the objectives is to accurately predict the velocity overshoot phenomenon. Although much progress has been made, most of currently existing HD models tend to overestimate the velocity overshoot [1]. This overestimation may become worse, for example, as the base width of BJT's reduces to 0.1 μm or less. In this paper, we propose to include a viscous term in the modeling of second moment energy tensor $\hat{U}$ as a possible mechanism for reducing the velocity overshoot.

II. MODELING OF $\hat{U}$

The macroscopic equation governing the velocity (or momentum) can be rigorously derived by taking the first moment of the Boltzmann transport equation and integrating over the entire momentum space. In the momentum transport equation so obtained, a second-order moment $\hat{U} = \left\langle \vec{v}\hbar\vec{k} \right\rangle$ sometimes called energy tensor or momentum flux tensor appears as a result of taking the moment. The simplest model for $\hat{U}$ is $k_B T_e \hat{1}$ where $\hat{1}$ is a unity tensor. A slightly more advanced model is given by $\hat{U} = \frac{1}{2} m_c \vec{V}\vec{V} + k_B T_e \hat{1}$ [2] where $\vec{V} = \langle \vec{v} \rangle$ and m_c is the effective mass. A recent Monte Carlo (MC) calibrated model gives $\hat{U} = \vec{V}\vec{P} + [\frac{2}{3}W + u(W)]\,\hat{1}$ where $\vec{P} = \left\langle \hbar\vec{k} \right\rangle$, $W = \left\langle \varepsilon(\vec{k}) \right\rangle$ and $u(W)$ is an empirical expression fitted to MC data [3].

In all the models for $\hat{U}$ mentioned above, none of them contains a viscous term. It is well known in gasdynamics that the Navier-Stokes equation includes a viscous term which is proportional to the gradient of velocity. For example, assuming a parabolic band structure and a heated displaced Maxwellian for the distribution function under the homogeneous field, an iterative solution for the inhomogeneous field yields a second-order viscous term for $\hat{U}$, e.g.. in one dimension, $(U_{xx})_{visc.} = -\bar{\eta}\tau_c m_c V_x^2 \frac{dV_x}{dx}$, where τ_c is the collision relaxation time and $\bar{\eta}$ is a dimensionless coefficient of order unity. A recent analysis based on extended thermodynamics [4] gives $(U_{xx})_{visc.} = -\frac{4}{3} k_b T_e \tau_\sigma \frac{dV_x}{dx}$

where τ_σ is the viscous stress relaxation time. In this work, we examine whether or not such a second-order effect should be included in the modeling of $\hat{U}$, and if so, what the approximate value of $\bar{\eta}$ based on the MC data is.

III. MODEL CONSISTENCY TEST

We focus our investigation on ultrasmall base width Si BJT's since the gradient of velocity near the base-collector junction is known to be very large in these devices. A one-dimensional BJT as shown in Fig.1 was used as a prototype for model testing.

Our MC model consistency test proceeds as follows [1]. First, we solve a HD model for the BJT to obtain an electric field profile within the device. Using this electric field, we perform a fixed-field MC particle simulation to obtain all the necessary macroscopic quantities such as $\vec{V}$, $\vec{P}$, W, $\hat{U}$, etc. To test the consistency of a model, say, $\hat{U} = \frac{2}{3}W\hat{1}$ or in one dimension $U_{xx} = \frac{2}{3}W$, we plot U_{xx} vs W using the position in the device as an implicit parameter. Since the average electron energy W always increases from its thermal equilibrium value W_o to a maximum and then falls back to W_o inside the device, the plot of U_{xx} vs W will trace the horizontal axis twice. If U_{xx} vs W traces a curve without a "hysteresis" loop, then U_{xx} is indeed a single-value function of W and $U_{xx} = \frac{2}{3}W$ is a consistent (good) model. In general, U_{xx} vs W will trace a loop and the larger the loop, the poorer the model. Thus, if U_{xx} is modeled as $U_{xx} = V_xP_x + \frac{2}{3}W + u(W)$, we should expect the plot of $(U_{xx} - V_xP_x)$ vs W to exhibit a very small hysteresis since $\frac{2}{3}W + u(W) = U_{xx} - V_xP_x$ is supposed to represent a single value function of W.

In the following,we compare two sets of MC data. One is without the viscous effect, i.e., the plot of $(U_{xx} - V_xP_x)$ vs W. Another, with the viscous effect included, is the plot of $(U_{xx} - V_xP_x + \bar{\eta}\frac{m_c}{q}\mu V_xP_x\frac{dV_x}{dx})$ vs W where μ is the electron mobility and $\bar{\eta}$ is a dimensionless parameter to be adjusted until the loop disappears. As shown in Fig.2, when $(U_{xx} - V_xP_x)$ is plotted against W, a large "hysteresis" loop exists in the high energy range $0.1eV < W < 0.3eV$ and a small one in the low energy range, $0.04eV < W < 0.1eV$. The relatively large loop is a result of a rather crude approximation for $\hat{U}$ by incorporating a tensorial component equal to $\vec{V}\vec{P}$ in the modeling of $\hat{U}$ [3]. This energy range is beyond the range in which velocity overshoot takes place. By including a viscous term we hope that the small hysteresis loop in the energy range $0.04eV < W < 0.1eV$ can be substantially reduced. This is indeed the case as when $(U_{xx} - V_xP_x + \bar{\eta}\frac{m_c}{q}\mu V_xP_x\frac{dV_x}{dx})$ is plotted against W, the small loop virtually disappears if $\bar{\eta}$ is chosen to be approximately $\frac{4}{3}$ (see. Fig.2).

Next, we compare the modeled $\hat{U}$ with the input data from the MC and the $\hat{U}$ directly obtained from the MC. In order to compare the two models, $U_{xx}^{(1)} = V_xP_x + \frac{2}{3}W + u(W)$ and $U_{xx}^{(2)} = V_xP_x + \frac{2}{3}W + u(W) - \bar{\eta}\frac{m_c}{q}\mu V_xP_x\frac{dV_x}{dx}$, we can input the MC data for V_x, P_x, W, etc. on the right-hand side of $U_{xx}^{(1)}$, $U_{xx}^{(2)}$ and compare both of them with the $U_{xx}^{(MC)}$ as a function of position. The comparison between the modeled U_{xx}'s with and without the viscous term and that of MC data is shown in Fig. 3. At first glance, the difference between the two models appears very small. However, since the electron velocity is given by $\vec{V} = -\frac{\mu}{q}\left[q\vec{E} + \hat{U}\cdot\nabla(\ln n) + (1-\lambda_p)\nabla\cdot\hat{U}\right]$ [3], it is easy to see that V_x is very sensitive to the slope, $\frac{dU_{xx}}{dx}$. Based on the MC consistency test [1], the model accuracy for predicting V_x is compared in Fig. 4 for the two U_{xx} models. The effect of the viscosity is seen to reduce the velocity peak in a better agreement with the MC result and is also confirmed by other investigation [5]. We also performed a similar study for $n^+ - n - n^+$ structures with active region less than 0.1 μm. Although inclusion of the viscous term with $\bar{\eta} \approx 2$ makes a

more accurate representation for U_{xx}, its effect on the velocity overshoot is not as significant as in the BJT's.

IV. CONCLUSION

From the MC consistency test, it seems to suggest that a more accurate modeling for $\hat{U}$ should include a viscous term. In one-dimensional case, we found that $U_{xx} = V_x P_x + \frac{2}{3}W + u(W) - \bar{\eta}\frac{m_c}{q}\mu V_x P_x \frac{dV_x}{dx}$ where $\bar{\eta} \approx \frac{4}{3} \sim 2$ fits best with the MC data. This range of coefficient for $\bar{\eta}$ also agrees with values predicted by the others [4]. Since the emphasis of this work is to see whether it is necessary to include the viscous effects in the simulation of narrow base with Si BJT's, the solution to the full set of HD equation including the viscous term was not attempted. A rigorous numerical solution to such a system of equations requires a solution strategy different from the conventional one. This is because when the viscous term is included, the order of differential equation representing the momentum conservation is raised by one and the added viscous term represents a singular perturbation in the limit of vanishing viscosity [5]. Although it is yet to be confirmed, we believe that the viscous effect may be important in modeling of advanced BJT's with ultranarrow bases because the gradient of velocity at the base-collector junction of such devices is usually very large.

ACKNOWLEDGEMENT

This work was supported in part by NSF Grant ECS-9003518 and by a research contract from the IBM SRDC East Fishkill Facility.

REFERENCES

[1] S. Ramaswamy and T.-w. Tang, IEEE Trans. Electron Devices, **ED-41**, 76 (1994).

[2] K. Bløtekjær, IEEE Trans. Electron Devices, **ED-17**, 38 (1970).

[3] T.-w. Tang, S. Ramaswamy, and J. Nam, IEEE Trans. Electron Devices, **ED-40**, 1469 (1993).

[4] A.M. Anile and S. Pennisi, Phys. Rev. B, **46**, 13186 (1992).

[5] A.M. Anile and S. Pennisi, *private communications.*

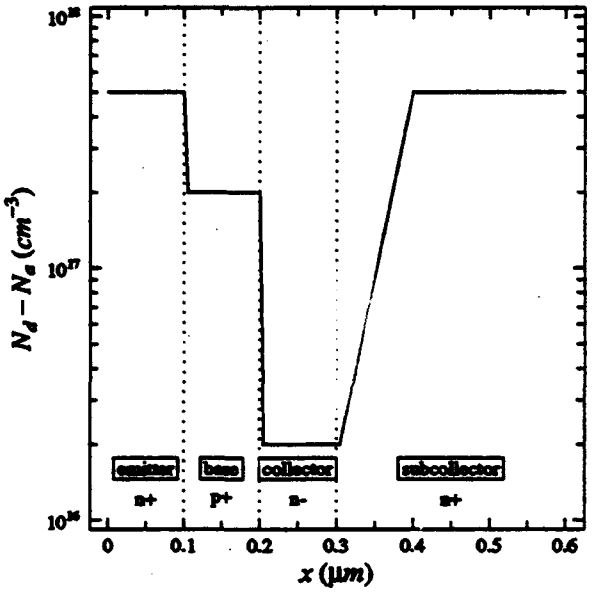


Figure 1: Doping profile of an *npn* BJT.

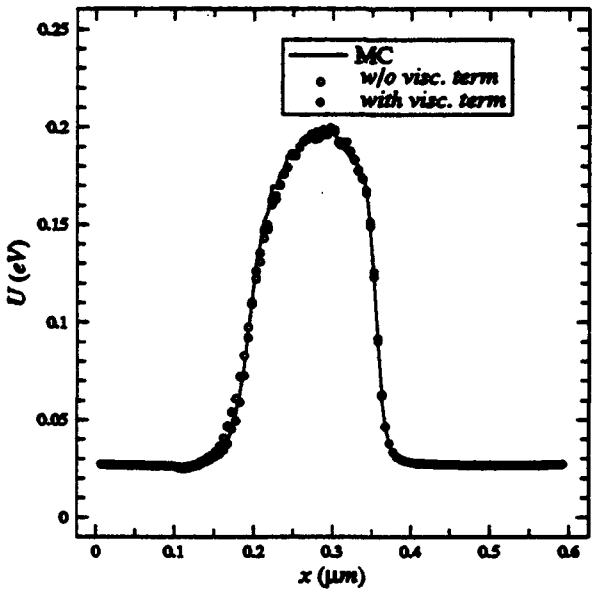


Figure 3: Modeling of U with and without the viscous term as a function of position.

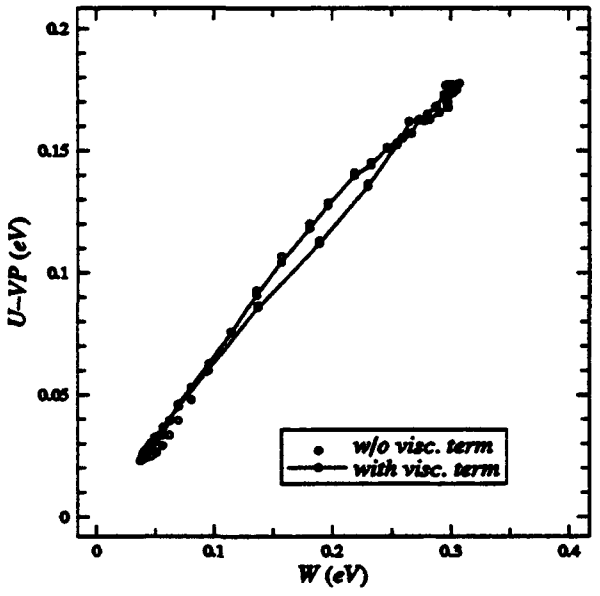


Figure 2: $U - VP$ with and without the viscous term as a function of the average energy W.

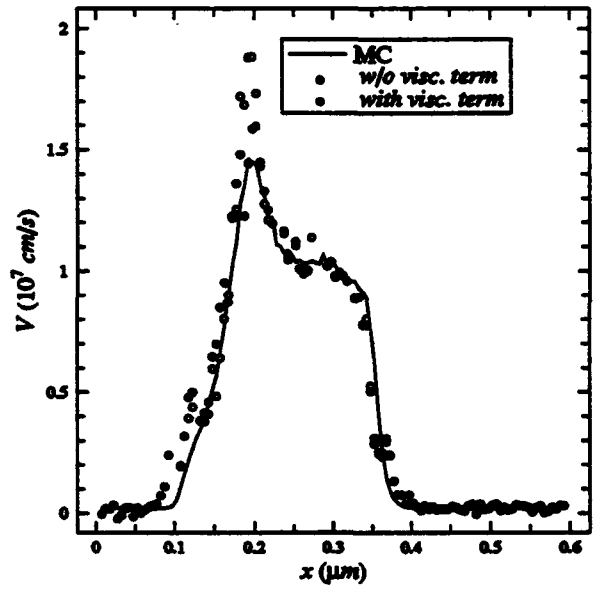


Figure 4: Modeling of V with and without the viscous term as a function of position.

Formulation of A Full Dynamic Transport Model for Heterostructure Devices *

R. Khoie
Department of Electrical and Computer Engineering
University of Nevada, Las Vegas
Las Vegas, NV 89154

Abstract

A Full Dynamic Transport Model consisting of the Momentum Conservation, Energy Conservation, and Particle Conservation Equations, along with Poisson Equation is presented. In this model, the velocity and energy of carriers are taken as *variables* of the system. The resulting system of equations are numerically solved for electrostatic potential, carriers energies, carries velocities, and carriers densities. Most Drift-Diffusion models are based on the fundamental assumption that the transport parameters such mobility, diffusion constant, and drift velocity are constant throughout the structure under any applied biasing conditions. In heterostructures, however, specially in $Al_xGa_{1-x}As/GaAs$ heterostructures, where the velocity-field characteristics exhibit the velocity overshoot, the assumption of constant mobility becomes seriously invalid. The Full Dynamic Transport Model (FDTM) presented here is a nonlinear model incorporating carriers velocities and energies, which eliminates the necessity of calculating the carriers mobilities and diffusion constants. The hot electron phenomena, such as velocity overshoot, have been accounted for by including the third moment of Boltzmann Transport Equation, which provides for energy conservation.

The results of the simulations of a one-dimensional p-n heterojunction are presented. The simulations are performed for different emitter doping levels. It has been found that by increasing the emitter doping from $10^{16}cm^{-3}$ to $5x10^{17}cm^{-3}$, the turn-on voltage of an $Al_{0.3}Ga_{0.7}As/GaAs$ heterojunction diode increases from $1.09V$ to $1.44V$ which is consistent with the results reported by others [1] – [4]. The Maximum velocity of electrons reduces from $2.5x10^7cm/sec$ to $1.8x10^7$ cm/sec, which is due to the increased collision of electrons with doping impurities. This is also the reason for increased electron average energy from $270mev$ to $640mev$. We also observed a significant velocity overshoot in the vicinity of the space charge region of the device. The results obtained are in good agreement with the experimental and simulated results reported by others [1] – [6].

I. The Full Dynamic Transport Model

Most Drift-Diffusion (DD) models presently available in the literature are based on a thermal equilibrium approximation.[3] – [6] The DD models are linear-velocity models in which the electrons are assumed to be subject to a drift force and a diffusive force, and the electron velocity is: $v_n = -\mu_n E - \frac{D_n}{n}\nabla n$. The main difficulty with this approach is in accurate estimation of electron mobility and diffusion constant. Also, in DD models, the effects of high field electron dynamics are neglected due to the thermal equilibrium assumption.

The Full Dynamic Transport Model (FDTM) presented here, is a nonlinear model involving the carriers velocities and average energies, which eliminates the necessity of calculating the mobility and

*This research was supported by the U.S. Army Research Office under ARO Grant No. DAAL 03-87-G-0004

diffusion constants. The particle, momentum, and energy conservations, and Poisson equations are given, respectively, by:

$$\frac{\partial n}{\partial t} + \nabla.(nv) = -R \tag{1}$$

$$\frac{\partial v_n}{\partial t} = -\frac{qE}{m_n} - \frac{2}{3m_n n}\nabla(nw_n) - v_n\nabla.v_n + \frac{1}{3n}\nabla(nv_n^2) + \frac{v_n}{\tau_m} \tag{2}$$

$$\frac{\partial w_n}{\partial t} = -q(v_n \cdot E) - v_n(\nabla w_n) - \frac{2}{3n}\nabla \cdot (nv_n(w_n - \frac{m_n}{2}v_n^2)) - \frac{\nabla}{n}Q_T - (\frac{\partial w}{\partial t})_{coll} \tag{3}$$

$$\nabla.E = \frac{q}{\epsilon}(p - n + C) \tag{4}$$

In the above equations the subscript n represents the variables associated to electrons; with v and w as carrier velocity and average energy, respectively.

In Eq. (1) the term R is the rate of recombination of electrons through traps. In Eq. (2), τ_m is the momentum relaxation time, the term $(-\frac{qE}{m_n})$, is the acceleration due to the crystal potential and the external electric field. The second term, $(-\frac{2}{3m_n n}\nabla(nw_n))$, represents the diffusive acceleration, which forces electrons to move in the direction that minimizes the concentration and energy of electrons. The third term $(-v_n\nabla.v_n)$, causes convective electron flow, giving electrons a tendency to move to an area where electrons move the fastest. The term $(\frac{1}{3n}\nabla(nv_n^2))$ is due to the kinetic energy of electrons. The last term, $(-\frac{v_n}{\tau_m})$, accounts for collisions. The right hand side of Eq. (3) describes how individual forces contribute to changes in the total energy of electrons: the first term, $(-q(v_n \cdot E))$ corresponds to the contribution of the electric field, and the second term, $(-v_n(\nabla w_n))$ is due to the convection of energy flow. The third term, $(-\frac{2}{3n}\nabla \cdot (nv_n(w_n - \frac{m_n}{2}v_n^2)))$, is a combination of the second and fourth terms in Eq. (2). The term $-\frac{\nabla}{n}Q_T$ accounts for the gradients of heat generation, and it is assumed to be negligible. This assumption has been shown to be valid when the distribution function is symmetrical about some mean value in the momentum space, which implies that the temperature is constant throughout the device [7]. The last term, $(-(\frac{dw}{dt})_{coll})$, is the change in energy due to collisions, and is represented by $\frac{w-w_0}{\tau_w}$, where τ_w is the energy relaxation time, and w_0 is the initial average electron energy equal to $\frac{3}{2}kT_o$, and T_o is the effective electron temperature.

In this paper, we study a steady-state case only, and make the assumptions that others [7]-[8] have made, namely, neglecting the convective term, $v\nabla v$, and the terms with the kinetic energy, $\frac{1}{2}mv^2$. The reason for neglecting $\frac{1}{2}mv^2$ is that the total energy : $w = \frac{3}{2}kT_o + \frac{1}{2}mv^2$, and at room temperature the kinetic energy is negligible compared to the thermal energy. Equations similar to Eqs. (1) through (3) are written for holes, providing a system of *seven* equations and *seven* unknowns which is numerically solved for: $\psi, n, p, v_n, v_p, w_n$, and w_p. Auxiliary equations (such as rate of recombination through traps) are used in each iteration cycle.

II. Results and Conclusions

A $n - Al_{0.3}Ga_{0.7}As/p - GaAs$ heterojunction device with $10^{16}/cm^{-3}$ of n-type emitter doping, and 1 μm of emitter thickness is simulated at 300°K. The doping level and thickness of the p-type base are also $10^{16}/cm^{-3}$, and 1 μm, respectively. The simulation program begins with an initial guess for the charge density (ρ), that is used to calculate the electrostatic potential, ψ, by solving the Poisson Equation. The remaining six differential equations are then numerically solved and values of carriers concentrations, energies, and velocities are calculated. A new value for ρ is recalculated from the new n, and p, and Poisson Equation is solved again. When convergence is achieved (with the current value of variables being within 1% of the previous iteration), we proceed to calculate the current densities.

Fig.(1a) shows the electrostatic potential, ψ, in the device for different applied voltages ranging from 1.2 V to -2.2 V. As expected, the width of the space charge region decreases as the forward applied voltage increases. Also, there is no discontinuity of ψ at the heterojunction. Figs. (1b), (1c), and (1d) show the electron concentration n, electron velocity v_n, and electron energy w_n, respectively, throughout the device for different applied voltages ranging from a forward bias voltage of $1.2V$ to a reverse bias of $-2.2V$. There is a discontinuity in the electron density at the heterojunction which is due to ΔE_c. In the n-type region the electron concentration is basically equal to N_D, the donor concentration. In the p-type region the electron concentration depends on the applied voltage.

The electron velocity (Fig 1c) shows the overshoot effect at both ends of the space charge region. There is a large change in the electron velocity at the heterojunction due to different characteristics of the two materials. Also, as the width of the space charge region decreases, the peak value of the electron velocity in the center of the space charge region decreases due to the fact that electrons do not reach the saturation velocity. At the center of the space charge region, the carrier velocity saturates, but the carrier energy increases due to the increase in the electric field.(Fig. 1d) This is because the total energy of the electrons depends on the potential and the kinetic energy. The potential energy dependents on the electric filed which is the highest at the center of the space charge region.

The $i - v$ characteristic of device is shown in Figure (2a). The turn-on voltage is about 1.0 V, and the current increases somewhat exponentially as the voltage increases. To investigate the effects of doping, we increased the doping levels of both sides of the junction from $10^{16}cm^{-3}$ to $5\text{x}10^{17}cm^{-3}$. As shown in Fig. (2b) the turn-on voltage increases from 1.09 V to 1.44 V as the doping level is increased, which is consistent with the results reported by others [1] – [4]. The maximum electron velocity decreases from $2.5\text{x}10^7 cm/sec$ to $1.8\text{x}10^7 cm/sec$ at the edge of the space charge region in the GaAs. This is due to the fact that the increased scattering rate with the ionized impurities tends to slow down the electrons. As the doping level increases, the electron average energy increases from $270mev$ to $640mev$. This is due to the field dependency of the electron energy. The results obtained from our model agrees well with those published by others. (See Tomizawa [9].) Specifically, the velocity overshoot effect that occurs at about $2.0kV/cm$ agrees well with those reported by [9]. The maximum electron velocity has been reported to be $0.7\text{x}10^7 cm/sec$ in $Al_{.3}Ga_{.7}As$ and $2.2\text{x}10^7 cm/sec$ in $GaAs$.[9] From our simulations we obtained a maximum electron velocity of $0.6\text{x}10^7 cm/sec$ for $Al_{.3}Ga_{.7}As$ and $2.5\text{x}10^7 cm/sec$ for $GaAs$. This maximum velocity reduces from $2.5\text{x}10^7 cm/sec$ to $1.8\text{x}10^7$ cm/sec, with increase in doping, which is due to the increased collision of electrons with doping impurities. This is also the reason for increased electron average energy from $270mev$ to $640mev$.

References

1. G. Gao, *IEEE Trans. on Electron Dev.*, Vol. ED-38, No. 2, pp. 185-195, 1991.
2. D. B. Slater, *IEEE Electron Dev. Lett.*, Vol. 12, No. 2, pp. 54-56, 1991.
3. P. D. Rabinzohn, *IEEE Trans. on Electron Dev.*, Vol. ED-38, No. 2, pp. 222-230, 1991.
4. G. Gao, *IEEE Trans. on Electron Dev.*, Vol. ED-37, No. 5, pp. 1199-1207, 1990.
5. Y. Hiraoka, J. Yoshida, *IEEE Trans. on Electron Dev.*, Vol. ED-35, No. 7, pp. 857-862, 1988.
6. K. Yokoyama, *IEEE Trans. on Electron Dev.*, Vol. ED-31, No. 9, pp. 1222-1229, 1984.
7. N. Goldsman J. Frey, *IEEE Trans. on Electron Dev.* Vol. ED-35, No. 9, pp. 1524-1529, 1988.
8. E. M. Azoff, *Solid State Electronics* Vol. 30, No. 9, pp. 913-917, 1987.
9. K. Tomizawa, *IEEE Trans. on Electron Dev.* Vol. ED-37, No. 3, pp. 519-529, 1990.

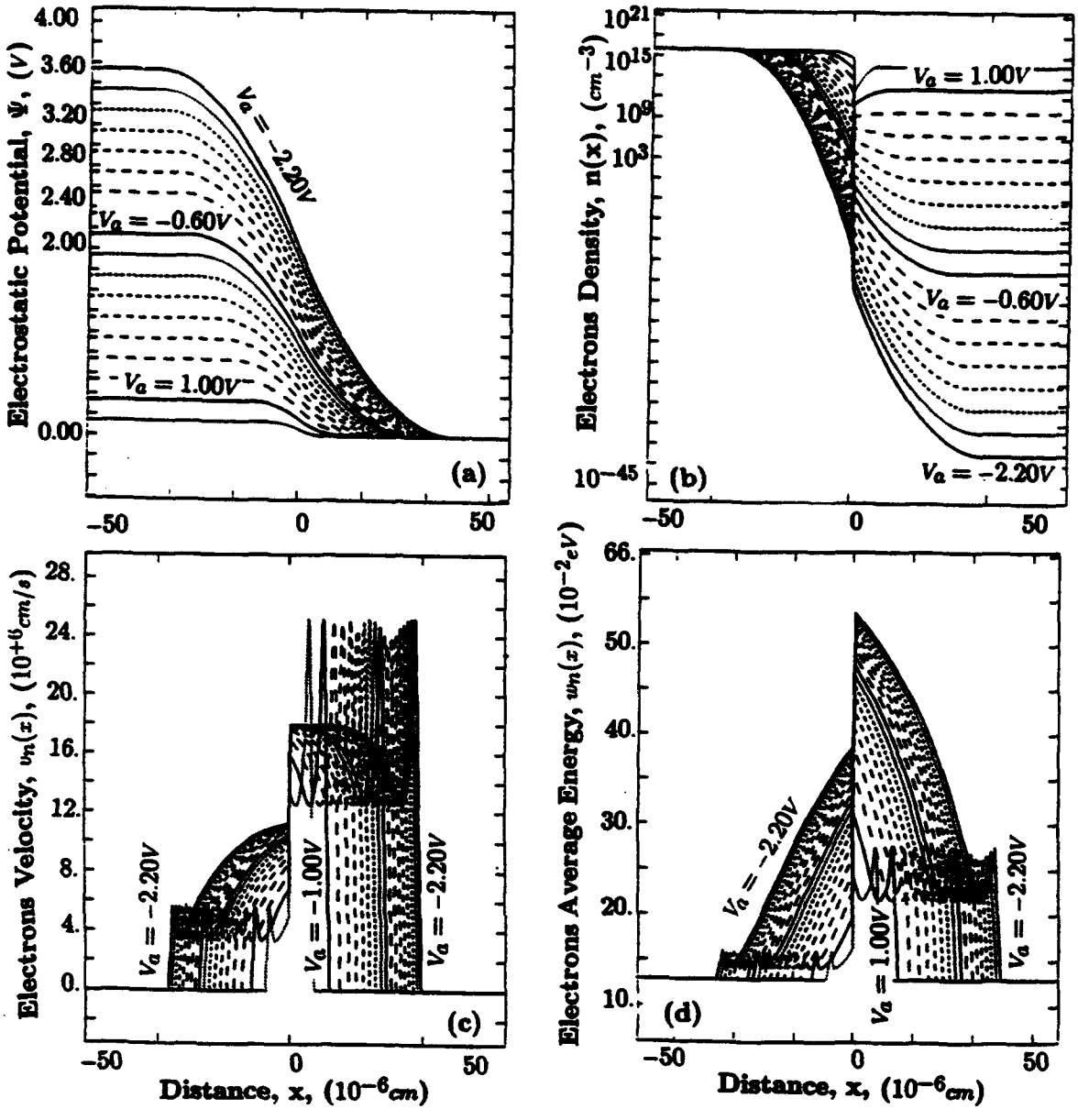


Figure 1: (a) Electrostatic potential, (b) electron density, (c) electron velocity, and (d) electron energy for different applied voltages ranging from forward +1.2 V to reverse −2.2 V, for the device with emitter doping of $10^{16}Cm^{-3}$.

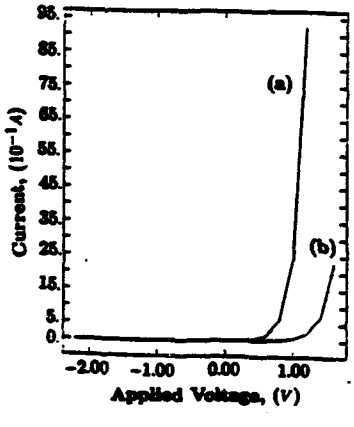


Figure 2: The $i - v$ characteristics of the device with emitter doping of (a) $10^{16}Cm^{-3}$, and (b) $5x10^{17}Cm^{-3}$.

Tunneling and its Inclusion in Analytical Models for Abrupt HBTs

Shawn Searles and David L. Pulfrey
Dept. of Electrical Engineering, University of British Columbia, Vancouver, B.C. V6T 1Z4, Canada.

Abstract

In an abrupt AlGaAs/GaAs HBT, transport in the Conduction Band Spike (CBS) can be the mechanism which limits the overall transport current within the HBT. In this paper closed-form analytic models are presented that describe the transport of carriers in the CBS. These models retain their connection to the physical attributes of the abrupt HBT, yet are simple enough to use in simulators such as SPICE.

I. INTRODUCTION

In an HBT with an abrupt emitter-base heterojunction, and a doping concentration in the narrow-bandgap base that is much larger than that in the wide-bandgap emitter, the Conduction Band Spike (CBS) is as shown in Fig. 1. As has been reported [1]-[5], this CBS plays a vital role in current transport within HBTs and can, in certain cases, completely determine the collector current [5]. Therefore, one must accurately model current transport in the CBS in order to accurately predict the performance of abrupt HBTs.

For the structure depicted in Fig.1, with base doping concentration around 10^{19} cm^{-3} and emitter doping around 10^{17} cm^{-3}, the relevant width of the CBS, as regards tunneling, is about 100 Å. As has been demonstrated by the aforementioned authors, based largely upon the work of Stratton, Padovani, and Christov [6]-[8], an accurate account of tunneling in abrupt HBTs is essential. The basic limitation of the published works regarding the modelling of CBS transport, is that the models can in general only be solved by appealing to numerical techniques; this hides the rich interplay that exists between the physical attributes such as doping concentration, temperature, effective mass, electron affinity, bias conditions, and the final transport model for the CBS. The work to be presented deals with the account of said tunneling in order to arrive at workable, analytic models for current transport across the CBS in abrupt HBTs.

II. THEORY OF CURRENT TRANSPORT IN ABRUPT HBTS WITH TUNNELING

Assuming that transport of electrons through the space charge region is not a limiting factor, then transport in the CBS is determined by both thermionic emission and tunneling. In such a case we find ([5], (32)) that the overall transport current J_C (which in general is equal to the collector current) within the HBT is given by:

$$J_C = q\gamma\upsilon\frac{n_{i,p}^2}{N_A}\exp\left[\frac{-\Delta E_{n0}}{kT}\right]\exp\left[q\frac{N_{rat}V_{BE}}{kT}\right] \tag{1}$$

where

$$N_{rat} = \frac{N_A}{N_A+N_D} \qquad \Delta E_{n0} = \Delta E_c - q\frac{N_D}{N_A}N_{rat}V_{bi} \qquad V_{bi} = \frac{kT}{q}\log\left(\frac{N_A N_D}{n_{i,p}^2}\right)+\Delta E_c$$

and using the notation and results of [1],

$$\gamma = 1+\exp\left[\frac{E_c(0^-)}{kT}\right]\frac{1}{kT}\int_{\max[E_c(x_p),0]}^{E_c(0^-)} D\left(\frac{E_x}{E_c(0^-)}\right)\exp\left[\frac{-E_x}{kT}\right]dE_x \qquad \upsilon = \left(\frac{kT}{2\pi m_n^*}\right)^{1/2} \tag{2}$$

with $N_{A(D)}$ being the base(emitter) doping concentration, $n_{i,p}$ the intrinsic carrier concentration in the base, ΔE_c the conduction band discontinuity, V_{bi} the built-in potential, V_{BE} the applied base-emitter potential,

and υ the effective electron velocity that results from integrating over all particle velocities with some component parallel to the direction of the transport charge flow.

γ is the tunneling factor [1] and (2) is valid as long as the energies E_x over which there is significant contribution to the tunneling current are well above the electron quasi-Fermi energy. Furthermore, due to the limits of integration for γ, it is assumed that the transmission coefficient $D(U)$ is unity above the CBS, i.e., the WKB approximation for the calculation of $D(U)$ is asserted. Using the depletion approximation for an abrupt metallurgical junction, and assuming a coincident metallurgical/hetero-junction interface, $D(U)$ is [1], [7]:

$$D(U) = \exp\left[\frac{2}{q\hbar}\sqrt{\frac{m_n^*\varepsilon_n}{N_D}}E_c(0^-)\left(\log\left(\frac{\sqrt{1-U}+1}{\sqrt{U}}\right)U-\sqrt{1-U}\right)\right]. \tag{3}$$

In (3), U is the normalised energy and is given by $U=E_x/E_c(0^-)$, where $E_c(0^-)$ is the height of the CBS and is given by $E_c(0^-)=qN_{rat}(V_{bi}-V_{BE})$, i.e., $U=1$ at the top of the CBS. In order to gain a familiarity with (2) and (3), Fig. 2 plots the normalised emission flux density (given by the integrand in (2)) that emerges to the right of the CBS for an abrupt HBT with the following material parameters: N_D: $5\text{x}10^{17}\,\text{cm}^{-3}$; N_A: $1\text{x}10^{19}\,\text{cm}^{-3}$; ε_n: $12.2\varepsilon_0$; 30% Al in the emitter; ΔE_c: 0.24 eV; $n_{i,p}$: $2.25\text{x}10^6\,\text{cm}^{-3}$; $\rightarrow \Delta E_i$: 77.3 meV; V_{bi}: 1.671 V; m_n^*: $0.091m_0$. Unexpectedly, the normalised energy U for the peak emission flux density is not a function of applied bias. After some manipulation of the integrand in (2) (using (3)), it is found that the energy for peak emission U_{max}, and the normalised peak emission flux density F_{max} are:

$$F_{max} = \exp\left[-\frac{qN_{rat}(V_{bi}-V_{BE})}{kT}\frac{\tanh(U_p)}{U_p}\right] \qquad U_{max} = \cosh^{-2}(U_p) \qquad U_p = \frac{q\hbar}{2kT}\sqrt{\frac{N_D}{m_n^*\varepsilon_n}}.$$

The fact that U_{max} is independent of the applied potential is interesting in that, relative to the top of the CBS, the emitted electron flux density is always centred at the same place. Discussion of this result will follow in Section III.

Now, given that U_{max} is independent of applied potential, that F_{max} has an exponential characteristic, and the emission flux density has a highly symmetric shape (Fig. 2), there promises to be a potentially simple analytic result for evaluating γ in (2). Through a series of transformations the normalised emission flux can be written as:

$$\int D(U)\,e^{\frac{-UE_c(0^-)}{kT}}\,dU = \int \frac{1}{r}\frac{dy}{dr}e^{C_1 y}dr \qquad C_1 = \frac{E_c(0^-)}{U_p kT} \qquad \begin{aligned} y &= r\cosh^{-2}(U_p+r)-\tanh(U_p+r) \\ U &= \cosh^{-2}(U_p+r) \end{aligned} \tag{4}$$

Equation (4) provides for the exact solution to the tunneling current. If the transform function $y(r)$ were invertible so that $r(y)$ could be determined, then (4) would yield the desired solution in the y domain. However, $r(y)$ cannot be determined analytically in an exact form, but does yield to an approximate form, e.g., the second-order expansion given in (5). Using this approximate $y(r)$ to solve the integral in (4), with limits of $\pm\infty$ (which implies that most of the emission flux should be contained within the limits specified in (2)), then γ is given by:

$$\gamma \approx 1+\sqrt{\frac{4\pi\sinh(U_p)\,U_p E_c(0^-)}{\cosh^3(U_p)\,kT}}\,e^{\frac{E_c(0^-)}{kT}\left(1-\frac{\tanh(U_p)}{U_p}\right)} \qquad \text{where} \quad y \approx -\frac{\sinh(U_p)}{\cosh^3(U_p)}r^2-\tanh(U_p). \tag{5}$$

Equation (5) is the simple analytic form for the tunneling factor γ that we desire; its simplicity suits it to implementation in simulators such as SPICE. Further simplification of γ is possible by dropping the factor of 1, such as would be appropriate in cases where the tunneling significantly exceeds the thermionic emission current.

III. DISCUSSION

Examination of (1) shows that J_C is proportional to γ. Thus, the quantum mechanical nature of the CBS directly manifests itself, through γ, in the determination of J_C. This result reaffirms the statement that modelling the current transport in the CBS is of paramount importance to the understanding of abrupt HBTs.

Further consideration of the subsidiary equations for U_{max} and U_p reveals, the following general traits: as U_p increases from 0 towards infinity, U_{max} tends from 1 towards zero, and tunneling becomes increasingly dominant over thermionic emission; as N_D increases, or ε_n decreases, the width of the CBS decreases and U_{max} becomes smaller, showing that tunneling is increasing; as m_n^* decreases the probability of tunneling should increase, as is confirmed by the associated reduction in U_{max}; finally, in the limit as $\hbar$ goes to zero, the system should evolve to a state that is purely describable by classical mechanics, and it is found that U_{max} goes to 1, which indicates that there is indeed no tunneling. Therefore, the general traits of the emission flux, as presented, follow physical expectations.

Before presenting the final form of J_C, with γ from (5) included, the error associated with the form given by (5) is illustrated via the plots shown in Fig. 3. Note that as T increases γ decreases; this is expected as more carriers can be thermally excited at higher T, and thus tunneling becomes less important relative to thermionic emission. The discrepancy between (5) and the exact form (2) at first decreases with bias. This is because the exact lower limit of integration (used in (2) and shown in Fig. 2) tends to the limit of $-\infty$ as V_{BE} increases. This improvement in accounting for the emission flux at energies below the maximum U_{max}, more than outweighs the discrepancy at higher energies which increases with bias (see Fig. 2 and note the placement of the upper limit). This latter discrepancy amounts to an inclusion of the thermionic-emission flux in the tunneling integral, i.e. a double-counting in γ of the emission flux density above the peak of the CBS. It is this double-counting that results in the increasing discrepancy between (5) and the exact form (2) at high biases.

The final form for J_C is achieved by substituting (5) into (1) to give:

$$J_C = q\upsilon\frac{n_{i,p}^2}{N_A}\sqrt{\frac{4\pi\sinh(U_p)\,U_p E_c(0^-)}{\cosh^3(U_p)\,kT}}\,e^{\frac{-\Delta E_{n0}}{kT}}\,e^{q\frac{N_{rat}V_{bi}}{kT}\left(1-\frac{\tanh(U_p)}{U_p}\right)}\,e^{\frac{N_{rat}\tanh(U_p)}{U_p}\frac{qV_{BE}}{kT}}. \qquad (6)$$

Examination of (6) shows that J_C is basically proportional to $\exp[qN_{rat}\tanh(U_p)V_{BE}/(U_pkT)]$ (this is achieved by disregarding the small variation with bias of $E_c(0^-)$ in the square root term of (6)). Thus we find the customary exponential relationship between J_C and V_{BE} that is found experimentally. However, we now realise that the injection index n is not 1 (as is given by Shockley boundary conditions) but is instead given by $n=U_p/(N_{rat}\tanh(U_p))$. For the device considered in Section II this gives $n = 1.13$, which is almost exactly what is found experimentally. In fact, the slightly larger values found for n experimentally can be accounted for by the bias dependence of the term in the square root of (6).

IV. CONCLUSION

We have achieved a tractable, analytic formulation for both the tunneling factor γ (5) and the transport current J_C (6), and both formulations are suitable for implementation in simulators such as SPICE. Finally, due to the analytic nature of these results, clear physical insight into the connection between material parameters and device operation is obtained, e.g., the new formulation for the injection index of J_C.

V. REFERENCES

[1] A.A. Grinberg, M.S. Shur, R.J. Fischer and H. Morkoç, "An investigation of the effect of graded-layers and tunneling on the performance of AlGaAs/GaAs heterojunction bipolar transistors", *IEEE Trans. Electron Dev.*, vol. 31, 1758-1765, 1984.

[2] M.S. Lundstrom, "An Ebers-Moll Model for the Heterostructure Bipolar Transistor", *Solid-State Electronics*, vol. 29, 1173-1179, 1986.

[3] A.A. Grinberg and S. Luryi, "On the Thermionic-Diffusion Theory of Minority Transport In Heterostructure Bipolar Transistors", *IEEE Trans. Electron Dev.*, vol. 40, 859-866, 1993.

[4] D.L. Pulfrey and S. Searles, "Electron Quasi-Fermi Level Splitting at the Base-Emitter Junction of

AlGaAs/GaAs HBTs", *IEEE Trans. Electron Dev.*, vol. 40, 1183-1185, 1993.

[5] S. Searles and D. Pulfrey, "An Analysis of the Importance of Space-Charge-Region Recombination in HBTs", *IEEE Trans. Electron Dev.*, vol. 41, 476-483, 1994.

[6] R. Stratton, "Theory of Field Emission from Semiconductors", *Physical Review*, vol. 125, 67-82, 1962.

[7] F. Padovani and R. Stratton, "Field and Thermionic-Field Emission in Schottky Barriers", *Solid-State Electronics*, vol. 9, 695-707, 1966.

[8] S. Christov, "Unified Theory of Thermionic and Field Emission from Semiconductors", *Phys. Stat. Sol.*, vol. 21, 159-173, 1967.

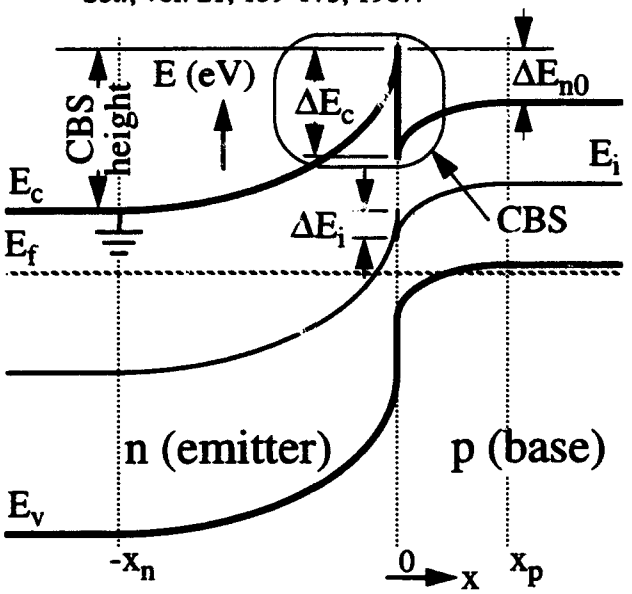


Fig. 1. Equilibrium Band Diagram for the abrupt emitter-base junction of a NPN HBT with a type I band alignment. Note: the reference potential is $E_c(x = -x_n) \equiv 0$.

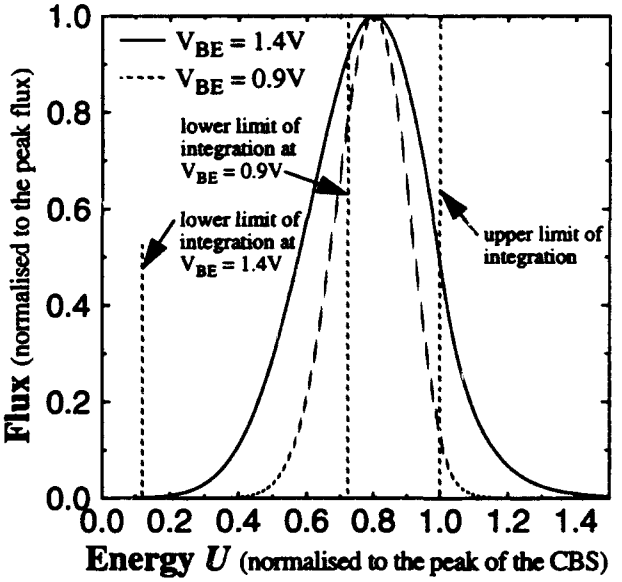


Fig. 2. Emission flux density for an $Al_{.3}Ga_{.7}As/GaAs$ abrupt HBT at two different forward biases. The material parameters are given in Section II.

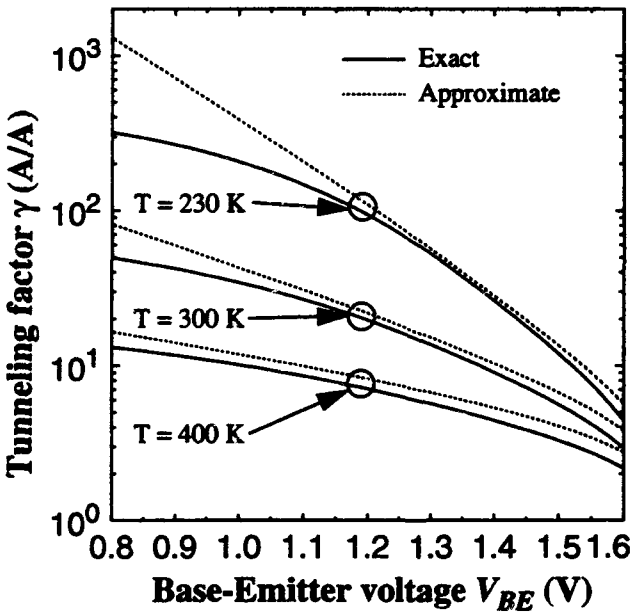


Fig. 3. Plot of the exact (2) and approximate (5) γ versus applied potential for the abrupt HBT detailed in Section II. Note: as the temperature increases γ decreases due to the expected increase in thermionic emission.

DEVICE MODELING USING HYDRODYNAMIC BALANCE EQUATIONS

J.Cai, H.L. Cui,.N.J.M. Horing, X.L. Lei
Department of Physics and Engineering Physics, Stevens Institute of Technology
Hoboken, New Jersey 07030
E. Lenzing, B.S. Perlman
U.S. Army Research Laboratory, Fort Monmouth, New Jersey 07703

ABSTRACT

We describe a semiconductor device modeling program based on hydrodynamic balance equations. This program is is capable of treating multiple carriers in compositionally nonuniform (such as heterostructures) and spatially inhomogeneous device structures, as well as high electric field and associated nonlinear effects. Unlike other balance equation based approaches to device modeling, where the various relaxation rates are treated phenomenologically, or imported from Monte Carlo calculations, our approach is self-contained in that these rates are calculated within the simulation program. The momentum and energy relaxation rates are cast in the form of electric field dependent frictional force and energy transfer functions, with full account of carrier-carrier interaction effects, such as dynamical screening/descreening. These effects are embodied in the dielectric function of the system, which is treated within the random-phase approximation here but can also include exchange-correlation effects. Another advantage of our balance equation approach is that arbitrary energy band structures can be treated, making it suitable for high-field and microwave applications. The simplicity of our technique permit fast and efficient modeling of device performance characteristics, requiring only a fraction of the CPU time needed for Monte Carlo simulations. We have tested the modeling program on simple devices such as an n^+-n-n^+ diode and have obtained good agreement with Monte Carlo simulations.

I. INTRODUCTION

An increasingly popular approach in device modeling is the hydrodynamic balance equation technique, which solves the first few moments of the Boltzmann equation. But the moments equations by themselves do not form a closed set of equations, requiring input of momentum and energy relaxation rates from outside the system. These relaxation rates are supplied from experimental measurements, or from Monte Carlo calculations, and sometimes they are simply taken to be constants. A third way of circumventing this difficulty is to postulate the distribution function with unknown parameters, and use balance equations to solve for these parameters.

Recently, a new balance-equation method for high-field transport in uniform system has been developed[1]. This has subsequently been generalized to weakly nonuniform systems[2]. In addition to the simplicity and transparency of its mathematical structure, the advantages of this method also include its generality of description of nonlinear transport in the presence of an electric field of arbitrary strength, and its ease of treating dynamic, nonlocal carrier-carrier scattering. All the important transport properties are expressed in terms of the carrier density-density correlation function which includes full carrier-carrier interaction within the random-phase approximation (RPA) or beyond.

In this presentation we will describe our balance equation based device modeling, giving particular attention to points of departure from other, more conventional balance equation based device modeling programs.

II. BALANCE EQUATIONS

Within the balance equation approach the following balance equations are used to describe the carrier transport. These are, the equation of continuity,

$$\frac{\partial n}{\partial t} + \vec{\nabla} \cdot (n\vec{v}) = 0, \tag{1}$$

the force/momentum balance equation,

$$\frac{\partial \vec{v}}{\partial t} + \vec{v}(\vec{\nabla} \cdot \vec{v}) = -\frac{2}{3}\frac{\vec{\nabla} u}{mn} + \frac{e}{m}\vec{E} + \frac{1}{mn}\vec{f}, \tag{2}$$

and the energy balance equation

$$\frac{\partial u}{\partial t} = \vec{v} \cdot \vec{\nabla} u = -\frac{5}{3}u(\vec{\nabla} \cdot \vec{v}) - w - \vec{v} \cdot \vec{f}, \tag{3}$$

along with the Poisson equation

$$\nabla^2 \phi = -\frac{e}{\epsilon_0}[n(\vec{R}) - N_D]. \tag{4}$$

These are supplemented by the expression of the average local kinetic energy density of the carrier

$$u(\vec{R}) = 2\sum_{\vec{k}} \epsilon_{\vec{k}} f_0[(\epsilon_{\vec{k}} - \mu(\vec{R}))/k_B T_e(\vec{R})], \tag{5}$$

and that of the local chemical potential $\mu(\vec{R})$ which is related to the local electron density $n(\vec{R})$ via the relation

$$n(\vec{R}) = 2\sum_{\vec{k}} f_0[(\epsilon_{\vec{k}} - \mu(\vec{R}))/k_B T_e(\vec{R})], \tag{6}$$

where $\epsilon_{\vec{k}} = \hbar^2 k^2/2m$ and f_0 is the Fermi-Direc function.

The resistive force and the energy loss rate are

$$\begin{aligned} \vec{f} &= n_I \sum_{\vec{q}} |u(\vec{q})|^2 \vec{q} \Pi_2(\vec{q}, \omega_0) \\ &\quad + 2\sum_{\vec{q},\lambda} |M(\vec{q},\lambda)|^2 \vec{q} \Pi_2(\vec{q}, \omega_0 + \Omega_{\vec{q}\lambda}) \left[N\left(\frac{\hbar\Omega_{\vec{q}\lambda}}{k_B T}\right) - N\left(\frac{\hbar(\Omega_{\vec{q}\lambda} + \omega_0)}{k_B T_e}\right)\right], \end{aligned} \tag{7}$$

$$w = 2\sum_{\vec{q},\lambda} |M(\vec{q},\lambda)|^2 \Omega_{\vec{q}\lambda} \Pi_2(\vec{q}, \omega_0 + \Omega_{\vec{q}\lambda}) \left[N\left(\frac{\hbar\Omega_{\vec{q}\lambda}}{k_B T}\right) - N\left(\frac{\hbar(\Omega_{\vec{q}\lambda} + \omega_0)}{k_B T_e}\right)\right], \tag{8}$$

where $\omega_0 = \vec{q} \cdot \vec{v}(\vec{R})$, $N(x)$ is the Bose-Einstein factor, n_I is impurity density, $\Omega_{\vec{q}\lambda}$ is the phonon frequency of wave vector $\vec{q}$ and branch index λ, $u(\vec{q})$ is the electron-impurity interaction potential, $M(\vec{q},\lambda)$ is the electron-phonon coupling matrix element, $\Pi_2(\vec{q},\omega)$ is the density-density correlation function of electrons which can be obtained within the RPA or beyond. Note that $\vec{f}$ and w depend on the position vector $\vec{R}$ through the quantities $n(\vec{R})$, $T_e(\vec{R})$, and $\vec{v}(\vec{R})$.

These equations will uniquely determine $\vec{v}(\vec{R})$, $T_e(\vec{R})$, $u(\vec{R})$, $n(\vec{R})$, $\mu(\vec{R})$, and $\phi(\vec{R})$, for given initial and boundary conditions. All these variables may be time dependent for transient or ac transport processes. Following standard procedures[3,4], these differential equations are turned into difference equations on a space-time grid. The resulting simultaneous nonlinear difference equations are solved using the Newton method[4].

III. AN EXAMPLE—MODELING OF AN n^+-n-n^+ DIODE

The process of device simulation developed here is applied to the simulation of a one-dimensional problem, an $n^+ - n - n^+$ Si diode. This is a symmetric, 0.55 μm structure, with the middle 0.25 μm doped to $N_D = 10^{15}\, cm^{-3}$, and the anode and cathode (each is 0.15 μm long) doped to $5 \times 10^{17}\, cm^{-3}$. There is some smooth grading in doping density at the junctions between the electrodes and the middle, low doping region.

We carry out our modeling for a lattice temperature $T = 300\, K$. In addition to ionized impurity scattering, we include nonpolar optical phonon scattering and deformation potential acoustic phonon scattering. All materials parameters are those of single crystal Si.

For a bias voltage of $0.5\, V$ we have calculated the steady-state carrier density, drift velocity, energy, and electrostatic potential, all as functions of position along the device length. These are depicted in Figure 1. In addition, we also present the calculated resistive force and energy-loss rate in Figure 2.

In summary, we have developed a semiconductor device modeling program, based on hydrodynamic balance equation approach to charge transport. Instead of the usual relaxation rates employed in traditional balance equation based modeling, our method relies on a resistive force function and an energy-loss rate function, which are calculated within the program.

REFERENCES

1. X.L. Lei and C.S. Ting, Phys. Rev. **B32**, 1112 (1985).
2. X.L. Lei, J. Cai, and L.M. Xie, Phys. Rev. **B38**, 1529 (1988).
3. C.M. Snowden, Ed, *Semiconductor Device Modeling*, (Springer-Verlag, Berlin, 1989).
4. K. Tomizawa, *Numerical Simulation of Submicron Semiconductor Devices*, (Artech House, Boston, 1993).

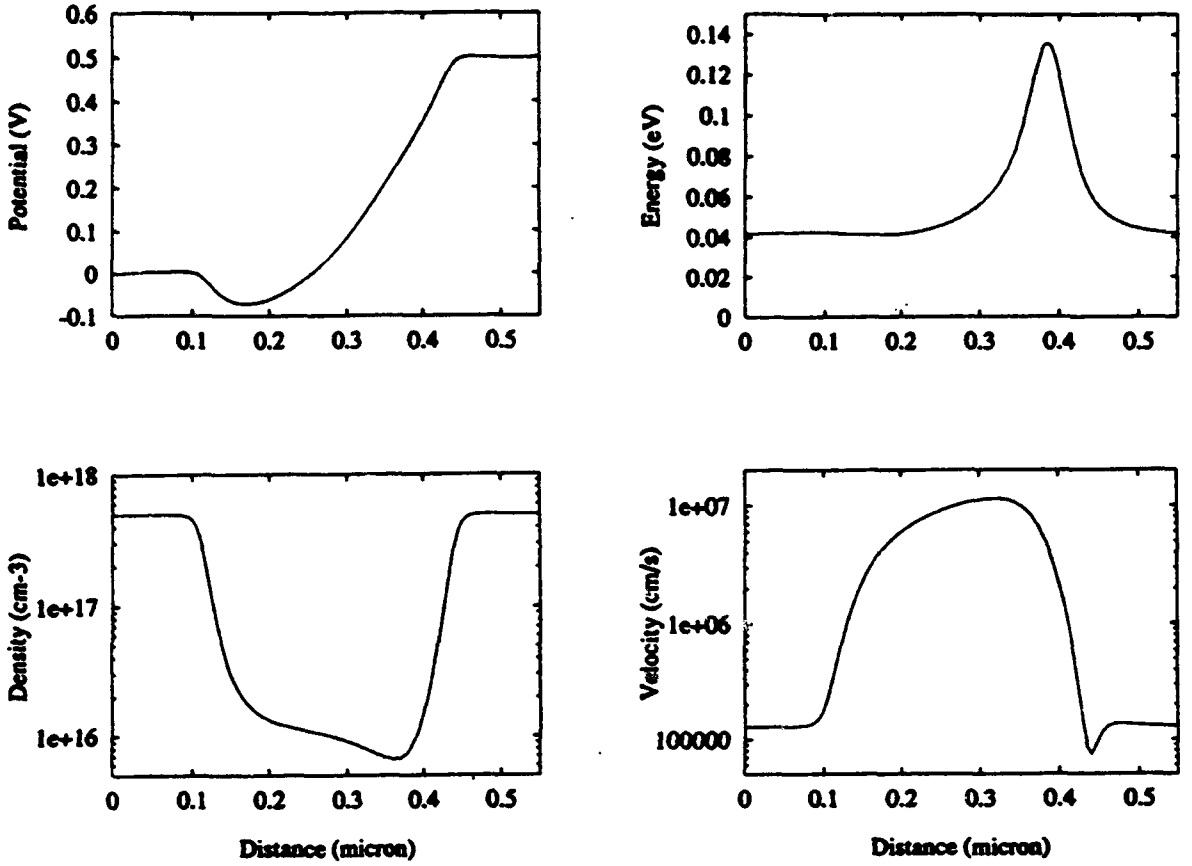


Figure 1: Electron density, drift velocity, energy, and electrostatic potential as functions of position for a bias of 0.5 V.

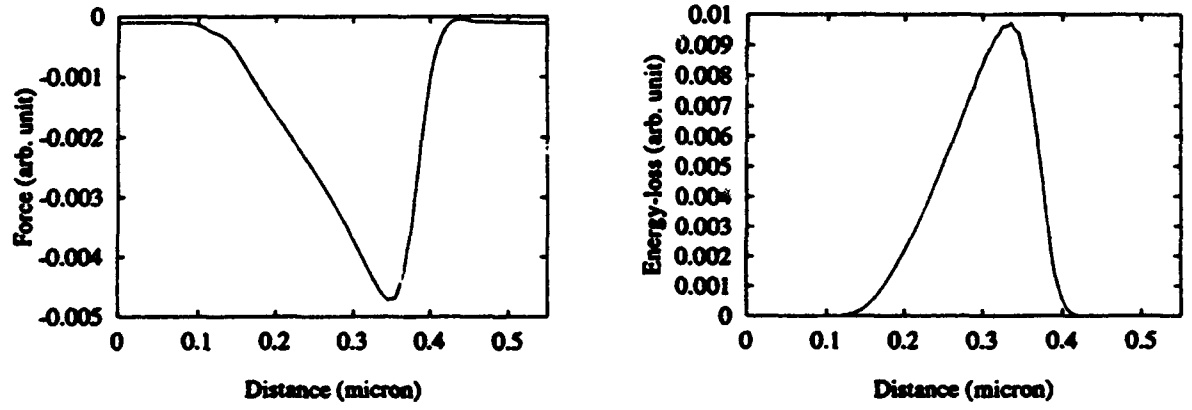


Figure 2: Resistive force and energy-loss rate as functions of position for a bias of 0.5 V.

Effect of Heavy Doping on Conduction Band Density of States: Implications for Device Modelling

C.G. Morton†, J. Wood‡, W. Batty‡, and C.M. Snowden†,

†*Department of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, United Kingdom.*
‡*Department of Electronic Engineering, University of York, York, YO1 5DD, United Kingdom.*

15th May, 1994

Abstract

In this paper, the doping induced distortion to the conduction band density of states is calculated by considering the many-body interactions of the electron-impurity system, following the work of Schwabe *et al*, [3]. The results demonstrate that, at a high impurity density, the majority of states within the band tail are extended in real space and therefore contribute to current flow. Thus, a hydrogenic donor description is inappropriate for devices such as HBTs and HEMTs. For doping densities typically occurring in MESFETs, however, there is still a significant portion of donor states which are localised. The assumption of complete ionisation of donors may, therefore, lead to errors when modelling MESFETs.

I Introduction

In recent years, the inclusion of heavily doped layers within heterostructure based devices has become more widespread. It is common for pseudomorphic HEMTs to incorporate a very narrow and heavily doped donor layer, often described as δ or planar doping, in order to supply electrons to the conducting channel. Typically a few monolayers are grown epitaxially with a sheet donor density of $5 \times 10^{12}\ cm^{-2}$ which translates to a density of $1 \times 10^{20}\ cm^{-3}$. It is common to assume complete ionization of donors in the case of MESFETs. Some authors have introduced a hydrogenic-like shallow donor level when considering parasitic charge build up in HEMTs. In this work, the distortion to the density of states (DOS) is accounted for via a multiple scattering approach described in [1] in order to consider the density of localised states for various donor concentrations.

II Theory

The interacting electron-impurity system can be described using the one-electron Green's function,

$$G(\mathbf{k}, E) = \frac{1}{E - \epsilon(\mathbf{k}) - \Sigma(\mathbf{k}, E)} \tag{1}$$

where E and $\mathbf{k}$ are the electron energy and wave-vector respectively. The function, $\epsilon(\mathbf{k})$, defines the eigenvalues of the unperturbed system,

$$\epsilon(\mathbf{k}) = \hbar^2 k^2 / 2m^* \tag{2}$$

Many-particle interactions are accounted for via the self-energy, Σ,

$$\Sigma(\mathbf{k}, E) = \Sigma_{xc}(\mathbf{k}, E) + \Sigma_{e-i}(\mathbf{k}, E) \tag{3}$$

where Σ_{ec} and Σ_{e-i} refer to electron-electron and electron-impurity interactions respectively.

The DOS is then derived from the spectral density,

$$A(\mathbf{k}, E) = \mp \frac{1}{\pi} \Im \{G(\mathbf{k}, E \pm i0)\} \tag{4}$$

$$\rho(E) = \frac{1}{\Omega} \sum_{\mathbf{k},\sigma} A(\mathbf{k}, E) \tag{5}$$

The self-energy is calculated using Klauder's 5th multiple scattering approximation, [2],

$$\Sigma_{e-i}(\mathbf{k}, E) \equiv U(\mathbf{k}, \mathbf{k}, E) - N_D V(0) \tag{6}$$

$$U(\mathbf{k}, \mathbf{q}, E) = N_D V(\mathbf{k} - \mathbf{q}) + \frac{1}{(2\pi)^3} \int d^3q' V(\mathbf{q}' - \mathbf{q}) G(\mathbf{q}', E) U(\mathbf{k}, \mathbf{q}', E) \tag{7}$$

Schwabe *et al* [3] point out that the exchange-correlation contribution to the self-energy results in a rigid shift to the band edge and so has been omitted from the calculation.

The impurity potential, which includes screening via the inverse screening length, κ, is given by,

$$V(\mathbf{q}) = -\frac{4\pi e^2}{\epsilon_s} \frac{1}{q^2 + \kappa^2} \tag{8}$$

The Separable Potential Approximation transforms the integral equation into a secular equation, the roots of which define the DOS at a particular E.

$$V(\mathbf{q} - \mathbf{q}') = ZC(d) V^{1/2}(q) V^{1/2}(q') \tag{9}$$

$$\Sigma(\mathbf{k}, E) = -\frac{4\pi}{k^2 + \kappa^2} N_D S(E) \tag{10}$$

$$\left[\frac{1}{2}(\kappa^2 - E) - \left\{\frac{(\kappa^2 + E)^2}{4} + 4\pi N_D S(E)\right\}^{1/2}\right]^{1/2} + \tag{11}$$

$$\left[\frac{1}{2}(\kappa^2 - E) + \left\{\frac{(\kappa^2 + E)^2}{4} + 4\pi N_D S(E)\right\}^{1/2}\right]^{1/2} +$$

$$Z - \frac{1}{S(E)} = 0$$

$$\rho(E) = \frac{1}{4\pi^2} \Im \left\{\frac{1}{S(E)} - Z - (\kappa^2 + E) \frac{S(E)}{1 - ZS(E)}\right\} \tag{12}$$

III Results and Discussion

Figure 1 shows the density of states for various donor concentrations. At low donor densities an impurity band is formed which slowly merges with the conduction band density of states, usually described by $CE^{1/2}$, until a band tail is formed. It is clear from the figure that the transfer from impurity band to band tail occurs at a low doping density of around 5×10^{15} donors cm^{-3}.

In order to investigate the spatial behaviour of the electron wavefunction, the spectral density has been considered in figure 2, defining the probability that an electron with energy E is in state k. The doping density has been used at which the transition from impurity band to band tail takes place. The transition from localised to free states can clearly be seen where the spectral density changes from a function distributed in k-space with a maximum at $k = 0$ to a function which exhibits a well defined maximum at some finite value of k. This follows from the uncertainty principle with a small value of Δk implying a state that is extended in real space. The degree of localisation can be estimated using $R = A(k = 0)/A_{max}$, [1].

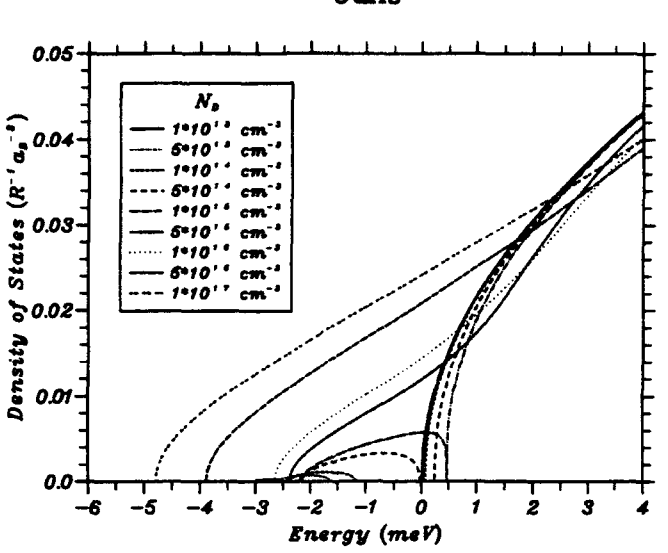


Figure 1: Effect of donor concentration on conduction band density of states.

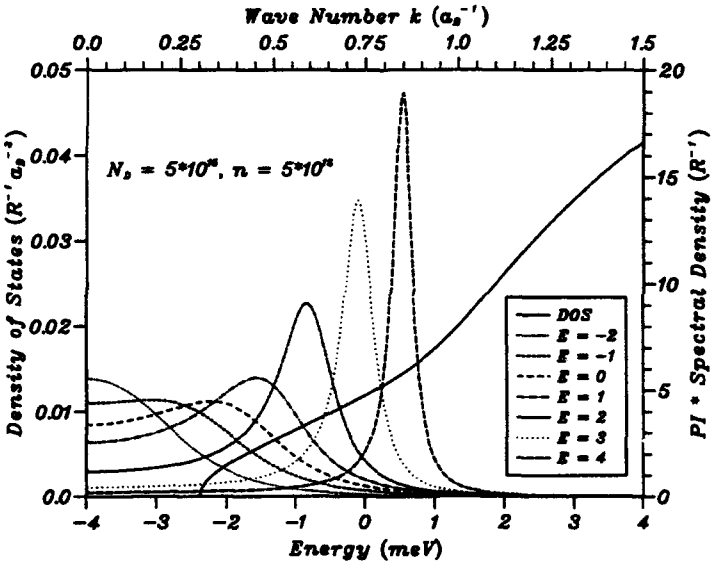


Figure 2: Spectral density distributions from localised to free states.

Figure 3 shows the ratio of localised state density to donor density as a function of impurity density. The localised state density is calculated by integrating the density of states function from the band tail edge to the point at which the transfer from localised to free states takes place. Serre and Ghazali define this transition at $R = 10\%$. As the impurity density is reduced in the low concentration regime, the localised state density approaches the donor density as predicted by the hydrogenic donor description. At higher doping densities, however, it is clear that the hydrogenic description is inappropriate with there being less than 10% of donor states which are localised (N_D greater than $1 \times 10^{18}\ cm^{-3}$). Interestingly, at doping levels typically occurring in MESFETs (10^{17} donors cm^{-3}) there is still between 10 % and 30 % of donor states which are localised. At high doping densities, greater than 1×10^{19}, the majority of states are extended in real space and therefore contribute to a conduction band current.

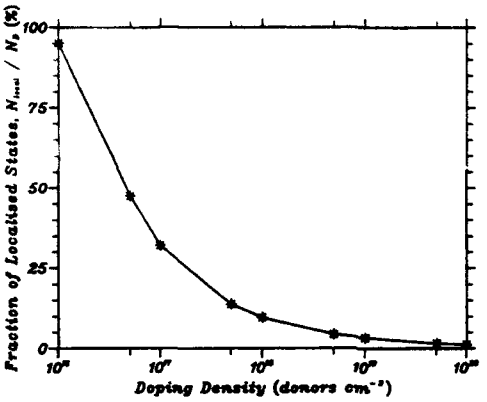


Figure 3: Fracton of localised donor states as a function of impurity concentration.

IV Conclusion

The analysis has shown that the hydrogenic donor model is invalid for doping concentrations typically occuring in HBTs, MESFETs, and HEMTs. Furthermore, the assumption that all donors are ionised appears to be incorrect for doping densities typically occurring in MESFETs. This approximation does, however, seem appropriate for HBTs and HEMTs where the doping density is typically greater than $5 \times 10^{18}\ cm^{-3}$.

Acknowledgement

C.G. Morton would like to acknowledge the Science and Engineering Research Council for providing a studentship.

References

[1] J. Serre and A. Ghazali. From band tailing to impurity-band formation and discussion of localization in doped semiconductors: A multiple-scattering approach. *Phys. Rev. B*, 28(8):4704–4715, 1983.

[2] J.R. Klauder. The modification of electron energy levels by impurity atoms. *Annals of Physics*, 14:43–76, 1961.

[3] R. Schwabe, A. Haufe, V. Gottschalch, and K. Unger. Photoluminescence of heavily doped n-InP. *Solid State Communications*, 58(7):485–491, 1986.

HYDRODYNAMIC MODEL AND JUMP CONDITIONS FOR DISCONTINUOUS MASS AND ELECTRIC POTENTIAL IN HETEROSTRUCTURE DEVICES

Emad Fatemi [1]
Departement de Mathematiques
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne Switzerland

Abstract

A heterostructure is composed of different materials where in each section effective mass of an electron and the minimum energy of the conduction band are different. To model dynamics of electrons in a heterostructure we derive the hydrodynamic equations for a gas of electrons with position dependent mass moving in a potential field. First we write the Boltzmann equation and then we derive the fluid equations by taking the first three moments of the Boltzmann equation and deriving a generalized hydrodynamic model. The jump conditions at the junction are obtained using two different methods. First method relies on using the conservative form of the equations. The second method is based on solving the problem of motion of one electron as it passes through the junction and deriving the necessary conditions on the density functions on two sides of the junction; By taking moments of the density functions we obtain the jump conditions. Jump conditions are obtained as continuity of flux for number of particles, energy, tangential momentum, and square of the tangential momentum.

I. INTRODUCTION

We are concerned with derivation of the hydrodynamic model for devices with a heterojunction. By the hydrodynamic model we shall mean the first three moments of the Boltzmann equation coupled to the Poisson equation [2]. We derive the hydrodynamic model for a gas of electrons with a differentiable position dependent mass and differentiable electric potential. At the interface between the two material there is a discontinuity in the effective mass of electrons and in the minimum energy of the conduction band. We study the problem of the jump conditions at the discontinuity.

A set of equations similar to the hydrodynamic equations were used by D. Widiger, K. Hess, and J. Coleman in [13] and [14]. The equations were derived by considering the first four moments of the Boltzmann equation and simplifying them in order to obtain a closed system for evolution of density and energy. They model a high electron mobility device (HEMT) by writing two sets of equations, a two dimensional system for the electrons in the bulk and a one dimensional system for the electrons in the channel and then coupling the two through flux relations.

A hydrodynamic model including the effects of the position dependent mass was derived by E. M. Azoff in [1]. Our derived equations are similar to his equations but he does not consider the discontinuity in the mass or the potential energy. T. Shawki, G. Salmer, and O. El-Sayed have done numerical simulations of 2D devices using finite difference methods for a set of equations similar to the ones derived by E. M. Azoff [11]. In their work they replace the discontinuities in the conduction band by a smoothed out profile and use an energy dependent relation for effective mass. A different approach was taken by T. Wang and C-H. Hsieh to model the heterostructure devices [12]. They solve a one dimensional hydrodynamic model in the channel coupled to a Schroedinger equation in the perpendicular direction and to a 2D Poisson equation. The problem of interface conditions for the hydrodynamic model has been considered by Schroeder, [9] and [10]. He assumes a perturbed Maxwellian for the density function and from that he derives the interface conditions by taking the moments.

In this paper we present the derived hydrodynamic model for a position dependent mass and the jump conditions at the interface between the two materials. We investigated the problem of electron transfer at the junction where there is a discontinuity in the effective mass and the potential energy. The jump conditions at the discontinuity can be derived using two different methods. One

[1]This research was supported by the Swiss National Science Foundation under grant no. 21-33862.92.

can derive the general conservation laws and write them in conservative form. The jump conditions are readily obtained using mathematical arguments. One can also solve the problem of one single electron passing through the junction and deriving the conditions on the density functions and then deriving the jump conditions by taking moments of the density functions on both sides. The two methods yield the same results.

II. QUANTUM MODEL AND BOLTZMANN EQUATION

An electron in a heterostructure consisting of two different semiconductors can be modeled by an electron in an electric potential $Q(\mathbf{x})$ such that

$$Q(\mathbf{x}) = Q^+(\mathbf{x}) \text{ for } x_1 \geq 0 \text{ and } Q(\mathbf{x}) = Q^-(\mathbf{x}) \text{ for } x_1 \leq 0. \tag{1}$$

Here $\mathbf{x} = (x_1, x_2, x_3)$ is the spatial coordinate variable and functions Q^+ and Q^- are periodic, possibly with different periods, corresponding to the lattice potential in each semiconductor. The Schroedinger equation for an electron in a heterostructure with an applied external potential ϕ is written as

$$-\frac{\hbar^2}{2m}\Delta\Psi(\mathbf{x}) + Q(\mathbf{x})\Psi(\mathbf{x}) + \phi(\mathbf{x})\Psi(\mathbf{x}) = \lambda\Psi(\mathbf{x}). \tag{2}$$

Although the problem is easy to state, analytic or computational solutions of the problem are quite complicated. We have obtained some partial results on the above problem [4].

Our results in this paper are based on modeling the electrons in each semi-space as a semi-classical particle. We assume that in each semi-space each electron has an effective mass corresponding to the underlying lattice potential and is moving in a potential field equal to the energy of the bottom of the conduction band plus the applied electric potential. The laws of motion for such a particle and the BTE equation for an ensemble of such particles are easy to derive. We let (x, y, z) be the position variable, (p, q, r) the momentum variable, t time, and $m(x, y, z)$ be the effective mass of an electron moving in a potential field $F(x, y, z) = E_c(x) + \phi(x, y, z)$, such that $E_c(x)$ is the energy level of the conduction band and discontinuous at $x = 0$ and $\phi(x, y, z)$ is the applied potential and differentiable everywhere. The Hamiltonian for motion of one electron is written as

$$H(x, y, z, p, q, r, t) = (p^2 + q^2 + r^2)/(2m) + E_c(x) + \phi(x, y, z). \tag{3}$$

Then the Boltzmann equation for a gas of such electrons is written as

$$\partial_t f + \partial_x(p/mf) + \partial_y(q/mf) + \partial_z(r/mf) +$$

$$\partial_q(-H_x f) + \partial_q(-H_y f) + \partial_r(-H_z f) = C(f, f) \tag{4}$$

where $f(x, y, z, p, q, r, t)$ is the density function,

$$\nabla H = (H_x, H_y, H_z) = -\nabla m(p^2 + q^2 + r^2)/(2m^2) + \nabla E_c + \nabla\phi, \tag{5}$$

and $C(f, f)$ is the collision operator of all the scattering mechanisms present. The fluid equations can be derived from the above Boltzmann equation. This derivation is standard for constant mass, [5], but here we also have calculated the extra terms coming from the dependence of the mass on position.

III. HYDRODYNAMIC MODEL

We present the derived hydrodynamic model of electrons in a heterostructure in this section, please see [4] for the details. We use as dependent variables:

- Density of Electrons, n
- Momentum, $mn\mathbf{u} = (mnu, mnv, mnw)$
- Energy, e_c
- Electric Potential, ϕ.

Temperature, T, is defined using the main variables by

$$\frac{3}{2}nT = e_c - \frac{1}{2}mn(u^2 + v^2 + w^2). \tag{6}$$

The other quantities are defined in terms of the above or are specified.

- charge of an electron, $e > 0$
- conduction band minimum energy, $E_c(x)$
- number of ionized donors, $N_D^+(x, y, z)$
- momentum relaxation time, τ_p
- and energy relaxation time, τ_w

Then the hydrodynamic model is written in the following form:

$$n_t + \nabla \cdot (n\mathbf{u}) = 0 \tag{7}$$

$$(mn\mathbf{u})_t + \nabla \cdot (mn\mathbf{u}^t \cdot \mathbf{u}) + \nabla(nT) - e_c\frac{\nabla m}{m} = en\nabla\phi + n\nabla E_c - \frac{mn\mathbf{u}}{\tau_p} \tag{8}$$

$$e_{ct} + \nabla \cdot (\mathbf{u}e_c + \mathbf{u}n\tilde{T} + (e\phi + E_c)n\mathbf{u}) = (e\phi + E_c)\nabla \cdot (n\mathbf{u}) - \frac{e_c - \frac{3}{2}nT_s}{\tau_w} + \nabla \cdot \mathbf{q} \tag{9}$$

$$\nabla \cdot (\epsilon\nabla\phi) = (N_D^+ - n). \tag{10}$$

The heat conduction vector can be calculated in the first approximation [5],

$$\mathbf{q} = \frac{5\tau nT}{2m}\nabla T + \frac{\tau n}{4m^2}(5m|\mathbf{u}|^2T - 25nT^2)\nabla m. \tag{11}$$

τ is the relaxation time coefficient of the density function and is approximately of order of τ_p. The electric potential satisfies the Poisson equation and ϵ has a discontinuity at the junction but ϕ and the normal component of $\epsilon\nabla\phi$ are continous at the junction. At the interface the following jump conditions are standard

$$[\phi] = 0, \quad [\epsilon\phi_x] = 0. \tag{12}$$

We use the notaion $[\phi] = \phi_2 - \phi_1$ to denote the jump in the variables across the junction. We suggest the following jump conditions for the fluid variables

$$[nu] = 0 \quad [mv] = 0 \quad [mw] = 0 \quad [T] = 0 \quad [m(u^2 + v^2 + w^2) + (3 + \beta(m))T + 2E] = 0. \tag{13}$$

The first jump condition is conservation of number of particles across the jump, the next two are conservation of tangential momentum, and the last two are consequences of conservation of energy, conservation of the square of the tangential momentum, and some assumptions about the shape of the density functions near the junction. The function N_D^+, number of ionized donors, is defined in terms of the electric potential and the Fermi level. β is defined

$$\beta = \frac{3\alpha}{\alpha + 2} \tag{14}$$

and αs for $2/3 < m1/m2 < 3/2$ are defined as

$$\alpha_1 = \frac{m_1/m_2 + 3 - 2m_2/m_1}{m_1/m_2 + m_2/m_1}$$

$$\alpha_2 = \frac{m_2/m_1 + 3 - 2m_1/m_2}{m_2/m_1 + m_1/m_2}. \tag{15}$$

We considered semi-classical dynamics of electrons in a semiconductor where effective mass and minimum energy of the conduction band are functions of position. We have derived the hydrodynamic model for such electrons from the Boltzmann equation. We considered in detail dynamics of electrons as they pass through the junction and the necessary conditions on the density functions at the junction where there is a discontinuity in the mass and electric field. The jump conditions were obtained by taking the moments of the density functions and writing them in terms of the fluid variables invoking the usual closure assumptions for fluid equations. The derived jump conditions for the conservation of mass and energy equations are satisfactory, but the jump conditions for the momentum equation is yet to be shown satisfactory. We have derived the jump conditions here but their numerical implementation and physical implications are subject of a future study.

References

[1] E. M. Azoff, Closed-form method for solving the steady-state generalized energy-momentum conservation equations, COMPEL, Vol. 6 No. 1, pp. 25-35, 1987.

[2] G. Baccarani and M.R. Wordeman, An investigation of steady-state velocity overshoot effects in Si and GaAs devices, *Solid State Electr.* **28**, pp. 407-416, 1985.

[3] E. Fatemi and F. Odeh, Upwind finite difference solution of Boltzmann-Poisson equation applied to electron transport in semiconductor devices, *JCP*, Vol. 108, No. 2, October 1993.

[4] E. Fatemi, Derivation of the hydrodynamic model and jump conditions for discontinuous mass and electric potential in heterostructure devices, to appear.

[5] Huang, Statistical Mechanics, New York, John Wiley, 1987, p. 104.

[6] M. S. Lundstrom and R.J. Schuelke, Numerical analysis of heterostructure semiconductor devices, *IEEE Transaction on electron devices*, Vol ED-30, No. 9, Sept 1983.

[7] E. W. Maby, Thermoionic emission currents in bulk-barrier heterojunctions, *Solid-state Electronincs*, **32**, No. 1, pp. 95-96, 1989.

[8] M. Mosko, I. Novak, and P. Quittner, On the analytical approach to the real space electron transfer in GaAs-AlGaAs heterostructures, *Solid-stae Electronincs*, **31** No. 3/4, pp. 363-366, 1988.

[9] D. Schroeder, The inflow moment method for the description of electron transport at material interfaces, *J. Appl. Phys.*, **72** (3), 1 August 1992.

[10] D. Schroeder, Three-dimensional nonequilibrium interface conditions for electron transport at band edge discontinuities, *IEEE Trans. Comp. Aid. Des.* **9**, 11, Nov. 1990.

[11] T. Shawki, G. Salmer, and O. El-Sayed, MODFET 2-D hydrodynamic energy modeling: optimization of subquarter-micron-gate structures, *IEEE transactions on electron devices*, Vol. 37. No. 1, Jan 1990.

[12] T. Wang and C-H Hsieh, Numerical analysis of nonequilibrium electron transport in AlGaAs/InGaAs/GaAs pseudomorphic MODFET's, *IEEE transactions on electron devices*, Vol. 37. No. 9., Sept. 1990.

[13] D. J. Widiger, I.C. Kizilyalli, K. Hess, J.J. Coleman, Two-dimensional transient simulation of an idealized high electron mobility transisitor, *IEEE transactions on electron devices*, Vol. ED-32 No. 6, June 1985.

[14] D.J. Widiger, K. Hess, and J.J. Coleman, Two-dimensional numerical analysis of the high electron mobility transistor, *IEEE electron device letters*, Vol. EDL-5 No. 7, July 1984.

QUANTUM DEVICE SIMULATION: OPEN BOUNDARY CONDITIONS FOR THE SCHRÖDINGER EQUATION

Craig S. Lent, Manhua Leng, and P. Douglas Tougaw
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

We discuss the problem of solving the Schrödinger equation on a finite domain for the current-carrying "scattering states" that are frequently of interest from a device viewpoint. We compare approaches based on the finite difference approximation with finite element and green function methods.

I. INTRODUCTION

In exploring theoretically the properties of mesoscopic structures it is frequently of interest to be able to calculate current densities and transmission coefficients for the case of purely ballistic transport with no inelastic scattering in the structure. In a typical idealized structure, current carrying leads with constant width are connected to a device region in which scattering from geometrical features of the potential landscape occurs. A straight-forward approach to the problem is to discretized the device region on a real-space mesh and numerically solve the equations which result from discretizing the Schrödinger equation for the wavefunction in that region. The desired current densities and transmission coefficients can then readily be obtained from the wavefunction directly. A difficulty which occurs is to express boundary conditions for the wavefunction at the interface between the leads and the device region. The wavefunction at the input is the sum of the incident and scattered wave, but the reflection (and transmission) coefficients are unknown at the beginning of the calculation. We have used a boundary-condition method, the Quantum Transmitting Boundary Method [1], and employed it in a finite element (FEM) discretization useful in many two-dimensional problems [2]. Here we compare the implementation of these boundary conditions in the finite difference method (FD) and FEM, and also compare with a recently developed Green function approach. We will restrict ourselves to one-dimensional problems for simplicity. The generalizations to higher dimensions have either been treated elsewhere in detail or are straight-forward.

II. BOUNDARY CONDITIONS ON THE WAVE FUNCTION

We examine first the problem of expressing the condition on the boundaries as a constraint on the wavefunction and its derivative. Figure 1 shows the geometry schematically. An arbitrary scattering potential in the device region from $x=0$ to $x=L$ results in scattered and transmitted plane waves. We

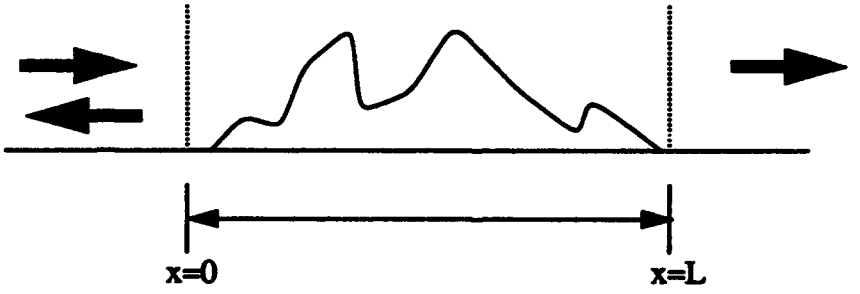


FIGURE 1. Schematic of 1D Schrödinger scattering problem.

(1)

assume a plane wave of unit amplitude is incident from the left. In the lead to the left of the device region the wavefunction has the form

$$\psi(x) = e^{ikx} + re^{-ikx}, \tag{1}$$

while in the right lead the wavefunction is,

$$\psi(x) = te^{ikx}. \tag{2}$$

Evaluating the derivative at the left boundary we have,

$$\psi'(0) = ik(1-r) = 2ik - ik\psi(0) \tag{3}$$

Similarly in the right lead we have

$$\psi(x) = te^{ikx}, \tag{4}$$

so that at the boundary with the right lead we have,

$$\psi'(L) = ik\psi(L). \tag{5}$$

The conditions at the boundaries of the discrete domain can therefore be expressed as follows:

$$\psi'(0) + ik\psi(0) = 2ik \tag{6}$$

$$\psi'(L) - ik\psi(L) = 0 \quad . \tag{7}$$

Equations (6) and (7) are the fundamental boundary conditions for open boundaries in one dimension. The inhomogeneous source term on the right-hand-side of equation (6) results in an incoming plane wave of unit amplitude. The generalization to higher dimensions must include the projection of the normal derivatives onto the various transverse mode of the leads and the existence of evanescent, exponentially decaying, modes.

II. FINITE DIFFERENCE FORMULATION

The boundary conditions in (6) and (7) can readily be expressed in terms of finite difference approximations to the derivatives at the boundaries. If the nodal points are numbered from 0 to N in the device region, the difference equations become

$$(ik - \frac{1}{h})\psi_0 + (\frac{1}{h})\psi_1 = 2ik \qquad (-ik + \frac{1}{h})\psi_N - (\frac{1}{h})\psi_{N-1} = 0 \tag{8}$$

where h is the distance between nodal points. The resulting FD equations can be expressed in terms of an a hopping energy parameter,

$$\Delta = \frac{\hbar^2}{2m^*}(\frac{1}{h^2}). \tag{9}$$

The resulting equations form a linear system with the term $2ik$ from (6) giving the right-hand-side vector. The equations can be written in the form

$$(H - EI)u = p, \tag{10}$$

where u is the column vector of the values of the wavefunctions at the nodal points. If we assume that at the boundaries the potential is zero, this equation becomes,

(11)

$$\begin{bmatrix} (ik-\frac{1}{h}) & (\frac{1}{h}) & & & & & & \\ -\Delta & (2\Delta-E+V_1) & -\Delta & & & & & \\ & -\Delta & (2\Delta-E+V_2) & -\Delta & & & & \\ & & -\Delta & (2\Delta-E+V_3) & \cdots & & & \\ & & & \cdots & \cdots & -\Delta & & \\ & & & & -\Delta & (2\Delta-E+V_{N-1}) & -\Delta & \\ & & & & & -\frac{1}{h} & -ik+\frac{1}{h} & \end{bmatrix} u = \begin{bmatrix} 2ik \\ 0 \\ 0 \\ 0 \\ \cdots \\ 0 \\ 0 \end{bmatrix}$$

The relationship between the wavevector, k, and the energy, E, bears some attention. For free (effective mass) electrons the dispersion is given by

$$E = \frac{\hbar^2 k^2}{2m^*}. \tag{12}$$

The finite difference approximation is equivalent to a tight-binding model and the dispersion is quadratic only in the limit when $h \rightarrow 0$. At any finite mesh size, the numerical dispersion is given by,

$$E_{FD}(k) = 2\Delta\left[1 - \cos(kh)\right] \tag{13}$$

III. FINITE ELEMENT FORMULATION

A finite element formulation begins by developing the weak form of the Schrödinger equation (S.E.). This is done by multiplying the original S.E. by an arbitrary test function $\overline{\psi}$ which obeys the same essential boundary conditions as the wavefunction, and integrating over the interval.

$$-(\frac{\hbar^2}{2m^*})\int_0^L dx\ \overline{\psi}(x)\psi''(x) + \int_0^L dx\ \overline{\psi}(x)V(x)\psi(x) = E\int_0^L dx\ \overline{\psi}(x)\psi(x) \tag{14}$$

Integrating by parts we obtain the weak form of the S.E.:

$$(\frac{\hbar^2}{2m^*})\int_0^L dx\ \overline{\psi}'(x)\psi'(x) + (\frac{\hbar^2}{2m^*})\left[\overline{\psi}(0)\psi'(0) - \overline{\psi}(L)\psi'(L)\right] + \int_0^L dx\ \overline{\psi}(x)V(x)\psi(x)$$

$$= E\int_0^L dx\ \overline{\psi}(x)\psi(x)$$

Notice that the derivative of the wavefunction at the boundary occurs naturally here. The constraint equations (6) and (7) can now be applied. When both the wavefunction and the test function are expanded in the basis of finite element shape functions (here we use linear functions), we obtain the matrix equation

$$(T + C + V - EM)u = p. \tag{15}$$

The matrices T and V represent the kinetic and potential energy operators in this basis. The "mass" matrix M results from the non-orthogonality of the FEM basis functions. The transmitting boundaries are represented by the matrix C. The structure of the FEM matrices is displayed below.

$$T = (h)\begin{bmatrix} \Delta & -\Delta & & & & & \\ -\Delta & 2\Delta & -\Delta & & & & \\ & -\Delta & 2\Delta & -\Delta & & & \\ & & -\Delta & 2\Delta & \dots & & \\ & & & \dots & \dots & -\Delta & \\ & & & & -\Delta & 2\Delta & -\Delta \\ & & & & & -\Delta & \Delta \end{bmatrix} \qquad M = (\frac{h}{6})\begin{bmatrix} 2 & 1 & & & & & \\ 1 & 4 & 1 & & & & \\ & 1 & 4 & 1 & & & \\ & & 1 & 4 & \dots & & \\ & & & \dots & \dots & 1 & \\ & & & & 1 & 4 & 1 \\ & & & & & 1 & 2 \end{bmatrix}$$

$$p = \left(\frac{k^2}{2m^*}\right)\begin{bmatrix} -2ik \\ 0 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix} \qquad C = (k^2)\begin{bmatrix} -ik\Delta & & & & & & \\ & 0 & & & & & \\ & & 0 & & & & \\ & & & 0 & & & \\ & & & & \dots & & \\ & & & & & 0 & \\ & & & & & & -ik\Delta \end{bmatrix} \tag{16}$$

The Green function operator for the S.E. can be expressed in terms of the FEM matrices.

$$G = (EM - T - C(k) - V)^{-1} \tag{17}$$

The wavefunction is then obtained by operating with the green function on the inhomogeneous "source" term, $u = G(-p)$.

The numerical dispersion relation for the FEM mesh is given by,

$$E_{FEM}(k) = 6\Delta\left[\frac{1-\cos(kh)}{2+\cos(kh)}\right] \tag{18}$$

A comparison between the dispersion relations for the continuum, FD, and FEM meshes is given in Figure 2.

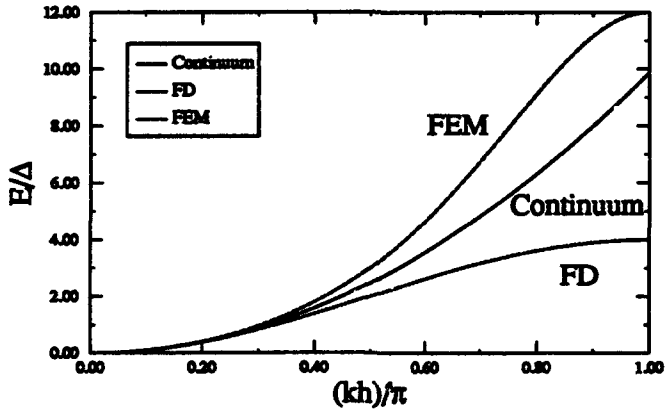


FIGURE 2. Comparison of dispersion relations for the continuum free electron model, the finite difference (FD) model, and the finite element (FEM) model.

IV. GREEN FUNCTION METHOD

An alternative to applying the conditions (6) and (7) on the wavefunction is to apply boundary conditions to the Green function. Let H_0 represent the Hamiltonian for the device isolated from the leads. Let V represent the interaction which couples the 0^{th} and N^{th} nodes in the device to the leads. The leads are assumed to be tight-binding (FD) chains of nodes with the same hopping parameter as in the device, but no potential energy. We can write the Green function as follows:

$$G = \frac{1}{(E - H_0 - V)} = \frac{1}{(G_0^{-1} - V)} \tag{19}$$

The Green function for the uncoupled system can be calculated easily, making use of the known solutions for the Green function of a semi-infinite lead. Because V couples only to nodes *-1* and *N+1* in the lead, we can perform the inversion shown in (19) using only our knowledge of the on-site Green function for the left and right leads:

$$g_L(-1,-1) = g_L(N+1,N+1) = \frac{-e^{ikh}}{\Delta}. \tag{20}$$

The green function for the coupled system can the be calculated by performing the matrix inversion shown below.

$$G = \begin{bmatrix} -\Delta e^{-ikh} & \Delta & & & & & & & \\ \Delta & (E-2\Delta-V_0) & \Delta & & & & & & \\ & \Delta & (E-2\Delta-V_1) & \Delta & & & & & \\ & & \Delta & (E-2\Delta-V_2) & \cdots & & & & \\ & & & \cdots & \cdots & \Delta & & & \\ & & & & \Delta & (E-2\Delta-V_N) & \Delta & & \\ & & & & & \Delta & -\Delta e^{-ikh} & & \end{bmatrix}^{-1} \tag{21}$$

Since only the value of the Green function, G(N,0), is required to find the transmission coefficient, one can solve a simple linear system rather than actually calculating the full inverse of the matrix. The transmission coefficient can then be calculated using the relation

$$T = \left(\frac{1}{h^2}\right)(\hbar v)^2 |G(N,0)|^2. \tag{22}$$

Generalization to higher dimension requires only the block-tridiagonal nature of the Hamiltonian.

V. ERROR ANALYSIS

Figure 3 shows an error analysis for each of the techniques discussed here. The transmission coefficient through a single barrier, for which exact results are available, was calculated for a range of energies from 0 to twice the barrier height. The error was calculated using the expression below.

$$error = \sqrt{\left(\frac{1}{N_E}\right)\sum_i^{N_E}\left(\frac{T(E_i) - T_{exact}(E_i)}{T_{exact}(E_i)}\right)^2} \tag{23}$$

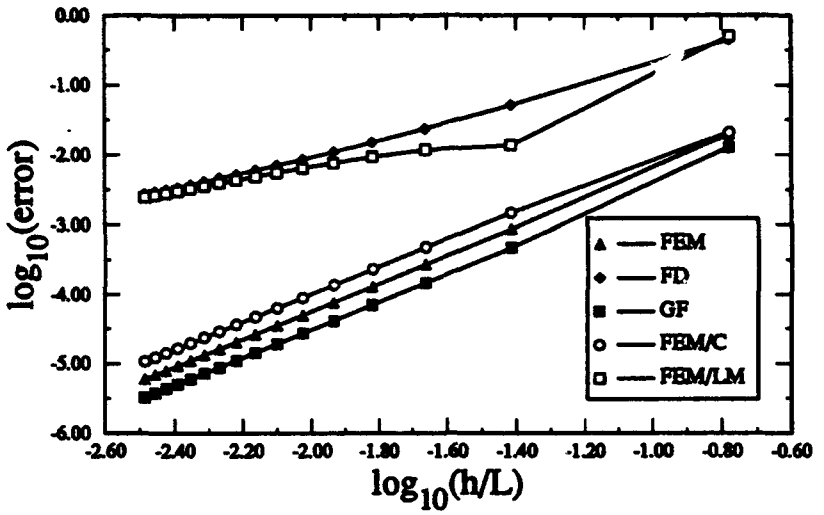


FIGURE 2. Comparison of the error in calculating the transmission coefficient using various numerical approximation. The comparison is made for the case of transmission through a single barrier.

For both FEM, Green function (GF), and FD techniques, the numerical dispersion relations were used. The curve labeled FEM/C employed the continuum dispersion for comparison. The curve labeled FEM/LM used the so-called "lumped mass" approximation for which the FEM matrices V and M are approximated by diagonal matrices. The results show that the GF method slightly improves on the FEM approach. One can see comparing (16) and (21) that the FEM matrix results actually is the second order expansion of the exponential in (20).

The GF and FEM approaches are sufficiently similar in performance that the choice of method will often be dictated by other strengths. The FEM is particularly well suited to complicated boundary shapes in higher dimensions and has been generalized to include applied magnetic fields [3]. The GF method is useful in small problems because analytical results for the matrix inversion can be obtained. Further, it can be used with Keldysh Green function to handle non-equilibrium and dissipative problems [4].

This work was supported in part by the Air Force Office of Scientific Research.

REFERENCES

[1] Craig S. Lent and D.J. Kirkner, *Journal of Applied Physics* **67**, 6353 (1990).
[2] Craig S. Lent and Manhua Leng, *Journal of Applied Physics* **70**, 3157 (1991); Manhua Leng and Craig S. Lent, *Superlattices and Microstructures* **11**, 351 (1992); Manhua Leng and Craig S. Lent, *Physical Review Letters* **71**, 137 (1993).
[3] Manhua Leng and Craig S. Lent, to appear in *Journal of Applied Physics*.
[4] M.J. McLennan, J. Lee and S. Datta, *Physical Review B* **43**, 13846 (1991).

APPLICATIONS OF 3D QUANTUM TRANSPORT SIMULATIONS

D. Z.-Y. Ting, S. K. Kirby, and T. C. McGill
Thomas J. Watson, Sr., Laboratory of Applied Physics
California Institute of Technology
Pasadena, California 91125, U.S.A.

Abstract

Quantum transport in low dimensional nanostructures is examined with an exactly solvable real-space three-dimensional supercell model. Applications to the following examples are discussed : (1) finite length quantum wires, (2) alloy clustering effects in single barrier tunnel structures, and (3) quantum dot arrays.

I. INTRODUCTION

We developed a flexible 3D model aimed at exploring issues relevant to quantum transport in nanostructures, including effects of reduced dimensionality and structural inhomogeneities. Using our method we have demonstrated that interfacial inhomogeneities in double barrier resonant tunneling diodes can induce lateral localization of wave functions [1]; strongly attractive impurities can produce additional transmission resonances [2]; and surface roughness in quantum dots can cause large fluctuations in transmission characteristics [3]. In this paper, we examine transport through single barrier tunnel structures with alloy clustering, and relate it to transport through quantum wires. We also examine transport through quantum dot arrays.

II. METHOD

We use a planar supercell tight-binding Hamiltonian and specify the active region of a structure as a stack of N_z layers perpendicular to the z-direction, with each layer containing a periodic array of rectangular planar supercells of $N_x \times N_y$ sites. Within each planar supercell, the potential assumes lateral variations as dictated by device geometry. Our method obtains exact scattering plane wave solutions [1, 2], subject to supercell periodic boundary conditions in the x- and y-directions, and open boundary conditions in the z-direction. Our method requires accurate and efficient solutions of large sparse linear systems, which is achieved using the quasi-minimal residual method [4].

III. APPLICATIONS

We apply our method to the following examples : (1) finite length quantum wires, (2) quantum dot arrays, and (3) clustering effects in alloy barriers. In all three cases, the band edge and effective mass values for well- and barrier-type materials used are : $E_C^W = 0$ eV, $m_W^* = 0.0673\ m_0$, $E_C^B = 1.05$ eV, $m_B^* = 0.1248\ m_0$; the choice of these material parameters nominally correspond GaAs and AlAs, respectively.

1. Finite Length Quantum Wires

We first examine finite-length quantum wire electron waveguides. We consider GaAs quantum wires surrounded on the sides by AlAs walls, and the ends by GaAs electrodes. The wires have

40Å×40Å cross-section, and wire lengths ranging from 50Å to 800Å. We study the dependence of quantum wire transmission properties on channel length. The transmission spectra in Fig. 1 show that as the quantum wire channel length increases, the number of transmission resonances increases, corresponding to an increasing number of modes in the wire. Note that in all the spectra shown, transmission coefficient tends to be quite small for electron energy below ≈ 0.3 eV. In Fig. 2 we plot the same set of transmission spectra on a semilogarithmic scale to reveal the sub-threshold behavior. We see that there is a cutoff energy (analogues to cutoff frequency in metallic waveguides for electromagnetic waves), and that the cutoff becomes sharper as the channel length increases.

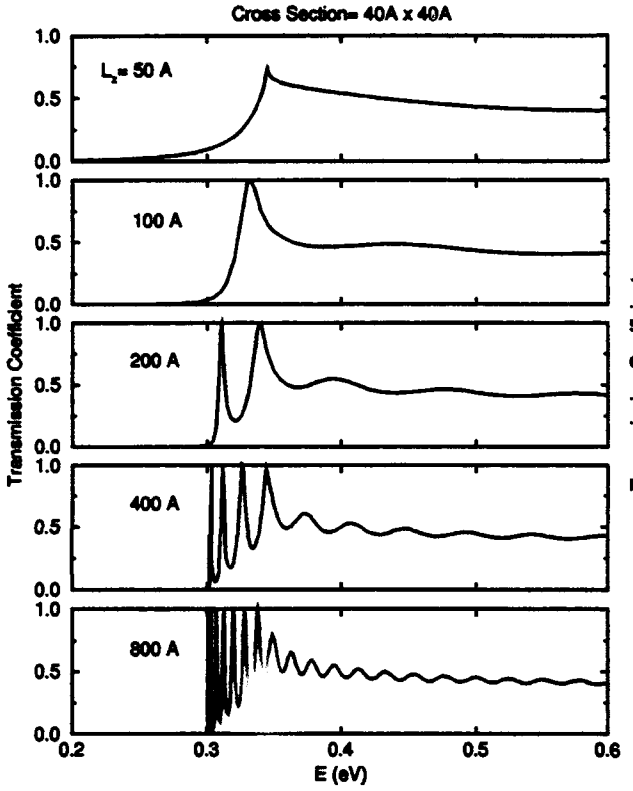


Fig. 1. Transmission coefficients for a set of quantum wire structures with various channel lengths.

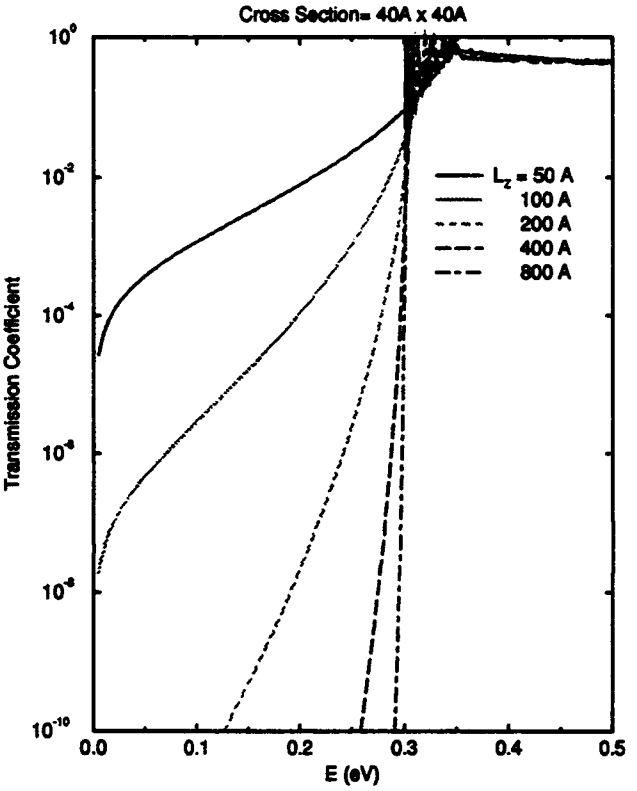


Fig. 2. Transmission coefficients for structures similar to those in Fig. 1, shown in semilog scale.

2. Alloy Clustering Effects in Single Barrier Tunnel Structures

We next consider tunneling characteristics of single barrier GaAs/$Al_{0.5}Ga_{0.5}As$ structures of varying thickness. It can be demonstrated that for totally random alloy configurations, the virtual crystal approximation yields transmission characteristics which are in agreement with supercell calculation results. However, if we allow the AlAs sites (equivalently, the GaAs sites) in the barrier to cluster, then tunneling characteristics can change significantly. Fig. 3 shows the transmission spectra for 50Å, 100Å, and 200Å thick barriers, with cluster size (average in-plane cluster "diameter") of $\lambda =$ 65Å. Note that the spectra show typical single barrier tunneling characteristics below a threshold energy (≈ 0.18 eV). Above the threshold, the even thick barriers becomes somewhat transparent. The threshold energy decreases as cluster size increase, as depicted in Fig. 4. The above-threshold behavior can be explained in terms of short wavelength electrons penetrating through the barrier via channels formed by GaAs clusters. The transport mechanism is analogous to that in finite

length quantum wires; a comparison between Fig. 3 and Figs. 1 & 2 shows qualitative similarities.

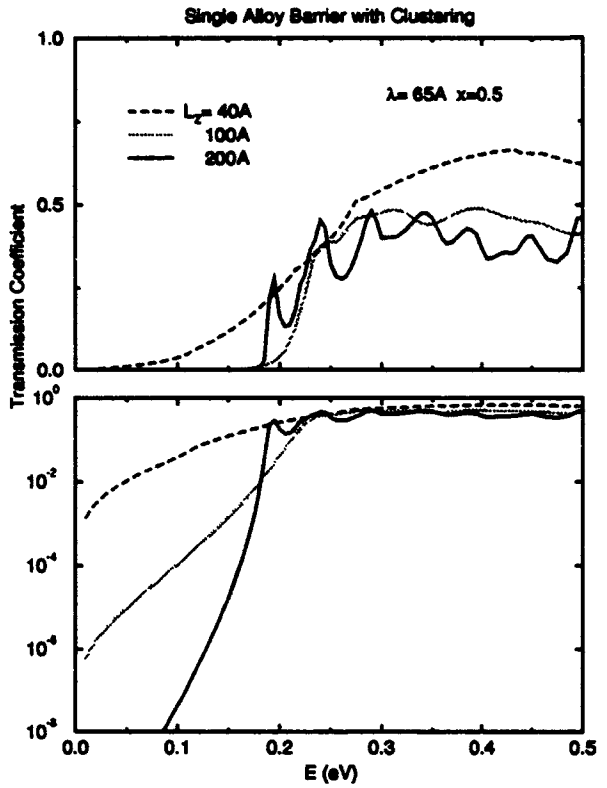


Fig. 3. Transmission coefficients a set of single alloy barrier tunnel structures with different barrier thickness.

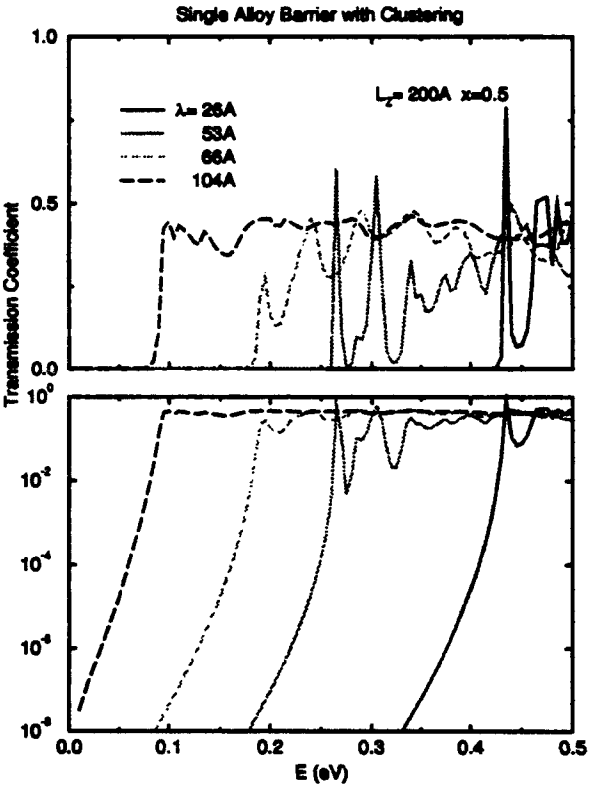


Fig. 4. Transmission coefficients for a set of single alloy barrier tunnel structures with varying degrees of clustering.

3. Quantum Dot Arrays

In the final example we study transmission properties of quantum dot arrays, which consist of $(40\ \text{Å})^3$ dots arranged in a 2D square lattice, embedded between a pair of 20 Å barrier layers. We consider the following three cases as illustrated in Fig. 5 : (1) isolated dots, where the dots are separated laterally by 40 Å barriers, (2) interacting dots, where the interdot barriers are 10 Å wide, and, for comparison, (3) the limiting case of zero interdot separation, which is simply a double barrier structure. Transmission spectra for these structures with various values of lateral incoming plane wave momentum are shown in Fig. 6. While all the spectra show resonances corresponding to the quantized levels in the quantum dots (quantum well), they differ significantly in their $\mathbf{k}_{||}$ dependence. The double barrier structure shows $\mathbf{k}_{||}$ dispersion similar to bulk GaAs, as expected. The array of isolated dots shows no $\mathbf{k}_{||}$ dispersion, due to 0D quantum confinement. The array of interacting dots can be considered as a 2D solid composed of interacting artificial atoms, forming its own band structure differing significantly from that of bulk GaAs. This is quite evident in Fig. 6. We note in particular that the splitting of the $n = 2$ peak due to the interaction of p-like bands.

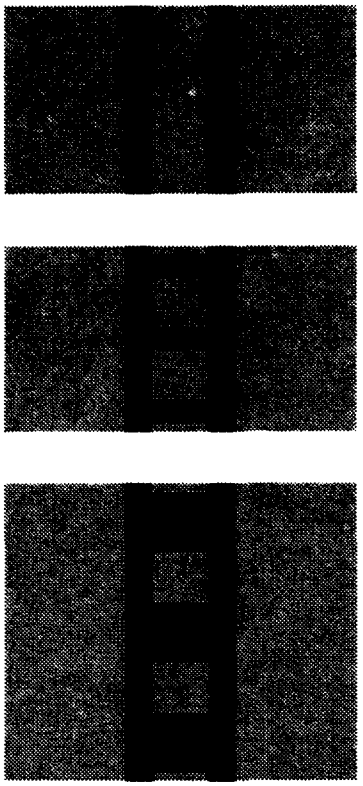


Fig. 5. Illustrations of closely-spaced and isolated quantum dot arrays. A double barrier structure is included for comparison.

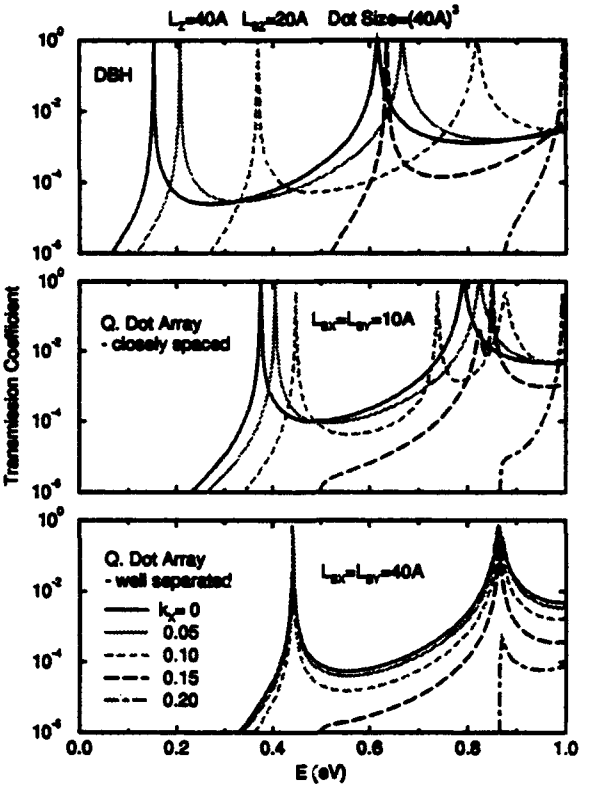


Fig. 6. Transmission coefficients a structures illustrated in Fig. 5.

IV. SUMMARY

We examine transport through single barrier tunnel structures with alloy clustering, and relate it to transport through quantum wires. This demonstrates that structural imperfections can not only produce additional scattering processes in a perturbative sense, but also alter quantized electronic states, leading to substantially modified transport properties. We also studied arrays of mesoscopic devices, where transport properties are strongly influenced by coherence among closely-spaced device structures.

Acknowledgment

The authors would like to thank Eric Van de Velde, R. K. Smith, and W. R. Frensley for helpful discussions. This work was supported by the U. S. Office of Naval Research (ONR) under Grant No. N00014-89-J-1141.

References

[1] D. Z.-Y. Ting, S. K. Kirby, and T. C. McGill, Appl. Phys. Lett. **64**, 2004 (1994).

[2] S. K. Kirby, D. Z.-Y. Ting, and T. C. McGill, Phys. Rev. B **48**, 15237 (1993).

[3] S. K. Kirby, D. Z.-Y. Ting, and T. C. McGill, To be published in Phys. Rev. B.

[4] R. W. Freund and N. M. Nachtigal, Numer. Math. **60**(3), 315 (1991). Technical Report 90.51, RIACS, NASA Ames Research Center, Dec. 1989.

SELF-CONSISTENT, QUASI-3D SIMULATION OF QUANTUM WAVEGUIDE COUPLERS

A. Galick*, M. Macucci†, U. Ravaioli* and T. Kerkhoven‡
* *Beckman Institute, University of Illinois*
405 N. Mathews, Urbana, Illinois 61801
†*Dipartimento di Ingegneria dell'Informazione*
Università degli Studi di Pisa
Via Diotisalvi, 2, I-56126 Pisa, Italy
‡*Department of Computer Science, University of Illinois*
Urbana, Illinois 61801

Abstract

A quasi 3D simulation of a quantum waveguide coupler has been performed. The Schrödinger and Poisson equations have been solved self-consistently in each of the 2D slices into which the device has been subdivided. A modified recursive Green's function algorithm is used to compute the waveguide and tunneling conductances.

I. DEVICE MODEL

We have studied a device model based on the structure reported by Eugster *et al.* in Ref. 1, assuming a layer arrangement (Fig. 1) for the shallow heterostructure as reported in Ref. 2. The gate geometry of our model corresponds faithfully to the real device in the central portion, where coupling between the two waveguides takes place, while differs in the outer regions (dashed lines in Fig. 2), where we have assumed semiinfinite quantum wires instead of tapering off from a 2DEG, for reasons of computational convenience. The model has been divided into 2D slices along the longitudinal direction, assuming that the potential profile is going to be constant within each slice. The Schrödinger and Poisson equations have been solved self-consistently in each slice, obtaining eigenvalues and eigenfunctions to be used for the conductance calculation. Our quasi-3D approach is based on the hypothesis of quasi-adiabatic variation of the potential along the longitudinal direction, which allows solving for the potential separately in each slice. Finally, the Green's functions for the whole structure are computed and from them we obtain the transmission coefficients and, consequently, the conductances, via the Landauer formula.

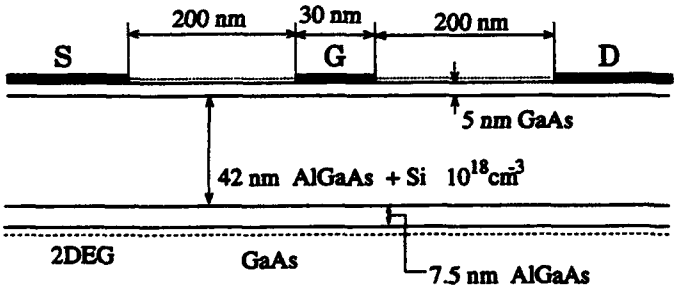


Fig. 1. Layer diagram of the heterostructure

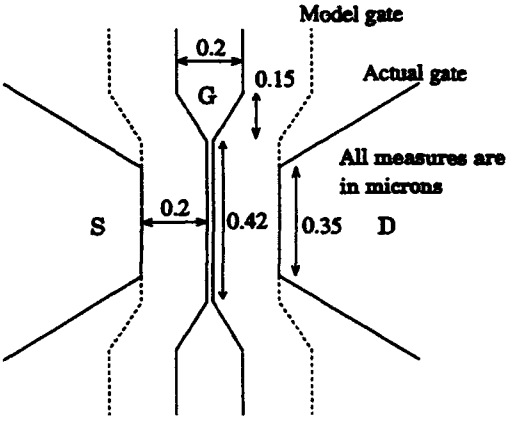


Fig. 2. Gate geometry

II. MODIFIED GREEN'S FUNCTION METHOD

In order to compute the transmission and reflection coefficients and thus the conductances in our model device, we have used a modified version [3] of the Recursive Green's Function Formalism [4,5,6].

The basic idea consists in computing the Green's functions for 1D chains with Dirichlet boundary conditions at their ends. Each 1D chain represents the propagation of a 2D transverse mode within a slice characterized by a constant transverse potential profile. Due to the invariance of the potential along the longitudinal direction, the various transverse modes do not couple within a single slice, therefore their representation with 1D chains is rigorously correct.

The Dirichlet boundary conditions imply isolation of each section from the neighboring ones. The Green's functions of the connected structure are evaluated by applying a perturbation corresponding to joining the ends of the chains belonging to different sections and removing the Dirichlet boundary conditions. The perturbed Green's functions are obtained from the Dyson equation, evaluating the effect of the perturbation to all orders

$$G = G_0 + G_0 \hat{V} G,$$

where G_0 is the Green's function for the unperturbed system, G the one for the perturbed system and $\hat{V}$ represents the perturbation potential, corresponding to coupling between neighboring slices.

The Green's functions for each 1D chain are computed with a discretization based on a tight-binding description of the device geometry. The present approach, however, differs from the one of Ref. 6, because we are not using a square tight-binding lattice. Usage of a square lattice is convenient when studying disorder induced phenomena, i.e. when rapid fluctuations of the potential occur in all directions. In the present calculation we use a tight-binding discretization only along the longitudinal direction parallel to the waveguides, while along the transverse direction we consider simply a number of modes sufficient to accurately describe the coupling between slices. This formulation of the problem yields an immediate advantage: it is possible to use a very fine discretization in the longitudinal direction, while keeping the number of transverse modes and therefore the size of the matrices to be inverted [6] down to reasonable values. A fine discretization in the longitudinal direction is important to obtain a good representation of the continuum energies: the actual tight-binding dispersion relation is cosinusoidal, only in the region around the origin it properly reproduces the parabolic dispersion relation of continuum.

The elements of the coupling matrix $\hat{V}$ are the mode overlaps between the transverse modes of the corresponding pair of slices, multiplied by the tight-binding hopping potential [6]. The mode overlaps are computed taking the discretized overlap integral between the wave functions relative to the modes being considered.

Numerical resolution of transverse modes in each slice requires a fine 2D grid in the quantum well (QW) region. The overlap integrals are computed on a common grid which must be fine over the QW regions of all slices and which therefore has very many lines. We use specialized methods to solve the resulting large, sparse eigenvalue problems.

III. SELF-CONSISTENT SCHRÖDINGER-POISSON SOLUTION

As in Ref. 7, the resolution of the transverse modes in each slice is an iteration to self-consistency of the Schrödinger equation for the wavefunctions and a nonlinear Poisson equation for semi-classical charges such as ionized dopants. We take the surface charge density to be constant between contacts at -3.3×10^{12} cm^{-2}. At 1.6 K, this Poisson equation is highly nonlinear, with effects such as acceptor freezeout in the substrate occurring practically discontinuously, which requires a grid refinement at the freezeout depth. The coupling between the Poisson and Schrödinger equations is also very dramatic, with slight perturbations in the potential resulting in a complete

change in the number and shape of the occupied wavefunctions. For this sensitive problem, we use a fixed-point iteration in the most sensitive quantity, the quantum electron density.

The iteration progresses by successive solution of the nonlinear Poisson and Schrödinger equations. It is clear that a fixed point of this iteration corresponds to a self-consistent solution. It is also clear that the nature of the iteration will be oscillatory, with underfull wavefunctions causing a deepening of the QW, leading to overfull wavefunctions, and vice versa. In the early iterations, we use adaptive underrelaxation to stabilize the oscillations. Close to the solution, we use a Jacobian-free approximate Newton method to accelerate the convergence to self-consistency. Our experience is that this is a very effective way to handle the nonlinearlity in the model [8].

The nonlinear Poisson equation is solved using a Newton method with inexact linesearch. We take zero-field boundary conditions in the air above the contacts and to the sides, Dirichlet boundary conditions in the contacts that include a Schottky barrier of 0.9 eV, and Dirichlet boundary conditions in the substrate for charge-neutrality, determined by a bisection search of the bandgap. The Jacobian is solved for the Newton direction using the Conjugate Gradient (CG) method on a reduced system obtained by block Gaussian elimination of a red-black reordering of the matrix from a 5-point discretization on a rectangular grid, as in [8]. This gives an order of magnitude speedup over straight CG.

The eigenvalue problem for the Schrödinger equation can be effectively solved with a version of RITZIT [9] modified to use column operations, providing the spectrum is first shifted to make the desired eigenvalues the largest in modulus. However, we have developed a more efficient Chebyshev-preconditioned Krylov subspace method. Both of these solvers are projection methods [10], which reduce the complexity of the eigenvalue problem by finding a small invariant subspace of a matrix rather than its entire spectral decomposition. Although the fine grid to resolve the transverse wavefunctions results in a large, sparse eigenvalue problem, only the few lowest energy levels that are occupied are relevant to the problem. The higher energy levels are squashed by Chebyshev preconditioning in both solvers, so that the subspace iteration in RITZIT and our Krylov subspace iteration converge specifically to the desired modes.

IV. RESULTS

We have been interested in simulating 1D-to-1D tunneling in this structure, therefore we have chosen electrode bias values tuned to obtain significant coupling between the wires. It turns out that appreciable coupling is reached only with a central gate bias of -0.6 V, which corresponds to the threshold for depletion of the 2DEG under an infinite gate. In our model device this does not lead to coupling between the outgoing leads, due to the depleting action of the source and drain electrodes that are in close proximity of the gate. In the real device geometry, at this gate voltage the two channels would probably be short-circuited, due to strong coupling far from the central region. This may be the explanation for the problems reported [11] in the observation of 1D-to-1D tunneling.

In Fig. 3 we report the results for the conductance between the ends of the same waveguide (upper curve) and between the end of one waveguide and the other end of the other waveguide (lower curve) vs. the length of the central slice, where most of the coupling takes place. Bias values are constant: V_G=−0.6 V, $V_D = V_S$=−1.65 V. As expected, we observe a substantially oscillatory behavior of the tunneling conductance for increasing length of the central section. Oscillations in the tunneling conductance have opposite phase with respect to the ones in the direct conductance in order for the total current to be constant. Even in this extreme bias condition the coupling never reaches $2e^2/h$.

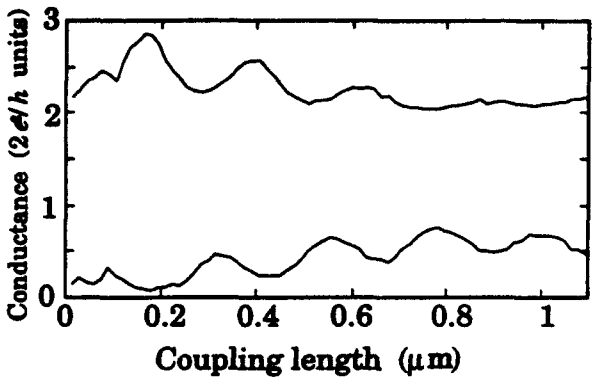


Fig. 3. Conductance between the end of the same waveguide (upper curve) and tunneling conductance (lower curve) vs. coupling length.

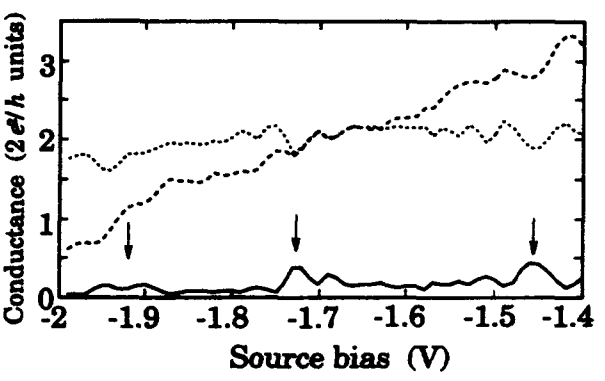


Fig. 4. Conductance of the drain waveguide (dotted line), of the source waveguide (dashed line) and tunneling conductance (solid line) vs. bias.

In Fig. 4 results are shown for the conductance of the source-side waveguide (dashed line), the drain-side waveguide (dotted line) and between the two waveguides (solid line) as a function of the bias of the source electrode. The gate and drain biases are kept constant at -0.6 V and -1.7 V, respectively. We observe peaks (indicated by arrows) of the tunneling conductance in correspondence with the opening of new modes in the source waveguide, in analogy with what has been observed experimentally for the 1D-to-2D tunneling. Conductance quantization for the source waveguide when the source bias is swept is rather poor, as in the experimental results of [1]. This may also be due to reflections in the bends, besides the effects of finite temperature (1.6 K) and of coupling.

ACKNOWLEDGMENTS

This work has been supported through the NSF grant ECS 91–20641. One of the authors (M. M.) acknowledges also support from the CNR (Italian National Research Council).

REFERENCES

[1] C. C. Eugster and J. A. del Alamo, Phys. Rev. Lett. **67**, 3586 (1991).

[2] M. J. Rooks, C. C. Eugster, J. A. del Alamo, G. L. Snider and E. L. Hu, J. Vac. Sci. Technol. B **9**, 2856 (1991).

[3] M. Macucci, U. Ravaioli and T. Kerkhoven, Superlattices and Microstructures **12**, 509 (1992).

[4] D. J. Thouless and S. Kirkpatrick, J. Phys. C **14**, 235 (1981).

[5] F. Guinea and J. A. Vergés, Phys. Rev. B **35**, 979 (1987).

[6] F. Sols, M. Macucci, U. Ravaioli, and Karl Hess, J. Appl. Phys. **66**, 3892 (1989).

[7] S. E. Laux and F. Stern, Appl. Phys. Lett. **49**, 91 (1986).

[8] T. Kerkhoven, A. Galick, U. Ravaioli, J. Arends, and Y. Saad, J. Appl. Phys. **68**, 3461 (1990).

[9] H. Rutishauser, Numer. Math. **16**, 205 (1970).

[10] Y.Saad, in Matrix Pencils, Proceedings, P. Havsbad, B. Kagstrom and A. Ruhe, Eds., 121 (1982).

[11] J. A. del Alamo, private communication

OPEN BOUNDARY CONDITIONS FOR MULTIDIMENSIONAL ELECTRONIC SCATTERING STATES*

Henry K. Harbury† and Wolfgang Porod
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

R. Kent Smith
AT&T Bell Laboratories
Murray Hill, NJ 07974-0636

Abstract

The detailed spatial variation of the electronic scattering states in open and unconfined mesoscopic systems is of interest in both the asymptotic far-field regime and the near-field regime in the vicinity of the scattering potential. We are interested in solving the multidimensional effective mass Schrödinger equation for the scattering states which are compatible with the outward Sommerfeld radiation condition. We present our findings in a comparison of two numerical solution methods which implement non-reflecting boundary conditions on an artificial boundary enclosing the multidimensional problem domain and are compatible with standard finite element techniques.

I. INTRODUCTION

It is of interest to study the near-field scattering states in open-boundary multidimensional ballistic structures. The importance of these scattering states is demonstrated in recent scanning tunneling microscope experiments on metals which support a 2DEG by surface state confinement [1,2]. The interference behavior near point scatterers and step edges is directly related to the two-dimensional local density of states of the electronic scattering system. It has also been pointed out [3] that local field effects in mesoscopic scattering systems are closely related to the electronic scattering states, which have been previously studied for Q1D ballistic structures [4] in the near field regime. To this end, we present a comparison of two numerical methods which implement non-reflecting boundary conditions which are compatible with the outward Sommerfeld radiation condition to solve the single-electron effective-mass Schrödinger equation.

II. SOLUTION METHOD

Two methods for implementing non-reflecting boundary conditions (NRBC) on an artificial boundary are presented. The first is an exact method which uses the known far-field solution for the asymptotic scattering state by matching the known partial-wave expansion to the near-field solution on the artificial domain boundary, and likewise matching the boundary normal derivatives. The Schrödinger equation is integrated by parts with the boundary normal derivatives inserted into the surface term which results in a non-local densely coupled boundary. The second NRBC method is an approximate boundary condition formulated from the application of operators which are, by construction, compatible with the outward Sommerfeld radiation condition in the far-field regime. The boundary condition is again inserted into the surface integral term but results in only a local tridiagonally coupled boundary. The model system is presented followed by a brief synopsis of both the non-local exact NRBC and the local approximate NRBC.

*This work has been supported by the Office of Naval Research and the Air Force Office of Scientific Research.
†H.K.H. is grateful for a fellowship from the Center for Applied Mathematics of the University of Notre Dame.

1. Model System

We seek explicit solutions to the multidimensional single-electron effective-mass Schrödinger equation,

$$-\frac{\hbar^2}{2}\nabla\cdot\left[\frac{1}{m^*}\nabla\psi_E(\mathbf{r})\right]+V(\mathbf{r})\psi_E(\mathbf{r})=E\psi_E(\mathbf{r}). \tag{1}$$

A two dimensional domain will be assumed for simplicity, although the method is easily generalized to three dimensional geometries. As shown schematically in Fig. 1, the scattering domain has an artificial boundary at a radius R_0 from the center of the scattering potential which encloses the solution domain, Ω_0. The potential is assumed constant outside the artificial domain boundary in the region labeled Ω_I, and is taken as the zero reference. As indicated in the figure, an electronic flux, injected by some outside source, is incident upon the mesoscopic scattering region, Ω_0. It will be assumed that the wave-function for the incident flux can be described by a plane wave of the form $\psi_{inc}=a\exp[i\mathbf{k}\cdot\mathbf{r}]$, where $k=|\mathbf{k}|=\sqrt{2m^*E/\hbar^2}$.

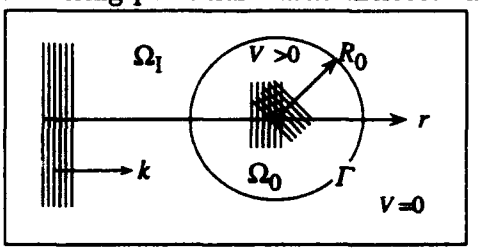


Figure 1. Schematic diagram of the solution domain enclosing the scattering potential.

2. Non-Reflecting Boundary Conditions

In this section a brief synopsis of the exact non-reflecting boundary condition for the solution of Eq. 1 will be presented and the reader is referred to the recent publication [5] by two of the authors for a detailed derivation. The formulation begins with the partial wave expansion of the known far-field solution to Eq. 1,

$$\psi(r\geq R_0,\theta)=\psi_{inc}+\psi_{scatt}=ae^{i\mathbf{k}\cdot\mathbf{r}}+\sum_{m=-\infty}^{\infty}i^m b_m H_m^{(1)}(kr)e^{im\theta}, \tag{2}$$

where a is the amplitude of the incident plane wave, $H_m^{(1)}$ is the Hankel function for the outgoing scattered wave with unknown amplitude b_m. The unknown coefficients, b_m in Eq. 2 may be eliminated through the use of the orthogonality of the angular modes. The known far-field solution is used to obtain the exact normal derivatives of the scattering state on the boundary. The integration by parts of Eq. 1 results in the surface integral term into which this boundary condition is inserted. The resulting formulation using the finite element approximation $\psi\approx\sum_i\psi_i u_i$, where u_i is the i^{th} shape-function, recast into matrix notation has the form of the linear system,

$$[\mathbf{T}+\mathbf{V}+\mathbf{C}]\,\boldsymbol{\psi}=\mathbf{p}, \tag{3}$$

where $\mathbf{T}_{ij}=\frac{\hbar^2}{2}\int_{\Omega_0}\nabla u_i\cdot\frac{1}{m^*}\nabla u_j\,d\Omega$ and $\mathbf{V}_{ij}=\int_{\Omega_0}u_i\,[V-E]\,u_j\,d\Omega$ are the kinetic and potential energy matrix element, respectively, $\mathbf{C}_{ij}=-\frac{\hbar^2}{2}\int_0^{2\pi}u_{i,\Gamma}\frac{1}{m^*}\left[k\sum_{m=-\infty}^{\infty}\frac{H_m^{\prime(1)}(kR_0)}{H_m^{(1)}(kR_0)}\frac{1}{2\pi}\left(\int_0^{2\pi}e^{-im\theta}u_{j,\Gamma}d\theta\right)\times e^{im\theta}\right]R_0d\theta$ is the matrix element which results from the boundary integral, and the right-hand-side vector $\mathbf{p}_i=\frac{\hbar^2}{2}\int_0^{2\pi}u_{i,\Gamma}\frac{1}{m^*}\left[ak\sum_{m=-\infty}^{\infty}i^m\left(J_m'(kR_0)-J_m(kR_0)\frac{H_m^{\prime(1)}(kR_0)}{H_m^{(1)}(kR_0)}\right)e^{im\theta}\right]R_0d\theta$ contains the incident plane wave contribution. The embedded integral around the boundary in $\mathbf{C}$ is inherently non-local and results in a numerical formulation whose discretized boundary is densely coupled.

The approximate non-reflecting boundary condition is based on the method of Bayliss and coworkers [6]. In solving Eq. 1 for the scattering states, a constraint on the artificial boundary is

imposed which restricts the solutions to the class of wave-functions which are compatible with the outward Sommerfeld radiation condition. The boundary constraint is exact in the limit of $R_0 \rightarrow \infty$ and has a radially decaying truncation error on the artificial boundary.

In the two-dimensional case, the scattered part of the wave-function may be expanded in the form $\psi_{scatt} \propto e^{ikr} \sum_{j=0}^{\infty} f_j(\theta)/(kr)^{j+\frac{1}{2}}$. A set of linear operators, B_m, compatible with the outgoing Sommerfeld radiation condition are constructed such that on the artificial boundary each m^{th} higher order term is eliminated when B_m operates on the far-field expansion, $B_m \psi_{scatt}|_\Gamma = \mathcal{O}\left(R_0^{-\frac{4m+1}{2}}\right)$. The set of operators are therefore defined as $B_0 = 1$, $B_m = (\partial_r - ik + \alpha_m/r)B_{m-1}$, where α_m is a constant chosen to eliminate the m^{th} term of the expansion. In the two-dimensional case, the first two non-trivial operators are given by:

$$B_1 = (\partial_r - ik + \tfrac{1}{2r}) \quad ; \quad B_1\psi_{scatt}|_\Gamma = \mathcal{O}(R_0^{-5/2}) \tag{4.a}$$

$$B_2 = (\partial_r - ik + \tfrac{5}{2r})B_1 \quad ; \quad B_2\psi_{scatt}|_\Gamma = \mathcal{O}(R_0^{-9/2}). \tag{4.b}$$

Expanding $B_2\psi_{scatt}|_\Gamma$ and implicitly using the original Schrödinger equation, $\frac{\partial^2\psi}{\partial r^2} = -\frac{1}{r}\frac{\partial\psi}{\partial r} - \frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2} - k^2\psi$, the result of operating with B_2 on ψ may be rearranged to obtain the boundary normal derivative of the wave-function: $\left.\frac{\partial\psi}{\partial r}\right|_\Gamma = \left\{\frac{r}{2(1-ikr)}\left[B_2\psi_{inc} + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2} - \frac{3\psi}{4r^2} + \frac{3ik}{r}\psi + 2k^2\psi\right]\right\}_\Gamma$. Similar to the derivation of the exact non-local NRBC, this boundary normal derivative may be used in the surface term resulting from the integration by parts of Eq. 1. Within the finite element method, the linear system given by Eq. 3 is again obtained, where now the boundary terms are given by $\mathbf{C}_{ij} = \frac{\hbar^2}{2m^*}\frac{4}{3}\eta\int_0^{2\pi}\left.\frac{\partial u_i}{\partial\theta}\right|_\Gamma\left.\frac{\partial u_j}{\partial\theta}\right|_\Gamma d\theta + \frac{\hbar^2}{2m^*}(\eta - \xi - 2\gamma)\int_0^{2\pi} u_i|_\Gamma u_j|_\Gamma d\theta$ and $\mathbf{p}_i = \frac{\hbar^2}{2m^*}\left\{-\gamma\int_0^{2\pi} u_i(cos\theta - 1)^2\psi_{inc}d\theta + \xi\int_0^{2\pi} u_i(cos\theta - 1)\psi_{inc}d\theta + \eta\int_0^{2\pi} u_i\psi_{inc}d\theta\right\}$, where ψ_{inc} is the incident wave, $\gamma = \frac{(kR_0)^2}{2(1-ikR_0)}$, $\xi = \frac{3ikR_0}{2(1-ikR_0)}$, and $\eta = \frac{3}{4}\frac{1}{2(1-ikR_0)}$. Unlike the exact NRBC method, this formulation does not contain any embedded integral terms in $\mathbf{C}$ and results in a numerical formulation whose discretized boundary is only tridiagonally coupled. This local NRBC, however, has an approximation error due to the neglected terms of the far-field expansion. This error may be made less than the inherent discretization error in the domain by using a sufficiently large radius.

III. PERFORMANCE

The approximation error of the local NRBC method is expected to improve with increasing radius and approach the exact solution compatible with the outward sommerfeld radiation condition. To compare the two algorithms, it is necessary to fix the discretization error of both the interior domain and the artificial boundary. To this end, a series of regular concentric circular meshes of increasing radius were generated such that the discretization error was held constant by keeping both kh_r and kh_θ equal between grids. The number of nodes was therefore increased with increasing domain radius while keeping the wavevector, k, constant. For the non-local NRBC computations, the infinite sums in the boundary condition were truncated to 40 terms and the linear system, Eq. 3, was solved to obtain the "exact" scattering states. Likewise, the local NRBC was implemented on the same set of problem domains, using the B_2 operator, and the system was solved for the "approximate" solutions. The reflection coefficients, b_m, were obtained from the "exact" solutions and Eq. 2 by integrating around the boundary using the orthogonality of the angular modes. The approximation error of the local NRBC method is determined from $B_m w = B_m(\psi_{scatt} - v)$, where v is the approximate solution computed by enforcing $B_m v = 0$ on the boundary and ψ_{scatt} is the

"exact" formulation of the scattered wave in Eq. 2 using the numerically computed b_m coefficients. Plotted in Fig. 2 is the norm of the computed error of the local NRBC, normalized for the number of nodes on the boundary, for both the B_1 (circles) and B_2 (triangles) operators as a function of increasing domain radius. Each point was obtained from solutions computed using different grids with increasing domain radius. Also shown are the computed fits for the next four higher order terms from Eq. 4.a, $y = c_1(kR_o)^{-5/2} + c_2(kR_o)^{-7/2} + \ldots$, and Eq. 4.b, $y = d_1(kR_o)^{-9/2} + d_2(kR_o)^{-11/2} + \ldots$, which are neglected by the B_1 and B_2 local NRBC approximations, respectively. The local approximation error also improves with increasing wavevector, k, [6] so one may expect better results with higher energy. These results demonstrate that the two NRBC methods approach the same solution at larger radii, neglecting any difference in the discretization error on the boundary. For the non-local method, the accuracy of the embedded boundary integral is sensitive to the number of boundary nodes. For solutions far from the scattering potential, where a large number of nodes must be placed on the boundary, the efficiency of the local NRBC makes it superior to the densely coupled non-local "exact" method. If a given 2-D mesh has N nodes and an effective bandwidth $\beta = \mathcal{O}(\sqrt{N})$, then for the non-local NRBC method the storage requirement is $\mathcal{O}(N\beta)$ non-zero matrix entries as opposed to the tridiagonally coupled local NRBC $\mathcal{O}(N)$ requirement. The difference in storage will also be reflected in the execution time. For the non-local NRBC, the execution time is $\mathcal{O}(N\beta^2)$ for a direct solve and approximately $\mathcal{O}(N\beta)$ for an iterative solution. The local NRBC also has an execution time of $\mathcal{O}(N^2)$ for the direct solve, but the mulplicative constant is much lower. The time for an iterative solve is approximately $\mathcal{O}(N \log N)$, which is feasible for three-dimensional problems.

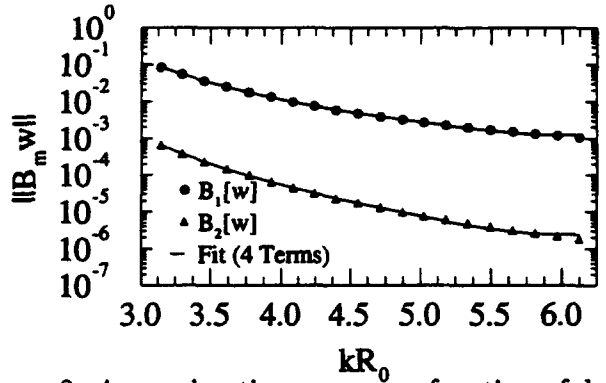


Figure 2. Approximation error as a function of domain radius, R_0, for $B_1 w$ (circles) and $B_2 w$ (triangles). Also shown are the fits for the next few expansion terms.

IV. SUMMARY

Two methods for implementing non-reflecting boundary conditions (NRBC) which are compatible with the outward Sommerfeld radiation conditions were compared for the solution of the single electron effective-mass Schrödinger equation. The exact NRBC results in a non-locally coupled boundary condition whereas the approximate NRBC method is local. The approximation error of the local method, however, requires that the artificial boundary be placed sufficiently far from the scattering center that the neglected higher order terms in the approximation have decayed. The non-local NRBC method has greater accuracy for an artificial boundary placed closer to the scattering potential. The efficiency in both storage and execution time of the local NRBC over the non-local NRBC suggests that, for realistic 3-D problems, only the local NRBC method is feasible.

[1] M. F. Crommie, C. P. Lutz, and D. M. Eigler, *Nature* **363**, 524 (1993).
[2] Y. Hasegawa and Ph. Avouris, *Phys. Rev. Lett.* **71**, 1071 (1993).
[3] R. Landauer, *Z. Phys. B - Condensed Matter* **68**, 217 (1987).
[4] H. K. Harbury, W. Porod, and C. S. Lent, *Superlat. Microstruc.* **11**, 189 (1992).
[5] H. K. Harbury and W. Porod, to appear in *J. Appl. Phys.*, May 15 (1994).
[6] A. Bayliss, M. Gunzburger, and E. Turkel, *SIAM J. Appl. Math.* **42**, 430 (1982).

QUANTUM DEVICE SIMULATION INCLUDING INTERACTIONS

S.Datta, G.Klimeck*, R.Lake**, Y.Lee+ and M.J.McLennan++
School of Electrical Engineering, Purdue University,
West Lafayette, IN 47907-1285.
*Present Address : *University of Texas at Dallas, **Texas Instruments,*
+Hiroshima University, ++AT&T Bell Laboratories.

Abstract

Electron-phonon and electron-electron interactions are commonly included in semiclassical device simulation programs. However, such interactions are often neglected in the simulation of quantum devices. In this talk we will describe a general approach to quantum device simulation based on the non-equilibrium Green's function (NEGF) formalism that allows us to include these interactions.

I. INTRODUCTION

The cornerstone of semiclassical transport theory is the Boltzmann equation

$$v.\bar{\nabla}f + (e\bar{E}/\hbar).\bar{\nabla}_k f + S^{out}f = S^{in}(1-f) \qquad (I.1)$$

which describes the behavior of the distribution function f(r,k). This description is based on a physical picture which views electrons as particles that move in the external electric field according to "Newton's laws" (the quotes are used as a reminder that bandstructure effects are included) and are scattered by the random microscopic fields arising from impurities, phonons or other electrons (described by the functions $S^{in}(r,k)$ and $S^{out}(r,k)$). The Boltzmann equation effectively combines semiclassical dynamics with a stochastic description of the scattering processes:

Boltzmann $\Rightarrow$ "Newton's laws" + Random scattering

This approach works quite well for most devices under most conditions. However, there are quantum devices like resonant tunneling diodes which cannot be described at all within this semiclassical framework. Their operation is based on quantum interference effects arising from the wave nature of electrons. Moreover, as devices shrink to smaller dimensions, it is expected that quantum interference effects will become increasingly significant even in the operation of conventional devices. In order to include these effects we need a quantum version of the Boltzmann equation which combines quantum dynamics with a stochastic description of the scattering processes. The non-equilibrium Green's function (NEGF) formalism (also referred to as the Keldysh formalism) provides us with just that:

NEGF $\Rightarrow$ "Schrodinger equation"+ Random scattering

In this talk we will briefly summarize this formalism. For more details we refer the reader to the cited references and the references therein.

II. BASIC CONCEPTS

Before we can introduce the NEGF formalism we need to discuss a few basic concepts. Consider a homogeneous conductor. In the semiclassical picture we can describe the electrons by specifying the distribution function f(k) which tells us the number of electrons occupying a particular state 'k'. But in the quantum mechanical picture this is not enough. We also need to specify the phase-relationship among the different states. One way to do this is to define a density matrix $\rho(k,k')$. The distribution function f(k) only gives us the diagonal elements of this matrix:

$$f(k) = [\rho(k,k')]_{k'=k}$$

The rest of the story is contained in the off-diagonal elements which cannot be neglected unless the phase-relaxation length is much shorter than the other length scales.

It will be noted that although we have used a representation in terms of k-states to define the correlation function, we can always transform to other representations using an appropriate unitary transformation. For example we could transform to a real space representation as follows:

$$\rho(r,r') = \langle r|\rho|r'\rangle = \sum_{k,k'} \langle r|k\rangle\langle k|\rho|k'\rangle\langle k'|r'\rangle\rangle$$

$$= \frac{1}{V}\sum_{k,k'} \rho(k,k')\exp[i(kr - k'r')] \qquad (V \equiv \text{normalization volume})$$

In principle it is possible to find a representation that diagonalizes the correlation function. In such a representation there are no phase-correlations to worry about and we could use semiclassical reasoning. In practice it may not always be convenient to find this special representation or to use it.

To include the time coordinate into this description, in general we need a two-time correlation function of the form $G^n(k,k';t,t')$. In steady-state problems, the correlation function depends only on the ***difference*** between the two times and can be Fourier transformed to yield

$$G^n(k,k';E) \equiv \int d\tau\, G^n(k,k';\tau)e^{-iE\tau/\hbar} \qquad (\tau \equiv t - t') \qquad \text{(II.1)}$$

One way to understand the Fourier transform relationship between the energy 'E' and the difference time coordinate (t-t') is to note that the wavefunction of a particle with energy E evolves in time with a phase factor of exp [-iEt / $\hbar$]. Consequently

$$\psi(t)\psi^*(t') \sim \exp[-iE(t-t')/\hbar]$$

This suggests that the Fourier transform of the correlation function with respect to (t-t') should yield the energy spectrum.

Some treatments of quantum transport are based on the equal time correlation function obtained from $G^n(k,k';t,t')$ by setting t' = t. It is straightforward to show that this is equivalent to integrating $G^n(k,k'; E)$ over all energy:

$$\left[G^n(k,k';t,t')\right]_{t=t'} = \int \frac{dE}{2\pi} G^n(k,k';E)$$

As a result the energy-resolved information is lost making it difficult to describe scattering processes which transfer electrons from one energy to another. In general we need to use the full two-time correlation function. Since our interest is confined to steady-state transport, the correlation function depends only on the time difference (t - t') and can be Fourier transformed to obtain $G^n(k,k'; E)$ as described above. This energy-dependent correlation function is what we will use in this talk.

In the semiclassical picture we can define a function $S^{out}(k,t)$ that tells us the rate at which electrons are scattered out of a state 'k' assuming it is initially full. In a quantum mechanical description we have to generalize this concept, too, to include phase-correlations:

$$S^{out}(k,t) \rightarrow \Sigma^{out}(k,k';t,t')$$

Once again for steady-state problems the outscattering function depends only on the difference time coordinate and can be Fourier transformed yield an energy-dependent outscattering function $\Sigma^{out}(k,k'; E)$.

In deriving semiclassical kinetic equations we usually balance the outflow of electrons against the inflow of electrons. The inflow of electrons can alternatively be viewed as an outflow of 'holes' (whose number is given by (1-f)). We use the quotes as a reminder that we are talking about holes in the conduction band itself (we are considering only one band) and not in some other valence band. To describe the outflow of holes in the quantum formalism we define a ***hole correlation function*** G^p and an ***inscattering function*** Σ^{in} (which is a hole outscattering function) using exactly the same argument as we used above for electrons.

Before proceeding further we should point out that we are using a notation that is slightly different from the standard notation in the literature. We have deliberately chosen the notation to reflect the physical meaning of these functions. The correspondence, however, is quite straightforward:

$$G^n, G^p, \Sigma^{in}, \Sigma^{out} \rightarrow -iG^<, +iG^>, -i\Sigma^<, +i\Sigma^>$$

This set of four functions G^n, G^p, Σ^{in} and Σ^{out} (which are the quantum analogs of the semiclassical concepts f, (1-f), S^{in} and S^{out}) provide us with the language needed to include phase-correlations into a transport theory. If we represent our device by a set of 'N' nodes (in real space or in momentum space or in some other representation), then each of these quantities is a matrix of dimensions (NxN) at a given energy 'E". From hereon we will not write the energy coordinate 'E' explicitly for clarity.

III. KINETIC EQUATION

The correlation function is related to the scattering function by the relation

$$G^n = G^R \Sigma^{in} G^A \qquad \text{(III.1)}$$

This equation is written in matrix notation and could be applied in any convenient representation. The Green's function G^R is calculated from a Schrodinger-like equation

$$\left[EI - H_0 - \Sigma^R\right] G^R = I \qquad \text{(III.2)}$$

where H_0 is the Hamiltonian operator describing the device and I is the identity matrix (the other new function G^A is just the Hermitian conjugate of G^R).

To understand the physical meaning of the Green's function we note that in ordinary quantum mechanics the wavefunction of an electron is described by a Schrodinger equation $\left[EI - H_0\right]\Psi = 0$. Comparing with Eq.(III.2) for G^R we note two differences. Firstly there is a delta function source term (I) on the right hand side of Eq.(III.2) suggesting that the function $G^R(r,r')$ be interpreted as the wavefunction at 'r' due to a delta function source at r' in the position representation. Similar interpretations are of course possible in other representations as well. Secondly there is an extra term Σ^R known as the ***self-energy***. It represents the effective potential that an electron feels due to its interactions with phonons, other electrons etc (which are not included in H_0).

It is interesting that we can rewrite the Boltzmann equation (Eq.(I.1)) in a form that looks a lot like Eqs.(III.1) and (III.2). We could define a Green's function g^R as follows

$$v.\vec{\nabla} g^R + (e\vec{E}/\hbar).\vec{\nabla}_k g^R + (S^{out} + S^{in})\, g^R(r,k;r',k') = \delta(r-r')\,\delta(k-k') \qquad \text{(III.3a)}$$

and express the distribution function in terms of this Green's function:

$$f(r,k) = \int dr'\, dk'\, g^R(r,k;r',k')\, S^{in}(r',k') \qquad \text{(III.3b)}$$

IV. SELF-ENERGY FUNCTIONS

In order to perform any concrete calculations based on the Boltzmann equation we need a recipe for calculating the functions S^{in} and S^{out}. These functions describe the physics of the interactions and the precise recipe depends on what interaction we want to describe and what approximation we wish to use. In the quantum formalism the same is true of the functions Σ^{in}, Σ^{out} and Σ^{R}. For a detailed description of different types of interactions we refer the reader to [2]. Here we will simply summarize the results for electron-phonon interactions in the self-consistent Born approximation (SCBA) and for electron-electron interactions in the Hartree-Fock approximation.

Phonon scattering in lowest order perturbation theory is described by

$$\Sigma^{in}(\vec{r},\vec{r}';E) = \int d(\hbar\omega)\, D(\vec{r},\vec{r}';\hbar\omega)\, G^{n}(\vec{r},\vec{r}';E-\hbar\omega) \qquad \text{(IV.1a)}$$

$$\Sigma^{out}(\vec{r},\vec{r}';E) = \int d(\hbar\omega)\, D(\vec{r},\vec{r}';\hbar\omega)\, G^{p}(\vec{r},\vec{r}';E+\hbar\omega) \qquad \text{(IV.1b)}$$

where the function D describes the spatial correlation and energy spectrum of the phonons ($\hbar\omega > 0$ corresponds to absorption and $\hbar\omega < 0$ to emission).

$$D(\vec{r},\vec{r}';\hbar\omega) = \sum_{q} |U_q|^2 \left\{ \begin{array}{l} \exp[-iq.(\vec{r}-\vec{r}')]\, N_q \delta(\omega-\omega_q) \\ \quad + \exp[+iq.(\vec{r}-\vec{r}')]\,(N_q+1)\delta(\omega+\omega_q) \end{array} \right\} \qquad \text{(IV.2)}$$

where N_q is the number of phonons with wavevector q and frequency ω_q and U_q is the potential felt by an electron due to a single phonon with wavevector q. Assuming that the bath of phonons is always maintained in thermal equilibrium N_q is given by the Bose-Einstein function. The self-energy function is given by

$$\Sigma^{R}(E) = \Gamma^{H}(E) + \frac{i}{2}\Gamma(E) \qquad \text{where} \quad \Gamma(E) = \Sigma^{in}(E) + \Sigma^{out}(E) \qquad \text{(IV.3)}$$

and $\Gamma^{H}(E)$ is the Hilbert transform of $\Gamma(E)$.

Electron-electron interactions in the Hartree-Fock approximation, do not give rise to any Σ^{in}, Σ^{out}. It only contributes to Σ^{R} :

$$\Sigma^{R}(\vec{r},\vec{r}';E) = U_H(\vec{r})\,\delta(\vec{r}-\vec{r}') + \Sigma_F(\vec{r},\vec{r}') \qquad \text{(IV.4)}$$

The first term is the Hartree potential:

$$U_H(\vec{r}) = \int d\vec{r}' \int \frac{dE}{2\pi} G^{n}(\vec{r}',\vec{r}';E) \frac{e^2}{4\pi\varepsilon\,|\vec{r}-\vec{r}'|} \qquad \text{(IV.5)}$$

The second term is the exchange potential:

$$\Sigma_F^{s}(\vec{r},\vec{r}') = -\int dE\, G^{ns}(\vec{r},\vec{r}';E) \frac{e^2}{4\pi\varepsilon\,|\vec{r}-\vec{r}'|} \qquad \text{(IV.7)}$$

The superscript 's' is added as a reminder that an electron only feels an exchange potential due to other electrons of the same spin.

V. TERMINAL CURRENT

In general we are interested in calculating the current that flows when a conductor is connected by to two (or more) contacts across which a potential difference is maintained by an external source. So far we have not worried about the leads connected to the conductor. One way to treat the leads is to impose an appropriate boundary condition on $G^n(r,r';E)$ and $G^R(r,r';E)$ when solving Eqs.(III.1) and (III.2), in the same way that we impose boundary conditions on the distribution function f(r,k) in semiclassical theory. Once we have solved for $G^n(r,r';E)$, we can calculate the current density J(r;E) throughout the conductor and then integrate over the cross-section to obtain the current in the contact [6,7].

An alternative approach that is often very convenient is to introduce the effect of the leads through functions Σ_m^{in}, Σ_m^{out} and Σ_m^R (defined for each lead 'm') and add it to the functions Σ_φ^{in}, Σ_φ^{out} and Σ_φ^R describing the interactions.

$$\Sigma^{in,out,R} = \Sigma_\varphi^{in,out,R} + \sum_m \Sigma_m^{in,out,R} \qquad \text{(V.1)}$$

Using a discrete lattice, or what is often referred to as the 'tight-binding' model, the self-energy function due to the leads can be written as [5]

$$\Sigma_m^R(i,j;E) = -t\,\phi_m(i)\,e^{ik_m a}\,\phi_m^*(j) \qquad \text{(V.2)}$$

where $\quad E = U_m + 2t(1-\cos(k_m a))$

Here 'i' and 'j' are points on a discretized lattice with spacing 'a' and $t = \hbar^2/2ma^2$. The self-energy is non-zero only for lattice sites that are adjacent to the lead with mode 'm' and ϕ_m represents the transverse wavefunction corresponding to mode 'm'. U_m is the potential energy in lead 'm'.

The inscattering and outscattering functions corresponding to the leads are given by

$$\Sigma_m^{in}(i,j;E) = f_m(E)\,\Gamma_m(i,j;E) \qquad \text{(V.3a)}$$

$$\Sigma_m^{out}(i,j;E) = (1-f_m(E))\,\Gamma_m(i,j;E) \qquad \text{(V.3b)}$$

where $\quad \Gamma_m(i,j;E) = \phi_m(i)\frac{\hbar v_m}{a}\phi_m^*(j) \quad$ and $\quad \hbar v_m = \partial E/\partial k_m = 2at\sin(k_m a)$

Here we have assumed that each mode 'm' in the leads is maintained in local equilibrium with some Fermi distribution $f_m(E)$.

It seems feasible to do something similar in semiclassical theory as well, namely, define functions $S_m^{in}(k)$ and $S_m^{out}(k)$ corresponding to each contact 'm'. The current in lead 'm' can then be calculated from the relation

$$I_m \sim \sum_k S_m^{in}(k)(1-f(k)) - S_m^{out}(k)f(k) \qquad \text{(V.5)}$$

We are not aware of anyone using this approach in semiclassical theory, but it has been used successfully in the quantum version. The quantum analog of Eq.(V.5) is given by

$$I_m = \frac{2e}{h}\int dE\, Tr\left[\Sigma_m^{in}G^p - \Sigma_m^{out}G^n\right] \qquad (V.6)$$

VI. EXAMPLES

To apply Eqs.(III.1a,b) to a specific conductor we could discretize the spatial coordinate into a discrete lattice with N points. All the matrices like G^n, G^R, Σ^{in} etc. are then of order (NxN). the energy coordinate 'E' too has to be discretized into a convenient number of nodes. At each energy node, we have to invert (NxN) matrices. The functions Σ^{in}, Σ^{out} and Σ^R then have to be recomputed and the calculation repeated till the process converges.At low temperature and bias a single energy node is adequate allowing us to handle conductors with many spatial nodes. So far we have applied this formalism to (1) two -dimensional conductors at low temperature and bias. This includes the study of the Hall effect (low and high magnetic fields) in ballistic as well as disordered conductors [6] and (2) one-dimensional conductors at room temperature and large bias. This involves the study of current flow and energy dissipation in single barrier and double barrier structures [7].

VII. FUTURE DIRECTIONS

The NEGF formalism provides a general framework for quantum transport comparable to that provided by the Boltzmann formalism for semiclassical transport. In this talk we have described how this formalism can be used to describe steady-state transport in mesoscopic devices. Although the basic ideas seem clear, much remains to be done in terms of incorporating realistic scattering models and bandstructure effects.

There are two areas where the basic concepts are not fully clear. One is the area of transient or ac response. The other is the area of transport in strongly interacting systems. The Coulomb blockade regime which has attracted much attention lately belongs to this category. The calculation of self-energy and scattering functions describing the interactions (Section IV) is based on perturbation theory which is not valid for strong interactions. Under these conditions we cannot use Eqs.(III.1) and (III.2) to calculate G^n, G^R etc. Alternative non-perturbative techniques are needed [8,9]. Interestingly the Boltzmann formalism too runs into similar difficulties when applied to strongly interacting systems. We then have to worry about higher order (two-particle, three-particle) distribution functions.

ACKNOWLEDGEMENTS

Parts of this article have been excerpted from a forthcoming book by one of the authors [10]. This work was supported by the National Science Foundation and the Semiconductor Research Corporation.

REFERENCES

[1] L.P.Kadanoff and G.Baym. *Quantum Statistical Mechanics*. Frontiers in physics lecture note series, Benjamin/ Cummings (1962).
[2] P.Danielewicz, Annals of Physics, vol.152, p.239 (1984).
[3] G.D.Mahan, Physics Reports, vol.145, p.251 (1987).
[4] F.S.Khan, J.H.Davies and J.W.Wilkins, Phys. Rev.B36, p.2578 (1987).
[5] C.Caroli, R.Combescot, P.Nozieres and D.Saint-James, J.Phys.C: Solid State Physics, vol.5, p.21 (1972); see also article by R. Lake in this proceedings.
[6] M.J.McLennan, Y.Lee and S.Datta, Phys.Rev.B43, p.13846 and p.14333 (1991).
[7] R.K.Lake and S.Datta , Phys. Rev. B46, p.4757 (1992); R.K.Lake, G.Klimeck and S.Datta, Phys.Rev.B47, p.6427 (1993).
[8] Y.Meir and N.S.Wingreen, Phys.Rev.Lett., vol.68, p.2512 (1992).
[9] P.A.Lee, Physica B189, p.1 (1993).
[10] S.Datta. *Electronic Transport in Mesoscopic Systems, Chapter 8*. Cambridge University Press (to be published).

ELECTRON MOBILITY IN QUANTUM WIRES LIMITED BY OPTICAL-PHONON SCATTERING

T. Ezaki, N. Mori, K. Taniguchi and C. Hamaguchi
Department of Electronic Engineering, Osaka University
2-1, Yamada-oka, Suita City, Osaka 565, Japan

U. Ravaioli
Beckman Institute, University of Illinois
Urbana-Champaign, Urbana, IL 61801, USA

Abstract

Electron mobility in a quantum wire of GaAs/AlGaAs heterostructure is calculated by using the Boltzmann transport equation and detailed balance equation, where the results of self-consistent calculations for the eigen states and energies of electrons in the quantum wire system are used to evaluate the conductivity. Resonant behavior of the conductivity is expected due to the resonant scattering of electrons between the subbands induced by longitudinal optical phonon scattering. The resonance is shown to depend on the magnitude of one-dimensional form factors which are proportional to the transition probability of electrons in quasi-one dimensional system.

1. INTRODUCTION

It has been pointed out that the suppression of small angle scattering in quasi-one dimensional structures results in an enhancement of electron mobility at low temperatures [1]. At high temperatures, however, electron-longitudinal optical (LO) phonon scattering will play an important role in quantum wires (QWs) of GaAs/AlGaAs. In QWs fabricated on GaAs/AlGaAs heterostructures, electrons are confined just below the hetero-interface and the gate electrodes fabricated on the surface form quasi-one dimensional electron gas (Q1DEG), where the electron density in channel area can be controlled by gate voltage. In other words, electronic states may be changed by the gate voltage, and thus we can tune the inter-subband energy to the optical phonon energy. Since the electron mobility in QWs at high temperatures is limited by the optical phonon scattering, an oscillatory behavior of electron mobility is expected when the gate voltage is changed [2, 3] or when a high magnetic field is applied [4, 5]. In the present work, we calculate the conductivity in a typical gated QW structure and show that the resonant behavior similar to magnetophonon resonance appears by changing the gate voltage without the presence of magnetic field. For this purpose we solve the Poisson and Schrödinger equations self-consistently and obtain one dimensional eigen states and then calculate electron mobility in a QW at high temperatures. The magnitude of the resonance depends strongly on the parity of the wave functions in the direction perpendicular to the heterointerface.

2. SELF-CONSISTENT CALCULATION

We consider a mesa etched quantum wire structure [6] shown in Fig 1. As shown in Fig 1, we choose y and z directions as parallel and normal to the interface, respectively, and the motion of electrons is quantized in these two directions. Up to 20 subbands are calculated self-consistently including different confinement in z direction, where we solved the Poisson and Schrödinger equations numerically by discretizing the structure in a nonuniform rectangular mesh. In Fig 2, we present calculated results of subband energies as a function of gate voltage. The subband index (n, m) is used in Fig 2, which represents node number n of the wave function in the y and node number m in the z directions. For example, eigen state $(0, 1)$ indicates the first quantization state in y direction

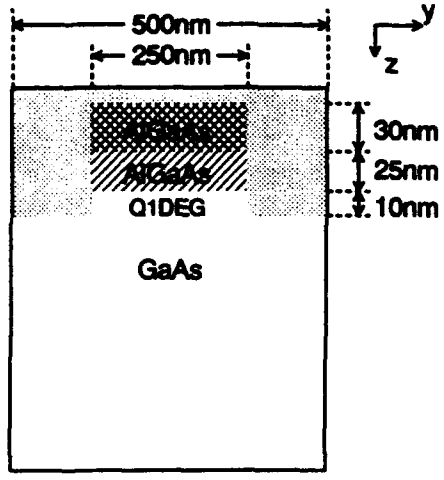


Figure 1: Cross-section of a mesa etched Al-gated QW structure used for the present calculations. The device consists of an unintentionally p-doped GaAs substrate ($N_A < 10^{14}cm^{-3}$), followed by an undoped AlGaAs spacer layer and an n-doped AlGaAs cap layer ($N_D = 1.5 \times 10^{18}cm^{-3}$).

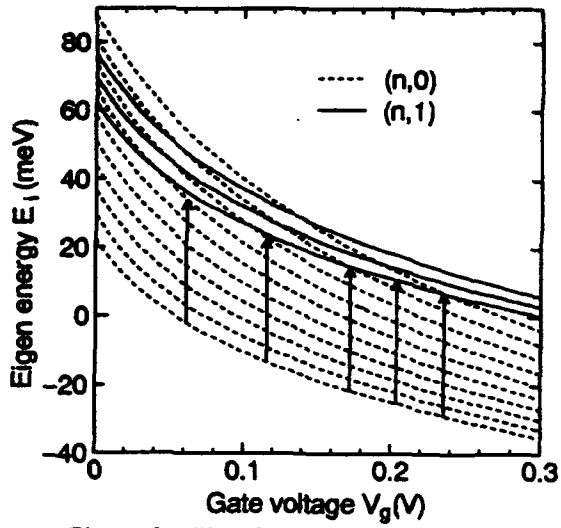


Figure 2: The electron eigen energies in the QW shown in Fig 1 at $T = 150$ K. The dotted lines represent the eigen states $(n, 0)$ and the dashed lines represent the eigen states $(n, 1)$. The arrows indicate relevant transitions. The index (n, m) represents the number of nodes n of the wave functions in the y direction and m in the z direction.

and the second quantization state in z direction. It is seen in Fig 2 that the eigen energies are lowered when the gate voltage is swept in the forward direction, and that many subbands move below the Fermi level (0 eV of the vertical scale), resulting in an enhancement of carrier population and the reduction of subband spacing. A typical result of the electron wave functions in the QW is shown in Fig 3, where the subband energies and potential profile are also plotted by dotted lines and dashed curve, respectively. As shown in Fig 3, the potential profile along z direction is quasi-triangular and the first peek (lefthand side) of the wave functions appears at almost the same position. Since the matrix elements (the form factors) are expressed by the overlap integral [4], we may expect a large value for the overlap integral between eigen states of same confinement in the y direction and different confinement in the z direction and the strong transition between the $(n, 0)$ and $(n, 1)$ states. In the following we present a calculated result of the conductivity in the QW structure.

3. CALCULATION OF CONDUCTIVITY IN A QUANTUM WIRE

The form factor $G_{ii'}$ and the scattering probability $W_{ii'}$ between subband i and i' are given by the following equations [4].

$$G_{ii'}(q_x) = \int d\rho_1 \int d\rho_2 K_0(q_x|\rho_1 - \rho_2|)\Psi^*_{i'}(\rho_1)\Psi_i(\rho_1)\Psi^*_i(\rho_2)\Psi_{i'}(\rho_2) \tag{1}$$

$$W_{ii'}(k,k') = \alpha\omega_0\sqrt{\frac{2\hbar\omega_0}{m^*}}\left(N_0 + \frac{1}{2} \pm \frac{1}{2}\right) G_{ii'}(q)\delta(\varepsilon(k') - \varepsilon(k) + E_{i'} - E_i \pm \hbar\omega_0) \tag{2}$$
$$(q = k - k'),$$

where $\Psi_i(\rho)(\rho = (y, z))$ is the wave function of the eigen state $i = (n, m)$ and K_0 is the modified

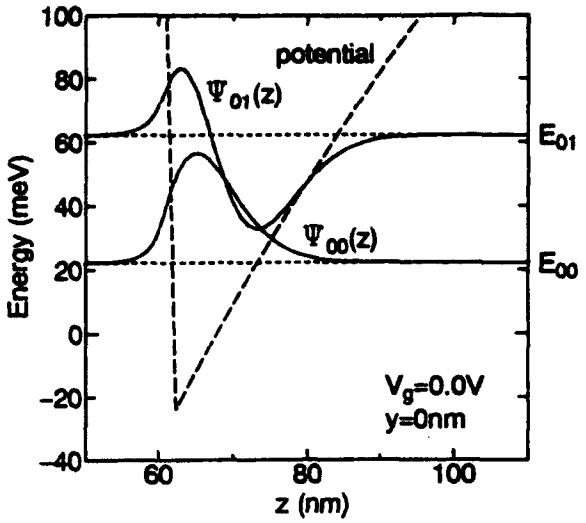


Figure 3: Calculated electron wave functions (solid lines) and potential profile (dashed line) along z direction of the QW shown in Fig 1 at $V_g = 0.0\,\mathrm{V}$. Ψ_{00} and Ψ_{01} are the wave functions of the $(0,0)$ state (ground state) and the $(0,1)$ state, respectively.

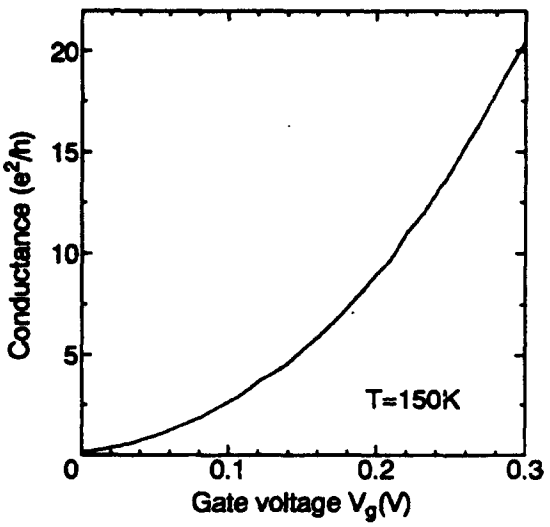


Figure 4: Calculated conductance as a function of gate voltage in the QW shown in Fig 1 at $T = 150\,\mathrm{K}$.

Bessel function of the second kind, α is the Fröhlich's coupling constant, $m^* = 0.067m_0$ is the electron effective mass and $\hbar\omega_0 = 36.2\,\mathrm{meV}$ is the LO-phonon energy in GaAs bulk, N_0 is phonon number, and E_i denotes subband energy of eigen state i. The $\pm$ sign in equation (2) corresponds to the emission and the absorption of LO-phonon, respectively.

The electron conductivity in a QW is calculated as a function of gate voltage using the Boltzmann transport equation and the detailed balance principle. Figure 4 shows the calculated result of conductance-gate voltage characteristic of a QW at $T = 150\,\mathrm{K}$. It is very interesting to point out that the calculated magnitude of the conductivity is very close to the experimental result of Ismail [2], where he used similar structure of a quantum wire. He observed a dip in the conductance-gate voltage characteristic, whereas the present calculation shows several weak dips in the region from 0.1 to 0.3 V. The increase in the conductivity with the gate voltage is interpreted in terms that the electron density increases with increasing the gate voltage. In order to see the weak structure in the conductance curve more clearly, we deduced the oscillatory components by deducting a smooth curve of the least square fit from the conductance. The oscillatory component thus obtained is plotted in Fig 5, where we see more detailed structure. Although the oscillatory structure is very complicated, some of them are well explained with the help of the results shown in Fig. 2, where the arrows show the resonant transition of the electrons (the length of the arrow is the LO phonon energy). From a comparison between Fig. 2 and Fig. 5, the oscillatory structures at about 0.2 and 0.24 V are ascribed to the resonant transition from the subband with the index $m = 0$ to the subbands with the index $m = 1$. Other transitions indicated by the arrows in Fig 2 are expected to be weaker because of the small value of the form factor as discussed in previous section. From the present work we find that the strength of transitions between the subbands in a QW structure strongly depends on form factors. We present the calculated form factors in Fig 6 for a gate voltage $V_g = 0.2\,\mathrm{V}$. It is clearly seen in Fig 6 that the form factor for the transition between the $(0,0)$ state and the $(0,1)$ state is large enough to dominate the reduction in the conductance because the energy separation between the subbands is very close to the LO phonon energy at this gate voltage. Similar situation occurs at the gate voltage $V_g = 0.24$. These oscillatory structures are found to be

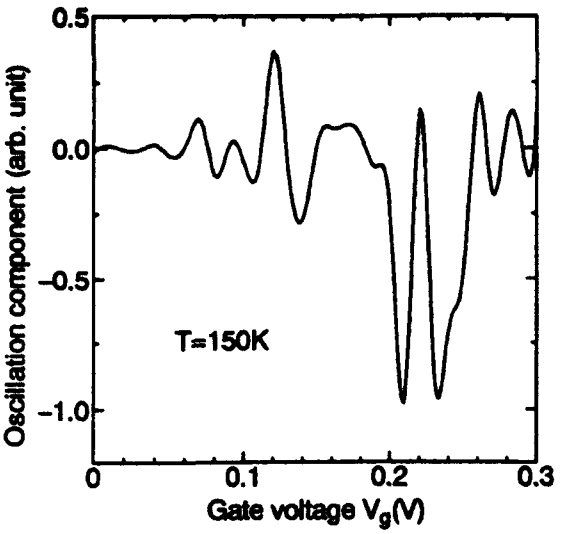


Figure 5: The oscillatory structures of the conductance at $T = 150\,\mathrm{K}$.

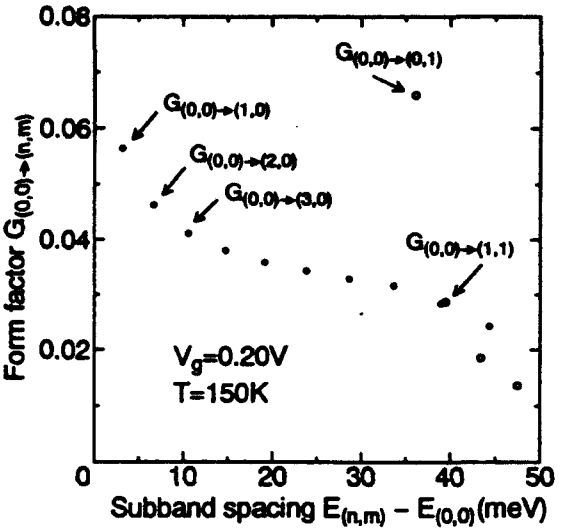


Figure 6: The form factors G_{0i} (see the text for the definition) between the ground state and the excited state in the QW at $V_g = 0.2\mathrm{V}$ (solid circles : $G_{(0,0)\to(n,0)}$, open circles : $G_{(0,0)\to(n,1)}$).

very weak compared with the experimental result reported by Ismail [2]. Taking into account the difficulty in obtaining a uniform QW structure, the energy subband structure of a real QW is not so sharp compared with the ideal one used in the present calculation and the broadened nature of the density of states will allow the LO phonon scattering in wider range of energy, resulting in a broadened conductance minima.

4. CONCLUSION

Self-consistent calculation was carried out to obtain electronic eigen states in a QW and the conductance was evaluated by calculating the electron mobility based on Boltzmann transport equation. The calculated conductance was found to exhibit minima as the resonant intersubband transition occurs. The dominant contribution to the minima arises from the intersubband transition between the (0,0) and (0,1) subbands in the QW structure used in the present calculations.

REFERENCES

[1] H. Sakaki, Jpn. J. Appl. Phys. **19** (1980) L735.

[2] Ismail: *Science and Technology of Mesoscopic Structures* (Springer-Verlag, Tokyo, 1992) p. 135.

[3] D. Jovanovic, J. P. Leburton, K. Ismail, J. M. Bigelow and M. H. Degani, Appl. Phys. Lett **62** (1992) 2824.

[4] N. Mori and C. Hamaguchi, Proc. Int. Workshop on Computational Electronics, University of Illinois at Urbana-Champaign, 1992, p. 261.

[5] G. Berthold, J. Smoliner, C. Wirner, E. Gornik, G. Böhm, G. Weimann, M. Hauser, C. Hamaguchi, N. Mori and H. Momose, Semicond. Sci. Technol. **8** (1993) 735.

[6] F. Hirler, J. Smoliner, E. Gornik, G. Weimann and W. Schlapp, Appl. Phys. Lett. **57** (1990) 261.

LOSS OF PHASE COHERENCE IN SEMICONDUCTOR HETEROSTRUCTURES DUE TO THE COULOMB INTERACTION

M. Žiger† and W. Pötz *

† *Universität Graz, Institut für Theoretische Physik, A-8010 Graz, Austria*

* *University of Illinois at Chicago, Physics Department, Chicago, IL 60607, USA*

ABSTRACT

We present a theoretical study of the dynamics of free photo-generated carriers in asymmetric GaAs-AlGaAs double wells. Photo-generation occurs on a sub-picosecond time scale and produces a coherent ensemble of electron-hole pairs in the wider well. The simultaneous thermalization and tunneling of electrons between the two wells is analyzed within the density matrix approach. The interplay between tunneling and Coulomb scattering is analyzed at several levels of approximation regarding free carrier screening. We find that the Coulomb interaction represents an effective agent to destroy phase coherence and to damp out charge oscillations. Nevertheless, our calculations predict that, if the free carrier Coulomb interaction represents the dominant dephasing mechanism, charge density oscillations associated with *free* carriers should be observable up to carrier sheet densities of about 10^{10} cm^{-2} .

I. INTRODUCTION

Various optical techniques have been used to monitor transport (tunneling) and thermalization of hot photo-generated carriers in quantum-well structures and superlattices. Recently, measurement of dipole radiation signals has been used to demonstrate both the existence of Bloch oscillations in superlattices and charge oscillations due to tunneling in semiconductor double wells.[1-4] To our knowledge, all charge oscillations which have been observed up to now have been attributed to excitons. No evidence for charge oscillations associated with free carriers has been presented so far. There are several reasons why excitons are more likely to exhibit charge oscillations than free carriers. Excitons are less exposed to structural imperfections in the double well. Moreover, any imperfections tend to detune the exciton levels and their optical excitation is inhibited. Excitons are neutral quasiparticles and thus interact with each other and free carriers more weakly than free carriers among each other, leading to longer phase coherence times. Finally, excitons are known to dominate the four-wave mixing signal, even when vastly outnumbered by free carriers.[5]

In this work we investigate the possibility of inducing charge oscillations of free carriers in double well structures. In particular, the role of free-carrier screening is investigated.

II. THEORY

We apply the density matrix approach to a situation in which a sub-picosecond laser pulse generates free electron-hole pairs of low to moderate densities in asymmetric GaAs-AlGaAs double-wells.[6,7] We consider an experimental situation identical to the one under which excitons have been created resonantly in double wells, except that here the laser energy of maximum intensity exceeds the energy gap by typically 10 to 20 meV.[2,4] As the hole bands involved in the excitation process are far off resonance, hole dynamics is neglected. Due to the low excess photon energy, optical phonon emission is unimportant. Here, the formation of excitons via LO phonon emission is neglected.

The problem reduces to a study of the time-evolution of an electronic one-particle density matrix

$$\begin{pmatrix} f_{LL}(k,t) = \langle b^\dagger_{Lk} b_{Lk} \rangle(t) & f_{LR}(k,t) = \langle b^\dagger_{Lk} b_{Rk} \rangle(t) \\ f_{RL}(k,t) = \langle b^\dagger_{Rk} b_{Lk} \rangle(t) & f_{RR}(k,t) = \langle b^\dagger_{Rk} b_{Rk} \rangle(t) \end{pmatrix}$$

with

$$\langle A \rangle = Tr\{\rho A\}(t)$$

for an electron observable A and density operator ρ. $f_{LL}(k,t)$ and $f_{RR}(k,t)$ are the electron distribution functions associated with left and right well, respectively. k denotes the magnitude

of the k-vector associated with in-plane motion. The off-diagonal element $f_{RL}(k,t)$ denotes the "polarization".

The time evolution of the density matrix originates from several sources:

• Firstly, laser generation of electron-hole pairs is incorporated as a generation term of the form

$$\frac{d}{dt} f_{LL}(k,t)|_{laser}.$$

As we are not concerned with ultra-short pulses this term should be adequate for the present purpose.

• Secondly, elastic tunneling between left and right well is taken into account within a two-subband approximation and the Hamiltonian

$$H_o = \sum_k \{\epsilon_{Lk} b^\dagger_{Lk} b_{Lk} + \epsilon_{Rk} b^\dagger_{Rk} b_{Rk} + V[b^\dagger_{Lk} b_{Rk} + b^\dagger_{Rk} b_{Lk}]\}$$

The basis states $|Lk\rangle$ and $|Rk\rangle$ are linear combinations of the two lowest eigenstates of the double well, $|+,k\rangle$ and $|-,k\rangle$ with eigenvalues $\epsilon_{+,k}$ and $\epsilon_{-,k}$, respectively. Here, $\epsilon_{\alpha,k} = \epsilon_\alpha + \frac{(\hbar k)^2}{2m^*}, \alpha = \pm, L, R$. Perfect interfaces are assumed. This provides a major reduction in complexity of the problem, but may be somewhat unrealistic in real structures.

• Thirdly, the Coulomb interaction between free carriers,

$$v = \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta,q,k,k'} v_{\alpha\beta\gamma\delta}(q) b^\dagger_{\alpha k+q} b^\dagger_{\beta k'-q} b_{\delta k'} b_{\gamma k},$$

$\alpha,\beta,\gamma,\delta = L, R$, leads to nonlinear terms in the equations of motion which tend to destroy phase coherence in the system.[8]

We apply a decomposition procedure of the structure $\langle \tilde{b}^4 \rangle(t) \approx \langle \tilde{b}^2 \rangle(t) \langle \tilde{b}^2 \rangle(t)$ to truncate the BBGKY hierarchy in the many-particle density matrix elements and arrive at a self-consistent and closed Markovian set of non-linear differential equations of first-order in time.[8]

The Coulomb matrix elements $v_{\alpha\beta\gamma\delta}$ are evaluated approximately for wave functions associated with infinitely deep wells. In particular we consider only matrix elements of the form $v_{\alpha\beta\alpha\beta}$. Termination of the BBGKY hierarchy at second order in v requires implementation of free-carrier screening by hand. Here, we adopt a commonly used short-cut and treat screening within the random-phase approximation (RPA). The retarded density fluctuation correlation function is evaluated within the plasmon-pole approximation (PPA)

$$D^o_\alpha(q,\omega) \approx D^o_{PPA\alpha}(q,\omega) = \frac{1}{v_c(q)} \frac{\omega^2_{pl\alpha}(q)}{(\omega + i\eta)^2 - \frac{q}{\kappa_\alpha}\omega^2_{pl\alpha}(q)}.$$

$\omega^2_{pl\alpha}(q) = (2\pi e^2 n_\alpha)/(\varepsilon_o m^*) q$ is the square of the plasmon frequency at sheet charge density n_α.[7] $\kappa_\alpha = \frac{2}{a^*_B} f_{\alpha\alpha}(0)$ is the $q = 0$ screening wave vector in two dimensions and a^*_B is the effective Bohr radius.

III. NUMERICAL RESULTS AND DISCUSSION

For a discussion of our numerical results we choose the left well L to be a $170\mathring{A}$ GaAs well and the right well R to be a $120\mathring{A}$ GaAs well, separated by a $17\mathring{A}$ AlGaAs barrier. Here, we consider an electric field which provides resonance between the lowest electronic subband associated with each of the (isolated) quantum wells. The duration of the excitation pulse was varied between "0" and 0.5 ps with an average carrier excess energy of up to $20meV$. The latter is to ensure validity of the two-subband approximation, as well as the neglect of optic phonon effects. The pulse width used was $4.2meV$. Particle densities between 10^9 and 10^{11} cm^{-2} have been considered.

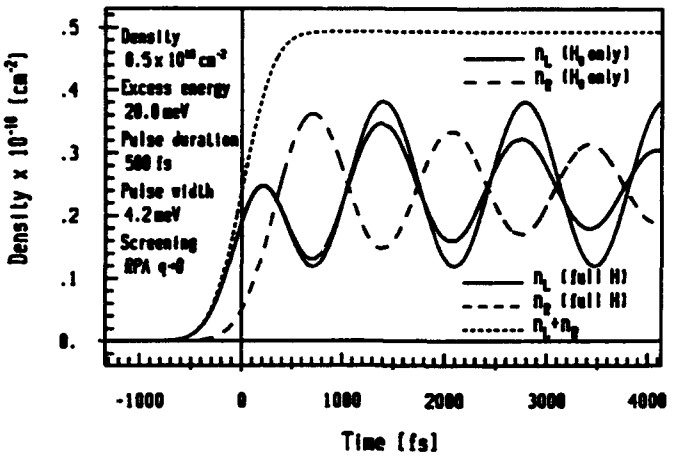


Fig. 1. Charge density in left and right well as a function of time.

A typical set of data is given in Figs. 1 to 3. Fig 1 shows the time–evolution of the total number of electrons in the left well, solid line, and in the right well, dashed line. The undamped solid line gives the carrier density in the left well in the absence of the free carrier interaction, while the dotted line gives the total number of carriers in the double well. At 5×10^9 carriers per cm^2, we observe charge oscillations, however, the Coulomb interaction provides strong damping of the latter. The polarization is plotted as a function of time in Fig. 2. It also displays a damped oscillatory behavior. Its second time derivative is proportional to the radiated electric field. In case of resonant exciton excitation, up to about 15 such oscillations have been observed in a similar structure.[2]

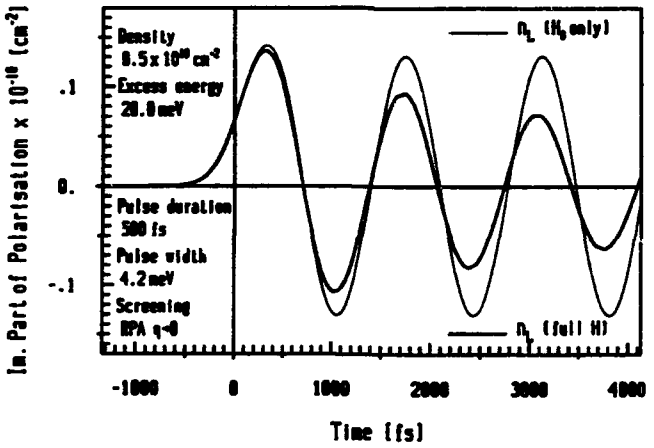


Fig. 2. Electron polarization $f_{LR}(t)$ as a function of time.

Simultaneous to the damping of the charge oscillations, thermalization among the electrons takes place on the time–scale of a few picoseconds. Fig. 3 (lhs) and Fig. 3 (rhs), respectively, give $f_{LL}(k,t)$ and $f_{RR}(k,t)$ as a function of k at selected times.

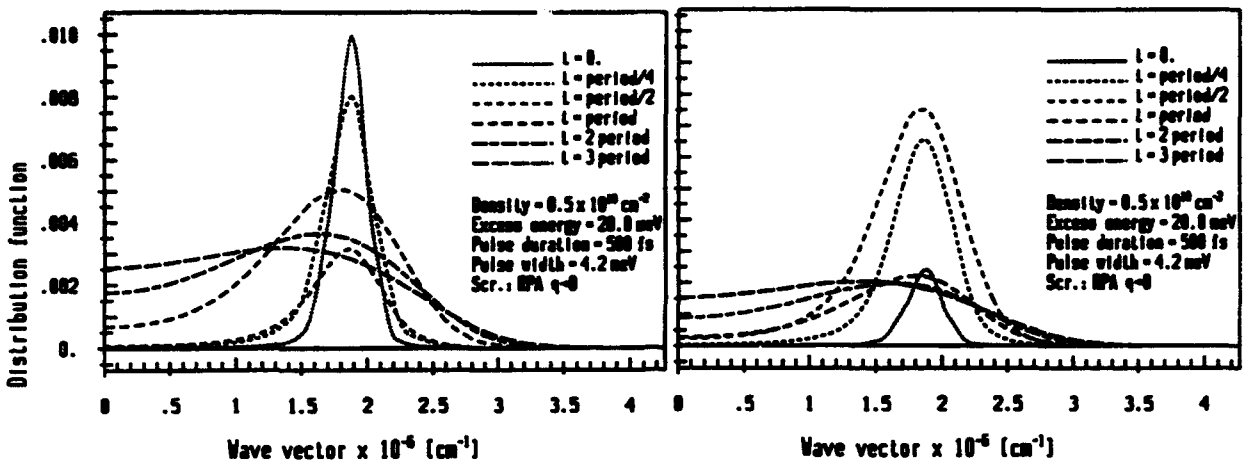


Fig. 3. lhs: distribution function $f_{LL}(k)$; rhs: distribution function $f_{RR}(k)$ at selected times.

IV. SUMMARY AND OUTLOOK

Our results can be summarized as follows. Charge oscillations due to *free* photo–generated carriers should be observable in asymmetric semiconductor double wells of high structural quality, provided that the inter–carrier Coulomb interaction provides the dominant dephasing mechanism. They should be observable up to about 10^{10} carriers per cm^2. Above this value the Coulomb interaction becomes so effective that it suppresses the onset of charge oscillations. Simultaneous to the destruction of phase coherence, the Coulomb interaction provides rapid thermalization of the photo–excited electrons.

It is remarkable that our results are rather insensitive to the screening model which is employed. The (dynamical) PPA, static RPA, Debye–Hückel, and Thomas–Fermi approximation produce practically identical results. This is largely due to the long–range nature of the Coulomb interaction and competing terms in the balance equations.

Further improvements, such as inclusion of electron–hole scattering, carrier–phonon interactions, and excitons, are desirable to clarify their role in the dephasing process.

ACKNOWLEDGMENT: This work has been supported by the US Army Research Office. W. P. wishes to acknowledge the hospitality received from the Institute for Theoretical Physics, University of Graz, Austria, during the early stages of this project.

V. REFERENCES

1. C. Waschke, H. G. Roscos, R. Schwedler, K. Leo, H. Kurz, and K. Köhler, Phys. Rev. Lett. **70**, 3319 (1993).
2. H. G. Roskos, M. C. Nuss, J. Shah, K. Leo, D. A. B. Miller, A. M. Fox, S. Schmitt–Rink, and K. Köhler, Phys. Rev. Lett. **68**, 2216 (1992).
3. M. S. C. Luo, S. L. Chuang, P. C. M. Planken, I. Brener, and M. C. Nuss, Phys. Rev. B **48**, 11043 (1993).
4. K. Leo, J. Shah, E. O. Göbel, T. C. Damen, S. Schmitt–Rink, W. Schäfer, and K. Köhler, Phys. Rev. Lett. **66**, 201 (1991).
5. D. S. Kim, J. Shah, J. E. Cunningham, T. C. Damen, W. Schäfer, M. Hartmann, S. Schmitt–Rink, Phys. Rev. Lett. **68**, 1006 (1992).
6. C. P. Slichter, "*Principles of Magnetic Resonance*", third edition, (Springer, Berlin, 1992).
7. H. Haug and S. Koch, "*Quantum Theory of the Optical and Electronic Properties of Semiconductors*", second edition, (World Scientific, Singapore, 1993).
8. M. Žiger, W. Pötz, and P. Kocevar, SPIE Proceedings 1993, to be published.

INCORPORATING SPATIALLY VARYING EFFECTIVE-MASS IN THE WIGNER-POISSON MODEL FOR AlAs/GaAs RESONANT-TUNNELING DIODES

K. K. Gullapalli and D. P. Neikirk
Microelectronics Research Center, MER 1.604 / 79900
The University of Texas at Austin, Austin, TX 78712 (512)-471-8104

Abstract

We present a single band equation of motion for the Wigner function, incorporating the effects of a spatially varying band structure. The transport equation is discrete in position, shedding light into the numerical aspects of the problem. While conventional upwind differencing to approximate the drift term was found adequate for $Al_{0.3}Ga_{0.7}As$/GaAs devices, it is completely unsatisfactory in modeling AlAs/GaAs resonant-tunneling diodes, particularly when the large change in effective-mass is included. Suggesting a new approach, meaningful steady state conduction curves for AlAs/GaAs diodes are presented for the first time.

Physical Model

With the need for studying effects of the detailed bandstructure such as Γ-X transfer in mind, the band structure is Fourier expanded: $E(k)=\sum_{n=1}^{\infty}\left(4\hbar^2/n^2 m^*_n a^2\right)\left[1-\cos(nka/2)\right]$, a being the lattice constant. The Brillouin zone is [$-2\pi/a$, $2\pi/a$]. Given that the masses m^*_n are spatially varying, we obtain the following equation of motion for the Wigner function:

$$\begin{aligned}\frac{\partial f}{\partial t}=&-\sum_{n=1}^{\infty}\frac{2}{n}\frac{\hbar\sin(nka/2)}{m^*_{n,\,GaAs}\,a}\left[\frac{f(q+na/4,k)-f(q-na/4,k')}{na/2}\right]+\frac{2}{\pi\hbar}\int dk'\,f(q,k')V(q,k-k')+\left.\frac{\partial f}{\partial t}\right|_{coll.}\\&-\sum_{n=1}^{\infty}\frac{4}{n^2}\frac{\hbar}{\pi a^2}\left\{\int dk'\,\sin\left(\frac{nk'a}{2}\right)\left[f\left(q+\frac{na}{4},k'\right)M^e_n\left(q+\frac{na}{4},k-k'\right)-f\left(q-\frac{na}{4},k'\right)M^e_n\left(q-\frac{na}{4},k-k'\right)\right]\right.\\&\left.+\int dk'\,\cos\left(\frac{nk'a}{2}\right)\left[f\left(q+\frac{na}{4},k'\right)M^o_n\left(q+\frac{na}{4},k-k'\right)+f\left(q-\frac{na}{4},k'\right)M^o_n\left(q-\frac{na}{4},k-k'\right)\right]\right\}\\&+\sum_{n=1}^{\infty}\frac{4}{n^2}\frac{2\hbar}{\pi a^2}\int dk'\,f(q,k')M^o_n(q,k-k')\end{aligned}\qquad(1)$$

where the first line in eq. 1 is the equation of motion if the effective-mass were uniform and

$$M^e_n(q,k)=\int dr\left(\frac{1}{m^*_n(r)}-\frac{1}{m^*_{n,\,GaAs}}\right)\cos[2k(q-r)],\quad M^o_n(q,k)=\int dr\frac{\sin[2k(q-r)]}{m^*_n(r)},\quad V(q,k)=\int dr\,v(r)\sin[2k(q-r)]$$

v(r) includes, in addition to the self-consistent potential, the Γ-Γ offset between the two materials.

Using the "minimal Hermitian form" ($\hat{H}=-\left(\hbar^2/2\right)\partial/\partial z(1/m^*)\partial/\partial z+v$)[1] to describe the effects of spatially varying effective-mass is inconsistent with the Weyl transform. By the Weyl correspondence rule, the Hamiltonian in position representation for a parabolic energy band is [2]:

$$\hat{H}=-\frac{\hbar^2}{8}\left[\frac{1}{m^*(z)}\frac{\partial^2}{\partial z^2}+2\frac{\partial}{\partial z}\frac{1}{m^*(z)}\frac{\partial}{\partial z}+\frac{\partial^2}{\partial z^2}\frac{1}{m^*(z)}\right]+v(z)$$

The equation of motion for the Wigner function in a parabolic band is just the $a\rightarrow 0$ limit of eq. 1.

Numerical Model

Computational resources limit the numerical treatment to nearest and second nearest neighbor coupling, and hence only two (appropriately chosen) components in eq. 1 can be included. For Γ-Γ tunneling the n=4 component is sufficient. The rate of change of the Wigner function $f(q, k)$ being determined only by its values at q, $q+a$, and $q-a$, eq. 1 is solved on the discrete phase-space given by $\{q_j \mid q_j = j\Delta; j = 1,2,\ldots,N_q\}$ and $\{k_n \mid k_n = \pi(2n-1-N_k)/(2N_k\Delta); n = 1,2,\ldots,N_k\}$ with $\Delta=a$ (5.6533 Å). We get a set of linear algebraic equations $\sum_{j'n'} L_{jn;j'n'} f_{j'n'} = b_{jn}$ where

$$L_{jn;j'n'} = -\frac{\sin(2k_{n'}\Delta)M^{e}_{j'n-n'}}{2N_k\Delta}\left(\frac{\delta_{j'j+1}-\delta_{j'j-1}}{2\Delta}\right) - \frac{\cos(2k_{n'}\Delta)M^{o}_{j'n-n'}}{4N_k}\left(\frac{\delta_{j'j+1}+\delta_{j'j-1}}{\Delta^2}\right)$$
$$+\frac{M^{o}_{j'n-n'}+2V_{j'n-n'}}{N_k}\delta_{j'j} + \frac{\delta_{j'j}}{\tau_f}\left(\frac{f^{eq}}{\sum_{n''} f^{eq}} - \delta_{n'n}\right) + T_{jn;j'n'} \qquad (2)$$

and b is the boundary contribution. The resulting matrix equation is solved using block LU factorization. T, the discrete drift term (the first term on the right-hand-side of eq. 1) will be discussed shortly. M^e, M^o and V are evaluated using fast sine and cosine transforms:

$$M^{e}_{jn-n'} = \frac{(-1)^{n'-n}}{2}\left(\frac{1}{m^*_{j+N_k/2}} + \frac{1}{m^*_{j-N_k/2}} - \frac{2}{m^*_{GaAs}}\right) + \left(\frac{1}{m^*_{j}} - \frac{1}{m^*_{GaAs}}\right) + \sum_{j'=1}^{N_k/2-1}\left(\frac{1}{m^*_{j+j'}} + \frac{1}{m^*_{j-j'}} - \frac{2}{m^*_{GaAs}}\right)\cos\left[\frac{\pi(n'-n)j'}{N_k/2}\right]$$

$$M^{o}_{jn-n'} = \sum_{j'=1}^{N_k/2-1}\left(\frac{1}{m^*_{j+j'}} - \frac{1}{m^*_{j-j'}}\right)\sin\left[\frac{\pi(n'-n)j'}{N_k/2}\right], \qquad V_{jn-n'} = \sum_{j'=1}^{N_k/2-1}\left(v_{j+j'} - v_{j-j'}\right)\sin\left[\frac{\pi(n'-n)j'}{N_k/2}\right]$$

For a parabolic energy band, the drift term appears as the spatial derivative of the Wigner function. It is then suggested that a stable numerical model be obtained by upwind differencing the drift term [1, 3, 4]. Here however, the drift term is already discrete in position. We obtain a numerically stable model by making the following approximation in T:

$$[f(q+\Delta)-f(q-\Delta)] \rightarrow [f(q+\Delta)-f(q-\Delta)] + \delta[f(q+\Delta)-2f(q)+f(q-\Delta)] \qquad (3)$$

where $|\delta| \ll 1$ so that the deviation from eq. 1 is small, and $k\delta < 0$ for stability (upwind bias). When $|\delta|=1$, we have first order upwinding, which is used at the device boundaries. When $|\delta|=0$, we have centered differencing. We apply an upwind bias only to T, the constant effective-mass drift term. The current density is defined to satisfy the discrete current continuity equation:

$$J_{j+1/2} = \frac{\hbar}{8N_k\Delta^2}\sum_n\left[\frac{\sin(2k_n\Delta)\left[(1\pm\delta)f_{j+1n}+(1\mp\delta)f_{jn}\right]}{m^*_{GaAs}} + \frac{1}{N_k}\sum_{n'}\sin(2k_{n'}\Delta)\left[M^{e}_{j+1n-n'}f_{j+1n'} + M^{e}_{jn-n'}f_{jn'}\right]\right]$$

The top sign is used for $k < 0$ and the bottom sign for $k > 0$. Since V(q,k-k') strongly couples the upwind and downwind flows, the fact that the differencing in eq. 3 is not transportive should not be of major concern. In any case, the exact equation of motion is not transportive either.

Simulation Results

The above model is first applied to the most commonly simulated $Al_{0.3}Ga_{0.7}As/GaAs$ resonant-tunneling diodes[1, 3, 4]. The conduction band offset is 0.27eV, $m^*_{GaAs}=0.067m_0$, and $m^*_{AlGaAs}=0.092m_0$. 30Å barriers sandwich a 50 Å well. The applied bias is dropped linearly across the double barrier quantum well structure. The contact is doped n-type at 2×10^{18} cm^{-3}. Figure 1 shows the flat-band results. The results due to Tsuchiya *et al*. [1] are also shown.

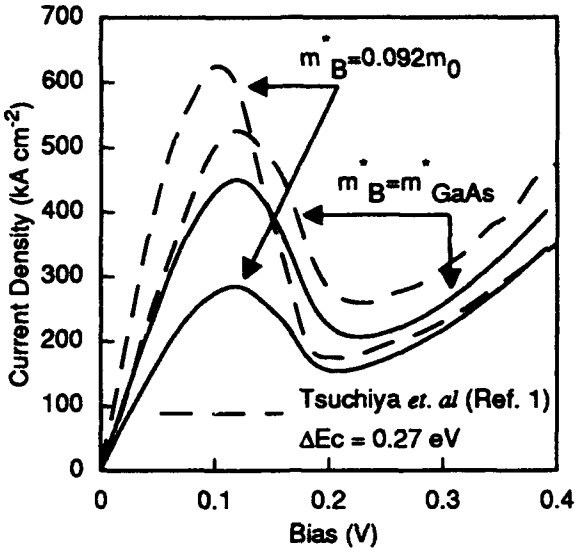


Figure 1: Wigner conduction curves with first order upwinding for $Al_{0.3}Ga_{0.7}As/GaAs$ diode. Collisions are ignored. $N_q = 80$, $\Delta = a$, $N_k = 64$ as in [1]. Solid lines: our model, dashed lines: Tscuhiya *et al.* [1]. The current density should decrease with increasing effective-mass in the barriers.

AlAs/InGaAs or AlAs/GaAs diodes are the choice for high speed applications due to their high peak current densities and peak-to-valley ratios [5-7]. The inadequacy of the first order upwinding begins to surface as we attempt to simulate AlAs/GaAs resonant-tunneling diodes. Here we consider our baseline AlAs/GaAs resonant-tunneling diode. The conduction band offset is taken to be 1.0eV, $m^*_{GaAs}=0.067m_0$, and $m^*_{AlAs}=0.15m_0$. 17 Å barriers sandwich a 50 Å well. On either side of the tunneling structure is a three step spacer layer consisting of 50 Å undoped GaAs

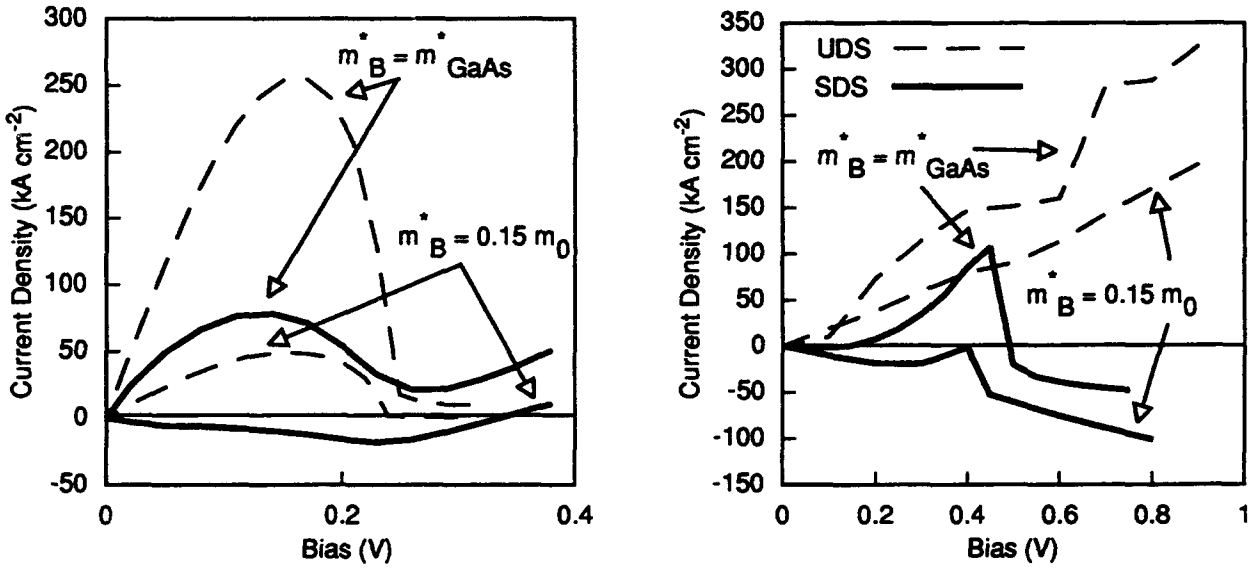


Figure 2: Conduction curves for AlAs/GaAs diode a) heavy lines: flat-band Wigner curves using first order upwinding. Schrödinger curves are shown as light dashed lines. b) self-consistent curves. The heavy lines are Wigner curves using second order upwinding (SDS), the light dashed lines using first order upwinding (UDS). Collisions are ignored. $N_q = 268$, $\Delta = a$, $N_k = 128$ in the Wigner calculations. The results are far from being satisfactory. Increasing N_k to 256 led to similar results. $\Delta = a/2$, $N_q = 536$, $N_k = 256$ does not help either.

closest to the barriers, 100 Å, $5x10^{16}$ cm^{-3} n-type GaAs and 100 Å, $6x10^{17}$ cm^{-3} n-type GaAs. The contact regions are $4x10^{18}$ cm^{-3} n-type GaAs. Figure 2a shows the Wigner conduction curves using

first order upwinding under flat-band conditions. For comparison the Schrödinger results are also shown. Figure 2b shows the self-consistent (potential is self-consistent to within 10^{-4}eV) Wigner conduction curves obtained by using first and second order upwinding. Flat-band, constant mass calculations have been reported for InGaAs/AlAs diodes using first order upwinding [8] and as can be seen from fig. 2a, under such conditions, the problems with the approach are not obvious.

To improve the fidelity of the numerical model to the exact equation, we use $|\delta| = 0.1$ in eq. 3. The resulting curves are shown in fig. 3a. The improvement over the other approaches is remarkable. For comparison, the Schrödinger-Poisson curves are also shown. Finally, including collisions in the relaxation time approach (τ =100fs) and using $|\delta| = 0.01$, the simulated and measur-

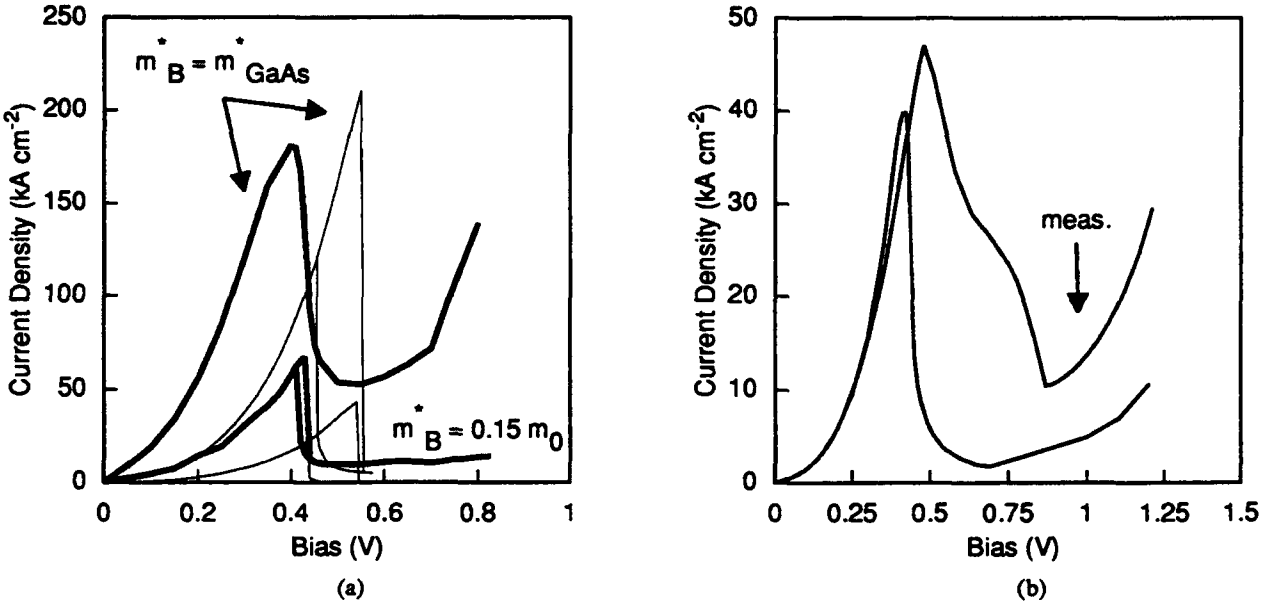


Figure 3: Improved Wigner conduction curves for AlAs/GaAs diode. a) heavy lines show the Wigner-Poisson results (collision free, $|\delta| = 0.1$) and the light lines are due to the Schrödinger-Poisson model. b) τ = 100fs, $|\delta| = 0.01$. $N_q = 268$, $\Delta = a$, $N_k = 128$. Also shown is a typical measured curve (300K) for our baseline.

ed curves are compared in fig. 3b. The poor agreement beyond the peak is an unresolved problem and has been the subject of intense discussions.

In conclusion, self-consistent steady state conduction curves for GaAs/AlAs resonant-tunneling diodes have been presented for the first time. The inclusion of the higher effective-mass in AlAs is essential. A new approach to obtaining meaningful conduction curves has been proposed and leads to much improved results.

Calculations were done on IBM RS/6000 models 590 and 320H. This work was sponsored in part by the Joint Services Electronics Program under Grant No. AFOSR 49620-92-C-0027, and by the Air Force Office of Scientific Research under AASERT Grant No. AFOSR F49620-93-1-0479.

References

1. H. Tsuchiya, M. Ogawa, and T. Miyoshi, IEEE Trans. Electron Devices, **38**, 1246 (1991).
2. S. R. De Groot, and L. G. Suttorp, *Foundations of Electrodynamics*. North-Holland, 1972.
3. F. A. Buot, and K. L. Jensen, Physical Review, **B 42**, 9429 (1990).
4. W. R. Frensley, Reviews of Modern Physics, **62**, 745 (1990).
5. D. H. Chow, J. N. Schulman, E. Ozbay, and D. M. Bloom, Appl. Phys. Lett., **61**, 1685 (1992).
6. V. K. Reddy, A. J. Tsao, and D. P. Neikirk, Electronics Lett., **26**, 1742 (21 1990).
7. T. P. E. Broekaert, and C. G. Fonstad, IEDM Technical Digest, 559 1989).
8. R. K. Mains, and G. I. Haddad, Journal of Applied Physics, **64**, 5041 (10 1988).

EVALUATION OF THE ELECTRON DENSITY OF STATES IN A SI-SiO_2 INTERFACE USING THE ZERO-TEMPERATURE GREEN'S FUNCTION FORMALISM[1]

Dragica Vasileska-Kafedziska, Paolo Bordone[2] and David K. Ferry
Center for Solid State Electronics Research, Arizona State University
Tempe, Az, 85287-6206, USA

Abstract

We develop the zero-temperature Green's function formalism to study transport in Si-SiO_2 inversion layer subject to both impurity and surface-roughness scattering. Surface-roughness is treated as a random potential scattering with a Gaussian correlation function. For the sake of simplicity, we assume that the electrons are scattered by randomly located but identical δ-function impurity potentials. The position of the subband minima and the electron concentration have been obtained by the self-consistent solution of the Poisson, Schrödinger and Dyson equations for each value of the effective transverse electric field. We give the analytical expression for the broadening of the electronic states in each subband, and the expression for the conductivity that includes the correction due to the normal particle-hole ladder diagram. In addition, the numerical results for the density of states function (DOS) for various values of the effective field are given. Finally, we present the numerical results for the mobility for various fitting parameters. The results for the mobility are in agreement with the experimental results of Kawaji obtained at 4.2 K in the region where surface-roughness dominates the transport properties of the system.

I. INTRODUCTION

We study transport properties of a (100) Si-inversion layer at zero temperature. We also give the results of the numerical self-consistent calculations for the density of states function, electron density and mobility for various fitting parameters and different effective fields. The dependence of mobility on the electron concentration N_s provides information for the strength of the considered dissipative mechanisms.

Our calculations are based on two major approximations. We assume that the effective-mass approximation is valid, so that we can use the effective masses and the dielectric constants of the perfect crystal. We also assume that the envelope functions for the inversion-electrons that satisfy the one-dimensional Schrödinger-wave equation vanish in the oxide. This is a valid assumption for moderately high surface fields. At very high surface fields, the wavefunction of the first subband extends less than 1 nm in the semiconductor and in this case the approximation probably fails.

Transport properties of Si-inversion layers at low temperatures are dominated by the elastic processes such as impurity and surface-roughness scattering. Surface-roughness is important only at high effective fields, where most of the inversion-electrons are trapped in the lowest subband.

The impurities are described by a random potential $u(\mathbf{R})$ with zero mean value and a correlator

$$\langle u(\mathbf{R})u(\mathbf{R}')\rangle = n_i U_o^2 \delta(\mathbf{R}-\mathbf{R}'). \qquad (1)$$

[1]Work supported in part by ONR.

[2] On leave from: Dipartimento di Fisica ed Instituto Nazionale di Fisica della Materia, Universita` di Modena, Via Campi 213/A, 41100 Modena, Italy.

where n_i is the impurity concentration. The <...> denote averaging over all impurity configurations. The strength of the impurity scattering is described through the constant U_o, equal to the matrix element for scattering from a single impurity.

Surface-roughness is introduced through a random local-potential term, proportional to the linear term of the Taylor expansion of the surface potential, of the form [1]

$$H_{sr}(\mathbf{R}) = f(\mathbf{r}) e E_s \quad , \tag{2}$$

where E_s is the surface field. The random function $f(\mathbf{r})$ that describes the deviation from the atomically flat surface is described by a two parameter Gaussian model, with autocorrelation function of the form

$$W_{sr}(|\mathbf{r}-\mathbf{r}'|) = \Delta^2 \exp\left[-\frac{|\mathbf{r}-\mathbf{r}'|^2}{\zeta^2}\right] . \tag{3}$$

Parameters Δ and ζ characterize the root-mean-square height of the bumps on the surface and the roughness correlation length , respectively.

From the coupled Dyson's equations for the retarded Green's function, we find that, within the diagonal approximation, the broadening of the electronic states for the n-th subband is obtained as a solution of the equation

$$\Gamma_n(\varepsilon_k, \varepsilon_F) = \frac{m^*}{4\pi\hbar^2} \sum_m \int_0^\infty d\varepsilon_q a_m(\varepsilon_q, \varepsilon_F) \times$$
$$\times \left\{ n_i U_o^2 O_{nm} + \delta_{nm} \pi (e E_s \zeta \Delta)^2 I_o\left(\frac{m^* \zeta^2}{\hbar^2}\sqrt{\varepsilon_k \varepsilon_q}\right) \exp\left[-\frac{m^* \zeta^2}{2\hbar^2}(\varepsilon_k + \varepsilon_q)\right] \right\} . \tag{4}$$

I_o is the modified Bessel function of the zeroth order, $a_m(\varepsilon_q, \varepsilon_F)$ is the spectral density function for the m-th subband, ε_F is the Fermi energy, ε_q is the kinetic energy and O_{nm} is the overlap factor. The details of this derivation are given in [2].

Within linear response, the expression for the conductivity can be summarized as

$$\sigma = \frac{e^2}{2\pi h} \sum_n (\varepsilon_F - \varepsilon_n) \int_0^\infty d\varepsilon_k \frac{a_n(\varepsilon_k, \varepsilon_F)}{\Gamma_n(\varepsilon_k, \varepsilon_F)} + \frac{e^2}{2\pi h} \sum_n (\varepsilon_F - \varepsilon_n) \frac{V_n}{1 - V_n} \int_0^\infty d\varepsilon_k \frac{a_n(\varepsilon_k, \varepsilon_F)}{\Gamma_n(\varepsilon_k, \varepsilon_F)} \tag{5}$$

where

$$V_n = \frac{m^*}{4\hbar^2} (e E_s \zeta \Delta)^2 e^{\frac{m^* \zeta^2}{2\hbar^2}(\varepsilon_F - \varepsilon_n)} \int_0^\infty d\varepsilon_q \frac{a_n(\varepsilon_q, \varepsilon_F)}{\Gamma_n(\varepsilon_q, \varepsilon_F)} e^{\frac{m^* \zeta^2}{2\hbar^2}\varepsilon_q} I_1\left[\frac{m^* \zeta^2}{\hbar^2}\sqrt{(\varepsilon_F - \varepsilon_n)\varepsilon_q}\right] . \tag{6}$$

I_1 is modified Bessel function of first order and ε_n is the subband energy. The first term on the right-hand side of (5) represents the Drude result. The second term gives the correction to the Drude conductivity due to the normal particle-hole ladder diagrams, as explained in [2]. This term yields a replacement of the relaxation time by a transport lifetime for the conduction electrons [3].

II. SIMULATION RESULTS

The self-consistent calculation of the coupled Schrödinger and Poisson equations gives the solution for the broadening of the electronic states according to (4). This is then used for the calculation of the density of states function and electron density. The process starts with an initial estimate for the potential energy profile and then solves all of the forementioned equations successively [4-6]. For the numerical solution of the Schrödinger equation, we have applied the Numerov algorithm [7], which is one order of magnitude more accurate that the fourth-order Runge-Kutta method. The matching tolerance for the wavefunctions was taken to be 10^{-5} . Finite-difference methods were used for the solution of the Poisson equation. We have used Gauss-Legendre integration for the energy integrations in (4-6) to speed the computation and decrease round-off errors.

The potential energy profile is given in Fig. 1. The Fermi energy corresponds to the zero-energy level on the figure. The electric field in the oxide is $2x10^6$ V/cm. The wave-functions for the first two subbands for the lowest valleys are shown in the insert. The corresponding electric field profile, is given in Fig. 2. The spectral density function due to impurities and surface-roughness scattering is presented in Fig. 3. The fitting parameters for surface roughness are: $\Delta = 0.2nm$ and $\zeta = 1.3nm$. In the numerical simulation for the broadening of the electronic states, instead of the surface field, we have used the average field that is felt by the electrons. The field in the oxide is the same as in Figs. 1-2. In Fig. 4, we present the form of the DOS function for various oxide fields. The fitting parameters for surface-roughness are the same as above. Due to quantum-size effects, we observe a change in the slope of the DOS curves near the subband threshold. This effect is more pronounced at higher electric fields, where surface-roughness dominates the transport properties of the system. The shift in the subband energies is due to the increase of the oxide field. The mobility curve, as a function of the inversion charge density, is given in Fig. 5 . The dots represent the experimental results obtained by Kawaji at 4.2 K [4]. In the region where surface-roughness dominates the transport (high inversion charge concentration), we achieve very good agreement with the experimental data. In the other region, the fitting failed because of the assumed simplified model for impurity scattering. In order to improve the results, we need to consider the Coulomb interaction properly. From the results presented in Fig. 6, we can deduce the relationship between the average field and the inversion density. We calculate that the average electric field varies as $E_{av}=e\,(0.52N_s+N_{depl})/(\varepsilon_o\varepsilon_{sc})$.

References

[1] S. M. Goodnick, D. K. Ferry, and C. W. Wilmsen, Phys. Rev. B **32**, 8171 (1985).
[2] D. Vasileska-Kafedziska, P. Bordone, and D. K. Ferry, submitted for publication.
[3] J. Bergmann, Physics Reports **107**, No. 1 (1984), p.1 (North-Holland, Amsterdam).
[4] T. Ando, A. B. Fowler, and F. Stern, Reviews of Modern Physics **54**, 437 (1982).
[5] F. Stern, Phys. Rev. B **5**, 4891 (1972).
[6] B. Vinter, Appl. Phys. Lett. **44**, 307 (1984).
[7] S. E. Koonin and D. C. Meredith, *Computational Physics* (Addison Wesley, New York, 1990).

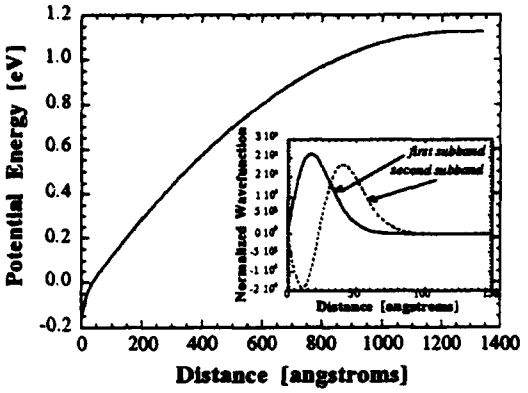


Fig.1 Potential Energy Profile

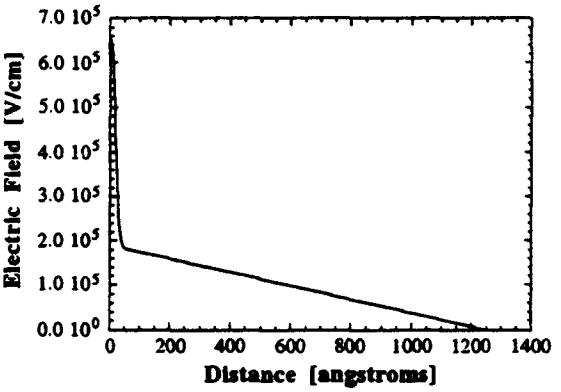


Fig.2 Electric field profile

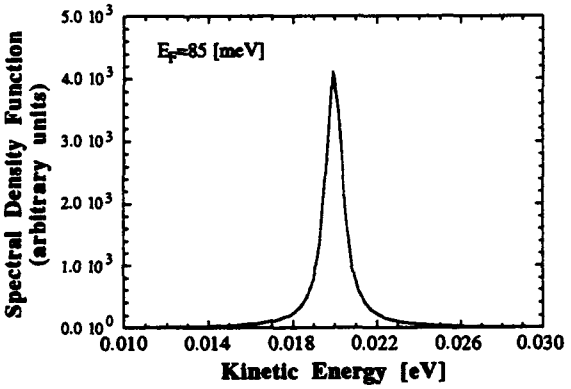


Fig.3 Spectral density function for the first subband vs kinetic energy

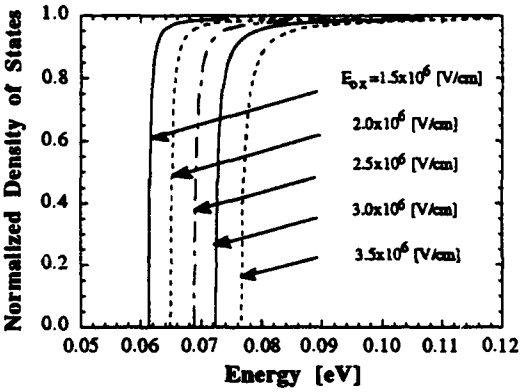


Fig.4 DOS vs Fermi energy for various oxide fields

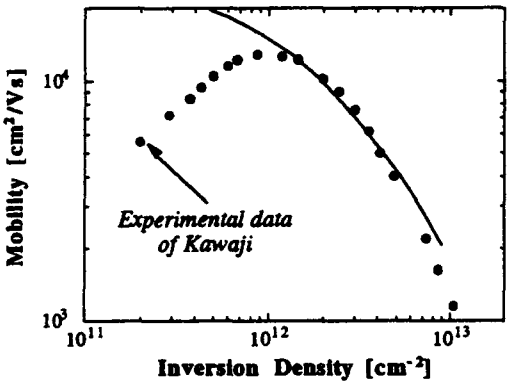


Fig.5 Calculated mobility vs inversion charge density

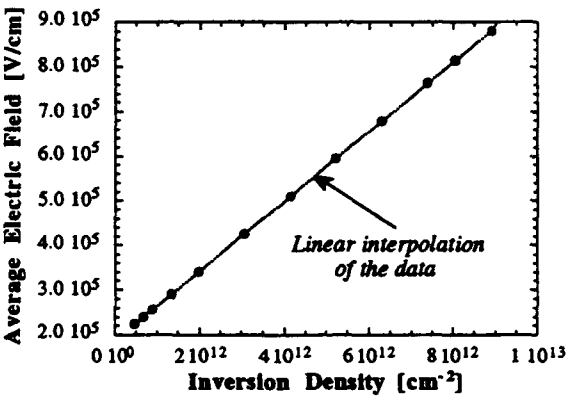


Fig.6 Average electric field vs inversion charge density

RESONANT TUNNELING CALCULATIONS VIA THE DENSITY MATRIX IN THE COORDINATE REPRESENTATION

H. L. Grubin and T. R. Govindan
Scientific Research Associates, Inc.
PO Box 1058
Glastonbury CT 06033

ABSTRACT

Solutions of the quantum Liouville equation in the coordinate representation, including dissipation, have be implemented for studying the double barrier resonant tunneling diode.

I. INTRODUCTION

Simulations of the quantum Liouville equation in the coordinate representation have been obtained for resonant tunneling structures. The coordinate representation equation includes dissipation represented via a quasi Fermi level. The two relevant equations (apart from Poisson's equation) are the equation of motion for the density matrix $\rho(x,x',t)$:

$$(1) \qquad i\hbar\frac{\partial\rho(x,x',t)}{\partial t}=-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial x'^2}\right)\rho(x,x',t)+\left[(V(x)-V(x'))-(E_F(x)-E_F(x'))\right]\rho(x,x',t)$$

and the equation constraining the quasi-Fermi level, the current density j, the position dependent scattering rate, $\Gamma(x)$, and the density $\rho(x)=\rho(x,x)$:

$$(2) \qquad E_F(x)-E_F(x')=-j\int_{x'}^{x}dx''m\Gamma(x'')/\rho(x'')$$

Each of these equations has been discussed in recent publications [1 ,2]. In particular, the algorithm used to solve these equations was discussed in [2]. Recent improvements summarized below have resulted in greater robustness and enable some of the calculations of this paper.

II. THE RESONANT TUNNELING STRUCTURE

The application of equations (1), (2) and Poisson's equation is to resonant tunneling structures. We treat a 200nm structure, with two 5 nm - 300 mev barriers separated by a 5nm well. The structure has a nominal doping of $10^{24}/m^3$ except for a central 50nm wide region where the doping is reduced to $10^{22}/m^3$. The effective mass is constant and equal to that of GaAs ($0.067m_0$); Fermi statistics are imposed; the ambient is 77K; and current is imposed through the density matrix equivalent of a displaced distribution at the boundaries (see [2]).

The signature of the RTD is it's current-voltage relation with the region of negative differential conductivity; for the structure considered this is displayed in figure 1. The current is numerically negligible until a bias of approximately 50 mev, with the peak current occurring at 260 mev, followed by a sharp but

modest drop in current at 270 mev. The interpretation of these results is assisted by figures (2) and (3) and the Bohm quantum potential:

$$(3) \qquad Q = -\frac{\hbar^2}{2m}\frac{\partial^2\sqrt{\rho(x)}}{\sqrt{\rho(x)}\,\partial x^2}$$

We have found, through an extensive number of numerical simulations, that the value of $V(x)+Q(x)$, between the barriers of an RTD is a measure of the position of the quasi-bound state.

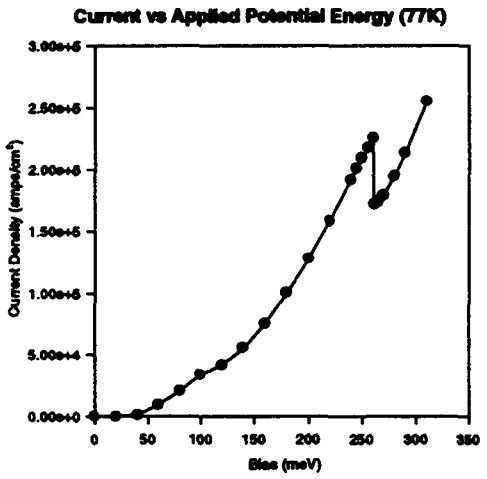


Figure 1. Current versus voltage for the resonant tunneling structure.

Consider figure 2 which displays the equilibrium self-consistent potential for the RTD. Also shown is the value of the equilibrium Fermi energy (approximately 54 mev) and the values, at five different values of applied potential energy, of $V(x)+Q(x)$ within the quantum well. At 100 mev the quasi-bound state is approximately equal to the equilibrium Fermi energy and significant current begins to flow. The current continues to increase until the bias equals 260 mev, where there is a sudden drop in current.

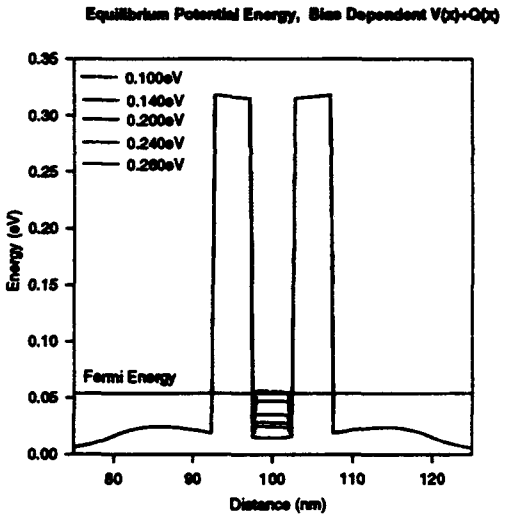


Figure 2. Equilibrium potential energy and the bias dependence of $V(x)+Q(x)$ within the quantum well. Legend denotes collector bias.

To see what is happening we blow up the region on either side of the emitter barrier, where we display values of $V(x)+Q(x)$ before the emitter barrier and within the quantum well (figure 3). Within the quantum well we see the quasi bound state decreasing as the bias on the collector is increasing. In the region prior to the emitter barrier where a 'notch' potential forms signifying charge accumulation, we see the formation with increased bias of a region where $V(x)+Q(x)$ is relatively flat. Of significance here is that for values of bias associated with the initial current increase the value of $V(x)+Q(x)$ within the quantum well is greater then its value in the emitter region. The current reaches a maximum at the cross-over where $V(x)+Q(x)$ in the emitter region and in the quantum well are approximately equal. (Implementation of an earlier algorithm, generally resulted in solutions oscillating between high and low values of current when this condition was reached). While it is tempting to associate $V(x)+Q(x)$ within the emitter region with a quasi-bound state, this association may be premature.

The distribution of potential energy $V(x)$ as a function of bias is displayed in figure 4, where the notch potential is deepened with increasing bias, signifying increased charge accumulation. This is accompanied by a smaller share of the potential drop across the emitter barrier, relative to the collector

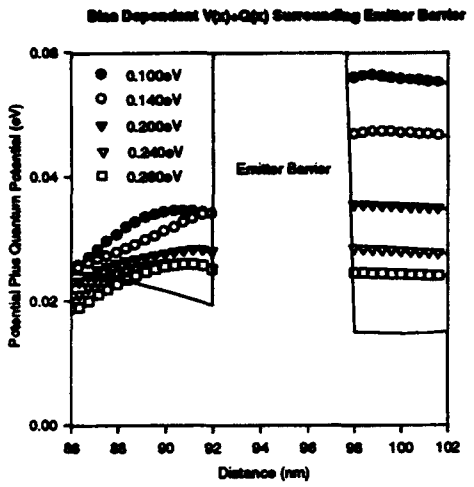


Figure 3. Blow up of figure 2 in the region surrounding the emitter barrier.

Explicit in this calculation is dissipation which is incorporated through the quasi-Fermi level. Within the vicinity of the boundaries the quasi-Fermi level is parallel to the conduction band edge. Indeed, for this calculation the quasi-Fermi level departs from the conduction band edge only within the vicinity of the barriers. The quasi-Fermi level is displayed in figure 5 at a bias of 260 mev, where we see that the quasi-Fermi level is relatively flat until the middle of the first barrier at which point there is a small drop in value followed by a flat region within the quantum well. There is a strong drop of the quasi Fermi level within the second barrier.

The charge distribution accompanying these variations in bias shows accumulation on the emitter side of the barrier along with charge accumulation within the quantum well. The increase in charge within the quantum well and adjacent to the emitter region is accompanying by charge depletion downstream of the second barrier, with the result that the net charge distribution throughout the structure is zero.

In all of the computations associated with figure (1), only one set of scattering rates was used. Variations in the quasi Fermi level were accompanied by variations in density and current which were all obtained in a self-consistent manner. Supplemental computations were performed in which the quasi-Fermi level was varied by altering the scattering rates. The calculations were applied to the post threshold case with values for the scattering rate

barrier region. In particular, comparing the slopes of the voltage drop across the emitter and collector barriers, it is apparent that larger fractions of potential energy fall across the collector barrier.

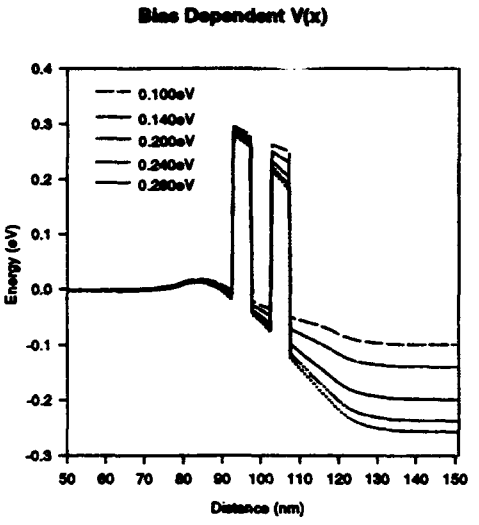


Figure 4. Potential energy V(x) as a function of collector bias

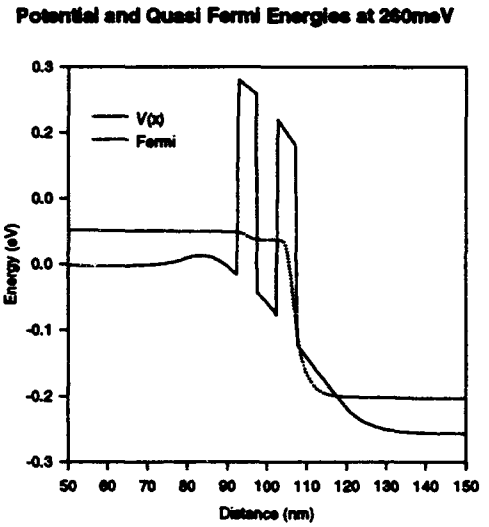


Figure 5 Potential and quasi-Fermi energy at a bias of 260 mev.

chosen so to provide a large drop in current. Indeed a current drop by greater than a factor of three was obtained followed by a shallow current increase with increasing bias. The significant difference leading to these changes was the manner in which the quasi-Fermi level changed. Rather than the shallow change depicted in figure 5, there was a larger change in the quasi-Fermi level across the first barrier (figure 6), a result similar to that obtained for single barriers [1].

III COMMENTS ON THE ALGORITHM

The calculations discussed in this paper were obtained from a new solution algorithm that was constructed for the quantum Liouville equation and permits a more convenient specification of boundary conditions, in particular when the device is under bias. The algorithm is based on a reformulation of the governing equations in which a higher order differential equation in the local direction $[(x+x')/2]$ is constructed from the quantum Liouville equation. The reformulated equation behaves like an elliptical equation in the local direction rather than the hyperbolic behavior of the quantum Liouville equation. With appropriate boundary conditions, solutions to the two forms of the quantum Liouville equations are equivalent. However the reformulated equation allows construction of a more robust algorithm that provides desired solution behavior at the contacts by boundary condition specification at both contacts.

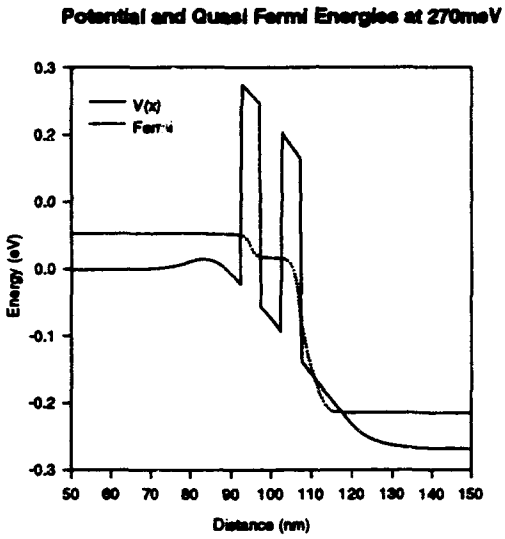


Figure 6. As in figure 5 but for enhanced scattering.

IV SUMMARY

The Liouville equation in the coordinate representation has been implemented for studying resonant tunneling structures. The results provide the first explicit relationship between the range of bias prior to the drop in current and the movement of the quasi-bound states. Additionally, the results provide the first explicit connection between quasi-Fermi levels and the magnitude of the peak to valley ratio in RTDs, and provide evidence that the behavior of RTDs is strongly controlled by dissipation.

V. ACKNOWLEDGMENT

This authors are grateful for the support of ONR and ARO.

VI. REFERENCES

1. H. L. Grubin, T. R. Govindan and M.A. Stroscio, *Semicond. Sci. Technol.* To be published (1994); and D. K. Ferry and H. L. Grubin, To be published in *Solid State Physics*, (H. Ehrenreich, ed) (1994).
2. H. L. Grubin, T. R. Govindan , J. P. Kreskovsky and M. A.Stroscio, *Solid State Electronics* **36**, 1697(1993).

QUANTUM TRANSPORT USING LIOUVILLEAN QUANTUM-FIELD DYNAMICS AND FUNCTIONAL APPROACH TO SELF-CONSISTENT MANY-BODY AND SCATTERING EFFECTS

F.A. Buot and A.K. Rajagopal
Naval Research Laboratory, Washington, D.C. 20375

Abstract

Quantum dynamics in Liouville space is used for discussing quantum transport in nanoelectronics. A synergism between the two formalisms treated here, namely, the "super" Green's function technique and the generalized functional approach is expected to pave the way towards more accurate self-consistent numerical calculations of many-body and scattering effects in nanolectronics and optoelectronics.

I. INTRODUCTION

There is a need for a self-consistent treatment of the nonlinear dynamics of interacting quantized fields, e.g., interactions between electrons, ions, and electromagnetic fields. These kinds of problems arise in high frequency and/or high power nanoelectronic and optoelectronic devices. So far, highly nonequilibrium situations are treated by assuming that all the subsystems other than the one of interest are behaving classically and/or in equilibrium condition. In this paper, non-equilibrium quantum transport is formulated in terms of the Liouville space (L-space) dynamics thus treating all the fields quantum mechanically on equal footing. This description also unifies classical and quantum statistical dynamics within the L-space dynamical framework [1].

The L-space formulation is equivalent to the nonequilibrium Green's function technique [2] originated by Schwinger, and Keldysh. However the major advantage of the L-space formalism is that it allows for a straightforward application of quantized field theoretical techniques since only real-time axis is used. This is in contrast to the double-time contour of the corresponding Hilbert space (H-space) formalism which leads to awkward calculational procedure for obtaining the self-energies of interest. Another major advantage of the L-space formalism is that it provides a common starting point for a many-body functional technique, which is rooted in the powerful density functional method for calculating many-body effects [3], and the real-time Green's function technique based on the Φ-derivable method [4] for the self-energies. The synergism of these two independent techniques is expected to yield a more powerful many-body functional technique for numerically simulating *bonafide* scattering effects.

For simplicity of presentation in what follows, we focus our discussion on the electron system. The corresponding discussion of the ions and electromagnetic fields involve a parallel treatment, which will be discussed elsewhere.

II. QUANTUM DYNAMICS IN LIOUVILLE SPACE

In Liouville space quantum dynamics, the density-matrix equation formulation of the quantum statistical dynamics in H-space becomes a dynamical equation in L-space defined as follows:

$$i\hbar\frac{\partial}{\partial t}|\rho\rangle\rangle = \mathscr{L}|\rho\rangle\rangle, \tag{1}$$

where ρ is the density matrix of the system in H-space, and $|\rho\rangle\rangle$ is its corresponding supervector in L-space. Note that $\mathscr{L}|\rho\rangle\rangle$ corresponds to the commutator $[\mathscr{H},\rho]$ in H-space. Here the Liouvillean $\mathscr{L}=\hat{\mathscr{H}}-\tilde{\mathscr{H}}$, where the "hat" and "tilde" superoperators are defined below. If the set $\{|n\rangle\}$ is an orthonormal basis in H-space in the number representation of the many-body states, then the corresponding set $\{\{||m\rangle\langle n|\rangle\rangle\}\}$ is an orthonormal basis supervectors in L-space. The annihilation and the creation quantum-field operators, $\psi, \psi^\dagger$ in H-space become the "hat"("tilde") annihilation, $\hat{\psi}\,(\tilde{\psi})$ and creation, $\hat{\psi}^\dagger\,(\tilde{\psi}^\dagger)$ operators in L-space. They are defined as follows:

$$\hat{\psi}^\dagger||m\rangle\langle n|\rangle\rangle=|\psi^\dagger|m\rangle\langle n|\rangle\rangle, \tag{2}$$

$$\tilde{\psi}^\dagger||m\rangle\langle n|\rangle\rangle=(-\sigma)^{m-n+1}||m\rangle\langle n|\psi\rangle\rangle, \tag{3}$$

$$\hat{\psi}||m\rangle\langle n|\rangle\rangle=|\psi|m\rangle\langle n|\rangle\rangle, \tag{4}$$

$$\tilde{\psi}||m\rangle\langle n|\rangle\rangle=(-\sigma)^{m-n}||m\rangle\langle n|\psi^\dagger\rangle\rangle, \tag{5}$$

where σ is -1 for bosons and +1 for fermions. Due to the doubling of operators in L-space corresponding to each operator in H-space, it is more convenient to introduce a two-component annihilation and creation operators in L-space. For fermions, we have

$$\Psi=\begin{pmatrix}\hat{\psi}\\ \tilde{\psi}^\dagger\end{pmatrix} \quad and \quad \Psi^\dagger=\begin{pmatrix}\hat{\psi}^\dagger & \tilde{\psi}\end{pmatrix}. \tag{6}$$

We also define a unit supervector as $|1\rangle\rangle=\sum_m|m\rangle\langle m|$, so that the average of an arbitrary operator A can be written as $\langle A\rangle=Tr\,\rho A=\langle\langle 1|\hat{A}|\rho\rangle\rangle$.

III. NON-EQUILIBRIUM GREEN'S FUNCTION IN L- SPACE DYNAMICS

A "super" or non-equilibrium Green's function in L-space is defined as

$$\mathscr{G}=\left\langle T\Psi_{\bar{H}}(t)\Psi^\dagger_{\bar{H}}(t')\right\rangle\!\!\rangle/i\hbar, \tag{7}$$

where T is the usual time ordering operator. In the above expressions, the superoperators are written in the "super"-Heisenberg representation, e.g.,

$$\Psi_{\bar{H}}(t,t_o)=\mathscr{U}(t_o,t)\,\Psi\,\mathscr{U}(t,t_o), \tag{8}$$

where $\mathscr{U}(t,t_o)=T\exp\left\{-\frac{i}{\hbar}\int_{t_o}^{t}\mathscr{L}\,dt'\right\}$. Thus, in the "super"-interaction representation in L-space, we can also write Eq. (7) as

$$\mathscr{G}=\langle\langle 1|\bar{S}(\infty,t)\Psi_{\bar{I}}(t)\bar{S}(t,t')\Psi^\dagger_{\bar{I}}(t')\bar{S}(t',-\infty)|\rho_{eq}\rangle\rangle\big/\langle\langle 1|\bar{S}(\infty,-\infty)|\rho_{eq}\rangle\rangle, \tag{9}$$

where $\bar{S}(t,t_o)$ is the "super"-S-matrix, obtained by substituting $\mathscr{L}$ in $\mathscr{U}(t,t_o)$ by $\mathscr{L}_I^{(1)}=\hat{\mathscr{H}}_I^{(1)}-\tilde{\mathscr{H}}_I^{(1)}$ in the "super"-interaction picture. In Eq. (9), the time axis is from $-\infty$ *to* $+\infty$ and therefore the theory is formally the same as for the "zero-temperature" Green's function. Similar equations for the respective Green's functions can be constructed for the ion and electromagnetic fields. The full dynamics of ion motion including phonons is described by the correlation function of the ion positions.

We will now develop the transport equations in L-space from the above Green's functions. Following the Keldysh approach in H-space, the transport equations for nonequilibrium plasmas and radiation has been given by DuBois [5]. A similar transport equation for a system of ions may be found in Kwok [6], which is based on the Green function associated with ion positions. In a

separate paper [7], we will derive the appropriate transport equations for the coupled system of electrons, ions, and electromagnetic fields.

IV. TRANSPORT EQUATIONS AND SELF-ENERGIES

In terms of the familiar correlation functions, $G^{>}$,and $G^{<}$, the matrix equation for the "super"-Green's function, $\mathcal{G}$, is exactly the same as the following expression

$$\mathcal{G} = \begin{pmatrix} G^{c} & -G^{<} \\ G^{>} & -G^{ac} \end{pmatrix}, \tag{10}$$

where G^{c} and G^{ac}, which can be expressed in terms of $G^{>}$ and $G^{<}$, are the chronological and antichronological Green's functions respectively. Equation (10) is exactly the same as the nonequilibrium matrix Green's function expression obtained by other authors [2], using the time contour formulation of Schwinger and Keldysh. Integro-differential transport equations for the matrix elements of $\mathcal{G}$, can be readily obtained from $\mathcal{G}^{-1}\mathcal{G} = \delta$ and its adjoint. We make use of the relations: $F^{(,)\dagger} = -F^{(,)}$, $F^{c\dagger} = -F^{ac}$, to obtain the transport equations for all the matrix elements of $\mathcal{G}$:

$$\begin{aligned} i\hbar(\partial/\partial t + \partial/\partial t')G^{>,<} = &\left[-\hbar^2\nabla^2/2m + \varphi_{eff} + \mathrm{Re}\,\Sigma^{r}, G^{>,<}\right] \\ &+\left[\Sigma^{>,<}, \mathrm{Re}\,G^{r}\right] + i\left\{A, \Sigma^{>,<}\right\}/2 - i\left\{\Gamma, G^{>,<}\right\}/2, \end{aligned} \tag{11}$$

$$i\hbar(\partial/\partial t + \partial/\partial t')G^{c} = \left[-\hbar^2\nabla^2/2m + \varphi_{eff} + \Sigma^{c}, G^{c}\right] + G^{<}\Sigma^{>} - \Sigma^{<}G^{>}, \tag{12}$$

$$i\hbar(\partial/\partial t + \partial/\partial t')G^{ac} = \left[-\hbar^2\nabla^2/2m + \varphi_{eff} - \Sigma^{ac}, G^{ac}\right] + \Sigma^{>}G^{<} - G^{>}\Sigma^{<}, \tag{13}$$

where G^{r} *and* Σ^{r} represent the retarded Green's function and its associated self-energy, and φ_{eff} is the effective potential. It is clear from the last two equations that the term $G^{<}\Sigma^{>} - \Sigma^{<}G^{>}$ and its counterpart describe effects beyond the finite-lifetime quasi-particle concept, and represent *bonafide* nonequilibrium scattering effects. These are similar to those occurring in the last two terms of Eq. (11) for $G^{>,<}$. The equation for $G^{>,<}$ is exactly identical to the Keldysh results [2], while the equations for G^{c} and G^{ac} also contain collision terms.

The "super" self-energy has formally the same functional form as that of the "zero-temperature" self-energy. In the L-space approach, each of the self-energy matrix elements is calculated using the equation of motion of the "hat" and "tilde" superoperators, which is a straightforward application of quantum field theoretical techniques. Similar transport equations are deduced [7] for the ion and electromagnetic fields from their respective Green functions and self-energies. The self-energies depend on all the field variables exhibiting the mutual interactions among the fields. Thus all the Green's functions become mutually coupled, requiring thereby a self-consistent analysis.

V. FUNCTIONAL TECHNIQUE IN L-SPACE DYNAMICS

The stationary action principle is the foundation of the time-dependent density functional theory of pure-state quantum mechanical systems. The "Schrodinger" Eq. (1), also provides a stationary action principle for nonequilibrium statistical mechanics.

We write the functional of the action in the form

$$W(t,t_o) = \frac{1}{2}\int_{t_o}^{t}\left\langle\left\langle \Phi(t')\left| i\hbar\frac{\partial}{\partial t} - \mathcal{L}\right|\rho(t')\right\rangle\right\rangle dt', \tag{14}$$

subject to the thermal equilibrium initial condition for the ρ. Thus by varying the left supervector and setting the result equal to zero, we obtain the superket Eq. (1). The factor 1/2 is chosen to account for the presence of "twins" (doubling) in L-space.

It is shown elsewhere [8], that $W(t,t_o)$ is a functional of averaged fields and they completely characterize the supervector $|\rho\rangle\rangle$. We use the "physical" functional given by $\langle\langle\Phi(t)| = \langle\langle 1|$ in Eq.(14) which is now stationary with respect to the variations of the average currents $J_\mu(\vec{r}t)$, average electromagnetic potentials $A_\mu(\vec{r}t)$, and average ion positions $\vec{R}(l\kappa t)$, where l represents the lattice point and κ labels the ion species, for a system of electrons, ions, and electromagnetic fields. Thus the stationarity of W leads to the equations:

$$\delta W/\delta J_\mu = 0, \quad \delta W/\delta A_\mu = 0, \quad \textit{and} \quad \delta W/\delta \vec{R} = 0. \tag{15}$$

The first of the equations in Eq.(15) leads to an effective one-particle Schrodinger equation, the second leads to an effective Maxwell's equation, and the third leads to an effective Newton's equation for the ions. In general, these equations involve "effective potentials", equal to the average potentials in addition to terms describing the mutual interactions with other fields.

By re-expressing these equations in terms of the Green function language, we can identify the terms corresponding to self-energies of the respective fields, given by the "super" Green function approach. Thus, we can incorporate the functional form of the appropriate diagrammatic expressions into a self-consistent scheme within the functional approach for calculating many-body effects which now includes the effects of scatterings. These are discussed in more detail by the authors in a separate paper [8].

VI. SUMMARY AND CONCLUDING REMARKS

The functional theory discussed here provides a self-consistent method for incorporating many-body and scattering effects in the self-energy to be used in the transport equations. By inserting the diagrammatic approximation to the self-energy in a self-consistent loop of the coupled equations in functional theory, a more accurate self-consistent self-energy can be generated. This algorithm may be considered as a generalization in device physics of the well-known self-consistent method of solving the Poisson equation along with the quantum transport equation [2]. For numerical simulation, it is desirable to take the Weyl transform [2] of Eqs.(11-13, 15). It is hoped that this work would lead to the numerical implementation of the algorithm proposed here in self-consistent analyses of nanoelectronics and optoelectronics problems.

ACKNOWLEDGMENT: This work is supported in part by the Office of Naval Research.

REFERENCES

[1] M. Schmutz, Z. Physik **B30**, 97 (1978)
[2] F.A. Buot and K.L. Jensen, Phys. Rev. **B42,** 9429 (1990); see also F.A. Buot, Phys. Reports **234**, 73 (1993)
[3] R.M. Dreizler and E.K.U. Gross, *Density Functional Theory* (Springer Verlag, NY, 1990)
[4] G. Baym, Phys. Rev. **127**, 1391 (1962)
[5] D.F. Dubois, *Lectures in Theoretical Physics,* (Gordon and Breach, NY, 1967) 469
[6] P.C.K. Kwok, in *Solid State Physics* **20**, 213 (1967)
[7] F.A. Buot and A.K. Rajagopal, *Quantum Transport and Nonequilibrium Time-Dependent Functional Theory for Coupled Interacting Fields,* in preparation
[8] A.K. Rajagopal and F.A. Buot, *A Nonequilibrium Time-Dependent Functional Theory for Coupled Interacting Fields,* submitted for publication

QUANTUM TRANSPORT AND COHERENCE IN BOUNDARY LIMITED ELECTRONIC DEVICES: RECURSIVE DYSON HAMILTONIAN GREEN'S FUNCTION AND FINITE ELEMENT TECHNIQUES IN HETEROSTRUCTURES

Clifford M. Krowne
Microwave Technology Branch, Electronics Science & Technology Division, Naval Research Laboratory, Washington, DC 20375-5347

Abstract

With the decreasing scale of electronic devices, increasing attention is being paid to the finite sized nature of the resulting structures. This necessitates particular efforts to take into account the constraining as well as limiting behavior of the device boundaries. Methods to efficiently formulate theoretically these effects, and then to construct user friendly, flexible, time efficient computer codes are becoming very important. Such issues will be addressed here as we discuss ballistic transport in 1D Aharonov - Bohm rings using a discrete tight - binding Hamiltonian, the 3D quantum transmission into and out of a multiport cavity in the linear single - particle weak - like variational formulation, and the nature of the above techniques in view of general variational and weighted residual techniques possible for extremely general media found in microwave and electromagnetic problems.

I. TRANSPORT IN 1D AHARONOV-BOHM HETEROSTRUCTURE RINGS

Electron transport through quantum interference device-like structures is of interest where microstructure fabrication allows multiple channels of quantum wave flow to occur. This is possible to achieve in finite width heterostructure waveguides. For narrow width channels, the channels may be considered 1D chains and analyzed in discretized form using tight-binding functions [1]. Non-local vector potential effects will alter relative phase information in the electron wave functions and cause varying interference behavior at channel intersections. For wide channels this tight-binding Hamiltonian technique, which yields recursive Green's functions, may not be the best way to approach the 2D nature of the problem. Finite element techniques which can easily allow for finite wall locations and arbitrary geometries may be a much more general method based upon solving the suite of physical field and transport governing equations. Section II discusses that method appropriate for 3D problems.

An infinite chain is intercepted at sites 0 and (N + 1) by a finite loop, with N' sites. By choosing a symmetric gauge, the vector potential, $\mathbf{A}(\mathbf{r})$ is given by:

$$\mathbf{A}(\mathbf{r}) = \frac{1}{2}\mathbf{H}\times\mathbf{r} \tag{1}$$

By the use of Peierl's substitution [2], the hopping potential which appears in the tight-binding Hamiltonian, becomes (h is Planck's costant divided by 2π):

$$\hat{V}_{ij} \rightarrow V\exp\left(\frac{ie}{hc}\mathbf{A}(\mathbf{r}_i)\cdot\mathbf{r}_j\right) \tag{2a}$$

Upon substitution of (1) and (2), and using well known vector identies, we find:

$$\hat{V}_{ij} = V\exp\left(\frac{ie}{hc}\frac{1}{2}\mathbf{H}\times\mathbf{r}_i\cdot\mathbf{r}_j\right) = V\exp\left(\frac{ie}{2hc}\mathbf{H}\cdot\mathbf{r}_i\times\mathbf{r}_j\right) \tag{2b}$$

Here, V_{ij} represents the hopping matrix element in a nearest-neighbor tight-binding formulation, so that $V_{ij} \neq 0$ for $i = j \pm 1$, and $= 0$ otherwise. To compute the effect of the loop induced scattering on the transmission amplitude, we use Dyson's equation [3]. $G_0(n, m)$ represents the

$$G(\alpha, \beta) = G_0(\alpha, \beta) + G_0(\alpha, 0)\hat{V}(0, 1')G(1', \beta) + G_0(\alpha, N+1)\hat{V}(N+1, N')G(N', \beta) \quad (3)$$

α an β indicate respectively the extreme left and right of the infinite chain, primes denote the loop locations, and 0 and $N + 1$ indicate respectively the intersections of the infinite chain and loop at the left and right. Unlike the field-free case [3],

$$\hat{V}(i, j) = [\hat{V}(j, i)]^* \quad (4)$$

We start the derivation with the propagator for an infinite one-dimensional tight-binding chain [4]:

$$G(1, m; z = E) = \left[2V\sqrt{1 - \left(\frac{E}{2V}\right)^2}\right]^{-1}\left[\frac{E}{2V} + i\sqrt{1 - \left(\frac{E}{2V}\right)^2}\right]^{|1-m|} \quad (5)$$

where we have used $E = 2V\cos ka$ for a tight-binding band, the site energy E_0 being set arbitrarily to zero. Setting $ka = \theta$, we find $E/2V = \cos\theta$, which yields:

$$G_0(1, m; E) = \frac{i}{2V\sin\theta}(\cos\theta + i\sin\theta)^{|1-m|} = \frac{ie^{i|1-m|\theta}}{2V\sin\theta} = G(0, 0; E)e^{i|1-m|\theta} \quad (6)$$

For a tight-binding chain of atoms, we start with a single site and add sites iteratively through the Dyson equation. For a single site [4],

$$G_0(1, 1; E) = \frac{1}{E} = \frac{1}{2V\cos\theta} \quad (7)$$

For two sites, we write the basic Dyson equation (previously used to generate equation (3)), where G_0 = Green function for the single site, and G = Green function for the two sites, and $\hat{V}$ = the hopping matrix element between the two sites:

$$G = G_0 + G_0\hat{V}G \quad (8)$$

Then applying (8) to both the diagonal and off-diagonal matrix element:

$$G(1, 1) = G_0(1, 1) + G_0(1, 1)\hat{V}_{12}G(2, 1) \quad (9a)$$

$$G(2, 1) = 0 + G_0(2, 2)\hat{V}_{21}G(1, 1) \quad (9b)$$

Substituting (9b) into (9a), we obtain:

$$G(1, 1)[1 - G_0(1, 1)\hat{V}_{12}G_0(2, 2)\hat{V}_{21}] = G_0(1, 1)$$

$$G(1, 1) = \frac{G_0(1, 1)}{1 - V^2G_0(1, 1)G_0(2, 2)} = \frac{1}{2V\cos\theta}\frac{4\cos^4\theta}{4\cos^2\theta - 1} = \frac{\sin 2\theta}{V\sin 3\theta} \quad (10)$$

Notice that the phase factors found in the hopping matrix elements in (2) cancel out in the (10) expression. To find the off-diagonal matrix element $G(1, 2)$ we again write:

$$G(1, 2) = G_0(1, 1)\hat{V}_{12}G(2, 2) \quad (11a)$$

$$G(2, 2) = G_0(2, 2) + G_0(2, 2)\hat{V}_{21}G(1, 2) \quad (11b)$$

Substituting (11b) into (11a),

$$G(1, 2) = \frac{G_0(1, 1)\hat{V}_{12}G_0(2, 2)}{1 - V^2G_0(1, 1)G_0(2, 2)} = \frac{V\exp\left(\frac{ie}{2hc}\mathbf{H}\cdot\mathbf{r}_1\times\mathbf{r}_2\right)}{4V^2\cos^2\theta}\frac{1}{4\cos^2\theta - 1}4\cos^2\theta = e^{i\gamma(1, 2)}\frac{\sin\theta}{\sin 3\theta} \quad (12)$$

where the argument of the phase factor has been replaced by $\gamma(1, 2)$.

For finite chains of arbitrary length N, we use the results for one and two sites and apply a recursive Green's function technique [5] to generate further sites. Continuation of this process leads to finding the transmission amplitude.

The transmission amplitude, t, is defined by:

The transmission amplitude, t, is defined by:

$$t = \frac{G(\alpha, \beta)}{G_0(\alpha, \beta)} = \frac{\text{amplitude for propagation through whole structure}}{\text{amplitude without upper loop}} \tag{13}$$

while the transmission coefficient is,

$$T_{NN'}(E, \phi) = t \cdot t^* \tag{14}$$

where E is the energy of the incident electron and ϕ is the ratio of the flux enclosed by the loop to the flux quantum $\phi_0 = hc/e$. The final result for $T_{NN'}(E, \phi)$ is:

$$T_{NN'}(E, \phi) = 4\left[\cos^2\frac{\phi}{2}\sin^2\alpha\cos^2\beta + \sin^2\frac{\phi}{2}\sin^2\beta\cos^2\alpha\right]\times$$
$$\left\{\left[\sin(2\alpha) + \cot\theta(\cos^2\alpha - \cos^2\beta)\right]^2 + \left[\cos\phi + 1 + \cot\theta\sin(2\alpha) - \left(3 - \frac{1}{2\sin^2\theta}\right)\cos^2\alpha - \left(\frac{1}{2\sin^2\theta} - 1\right)\cos^2\beta\right]\right\}^{-1} \tag{15}$$

where

$$\alpha = \frac{N + N' + 2}{2}\theta \tag{16a}$$

$$\beta = \frac{N' - N}{2}\theta \tag{16b}$$

and where θ is related to the energy in a single tight-binding band:

$$E = 2V\cos\theta \tag{17}$$

It can be verified easily that (15) reproduces the result for the field-free case derived by Ginuea and Verges [3] when $\phi = 0$.

II. 2D AND 3D MODELLING OF ELECTRON WAVE PROPAGATION IN HETEROSTRUCTURE DEVICES USING FINITE ELEMENTS

The description starts with the Schrodinger equation in 3D

$$-\frac{h^2}{2m}\left(\hat{P} - \frac{e}{c}A\right)^2\Psi(x, y, z) + V(x, y, z)\Psi(x, y, z) = E\Psi(x, y, z) \tag{18}$$

where h is Planck's constant divided by 2π. For a uniform magnetic field, the gauge is nonunique which leads to a A choice. Following Wang et. al. [6], A = (- By/2, Bx/2) and

$$-\frac{h^2}{2m}\nabla^2\Psi(x, y, z) + \hat{V}(x, y, z)\Psi(x, y, z) = E\Psi(x, y, z)$$
$$\hat{V}(x, y, z) = V(x, y, z) - \frac{iehB}{2mc}\left(y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right) + \frac{e^2B^2}{8mc^2}(x^2 + y^2) \tag{19}$$

Performing the 3D integral over the device volume, with an input port on the left at $x = x_1$ and an output port at the bottom at $y = y_2$, and using Green's theorem,

$$\frac{h^2}{2m}\iiint_{\Omega_0}\nabla\tilde{\Psi}\cdot\nabla\Psi d\Omega + \iiint_{\Omega_0}\tilde{\Psi}(\hat{V} - E)\Psi d\Omega = \frac{h^2}{2m}\sum_{i=1}^{2}\oiint_{\Gamma_i}\tilde{\Psi}\nabla\Psi\cdot\hat{n}d\Gamma \tag{20}$$

Apply a finite element discretization in the cavity region.

$$\Psi(x, y, z) = N(x, y, z)U \quad ; \quad \tilde{\Psi}(x, y, z) = N(x, y, z)\tilde{U} \tag{21}$$

Here N(x, y, z) is the global shape function, U is a vector of all the unknown wave function Ψ values on the nodes, and U tilde the trial U. Then (20) becomes

$$\tilde{U}(T + V)U = \frac{h^2}{2m}\sum_{i=1}^{2}\oiint_{\Gamma_i}\tilde{\Psi}\nabla\Psi\cdot\hat{n}d\Gamma \tag{22}$$

The surface integral in (22) which includes a depth d in the z-direction, requires a minor modification to convert from a line to a surface integral. The result is

$$\tilde{U}(T + V)U = \tilde{U}_1P - \tilde{U}_1C^1U_1 - \tilde{U}_2C^2U_2 \tag{23}$$

where

$$P_i = -\frac{h^2}{2m}\oiint_{\Gamma_1}\phi_i^1(x, y, z)\left[\mathbf{h}^1(y, z) - \mathbf{h}^2(y, z)S^{-1}\tilde{S}\right]ad\Gamma \tag{24}$$

$$C_{ij}^1 = \frac{h^2}{2m}\oiint_{\Gamma_1}\phi_i^1(x, y, z)\sum_{m, n}h_m^2(y, z)S_{mn}^{-1}T_{nj}^1d\Gamma \tag{25}$$

$$C_{ij}^2 = \frac{h^2}{2m}\oiint_{\Gamma_2}\phi_i^2(x, y, z)\sum_{m, n}g_m(x, z)R_{mn}^{-1}T_{nj}^1d\Gamma \tag{26}$$

In (23), U_1 and U_2 only correspond respectively to the subset of nodes on port surfaces 1 and 2. Since P_i is known due to the specified incident wave amplitude vector **a**, we know that the partial global vectors U_1 and U_2 can be moved to the left hand side of (23) after proper assembly procedure [7] and [8]

$$\tilde{U}_1^TP = \tilde{U}^T\tilde{P} \quad ; \quad \tilde{U}_1^TC^1U_1 = \tilde{U}^T\tilde{C}^1U \quad ; \quad \tilde{U}_2^TC^2U_2 = \tilde{U}^T\tilde{C}^2U \tag{27}$$

as done in [9] and [10]. The solution follows immediately.

$$(T + V + \tilde{C}^1 + \tilde{C}^2)U = P \tag{28}$$

REFERENCES

[1] W. R. Grise, C. M. Krowne, F. A. Buot, "Ballistic Transport in 1 Dimensional Aharonov - Bohm Rings," to be published.

[2] R. Peierls, Z. Phys., "Zur Theorie Des Diamagnetismus Von Leitungselektronen, vol. 80, pp. 763-791, 1933.

[3] F. Guinea and J. Verges, Phys. Rev. B, "Localization and Topological Disorder," vol. 35 pp. 979-986, 1987.

[4] E. N. Economou, Green's Functions in Quantum Physics, 2 nd ed., Springer-Verlag, 1983.

[5] F. Sols, M. Macucci, U. Ravaioli, and K. Hess, J. Appl. Phys., "Theory for a Quantum Modulated Transistor," vol. 66, p. 3892-3906, 1989.

[6] Y. Wang, J. Wang, and H. Guo, "Magnetoconductance of a Stadium-Shaped Quantum Dot: A Finite-Element-Method Approach," Phys. Rev. B, vol. 49, pp. 1928-1934, 1994.

[7] C. M. Krowne, "Vector Variational and Weighted Residual Finite Element Procedures for Highly Anisotropic Media," IEEE Trans. Antennas and Propagation, vol. AP-42, in press, 1994.

[8] D. S. Burnett, Finite Element Analysis, Addison-Wesley, 1988.

[9] C. S. Lent and D. J. Kirkner, "The Quantum Transmitting Boundary Method," J. Appl. Phys., vol. 67, pp. 6353 - 6359, May 1990.

[10] D. J. Kirker, C. S. Lent, and S. Sivaprakasam, "The Numerical Simulation of Electron Transmission Through a Two-Dimensional Quantum Device by the Finite Element Method," Intern. J. Numerical Meth. Engin., vol. 29, 1527-1537, 1990.

CONFINED ACOUSTIC PHONON CONTROLLED RELAXATION TIMES IN FREE STANDING QUANTUM WELLS

N. Bannov, V. Aristov, and V. Mitin
Department of Electrical and Computer Engineering
Wayne State University, Detroit, MI 48202

Abstract

The scattering rates for electron – confined acoustic phonon interactions in free-standing quantum wells are calculated numerically. We have considered the relaxation times in the test particle approximation as well as in the approximation corresponding to the kinetic equation solution through the polar functions expansion. The quantisation of acoustic phonons in free-standing quantum wells results in peculiarities in transport coefficients.

I. INTRODUCTION

Free-standing quantum structures attract considerable attention because they are promising for optoelectronic and electronic applications, as sensitive sensors, and for probing the local properties of solids. Furthermore, the free-standing structures are very interesting physical objects which display new physical phenomena and they are challenging objects for nanotechnology. The main feature of the acoustic phonon subsystems in free-standing structures is the quantization of the acoustic phonon wavevectors in the direction of the confinement. This quantization is responsible for the peculiarities of the acoustic phonon interactions with electrons and photons displayed as a set of peaks in the differential conductances [1] and luminescence spectra [2] of quantum microstructures. In this paper we concentrate our attention on the electron - confined acoustic phonon interactions in free-standing quantum wells (FSQWs) and on the peculiarities of the transport coefficients due to the acoustic phonon quantization. We have developed a model of the electron scattering by confined acoustic phonons interacting through the deformation potential. This model is used in collision integral of the kinetic equation. We solved the kinetic equation both in the test particle approximation and reducing it to the Fredholm equation of the second kind.

II. ELECTRON – CONFINED ACOUSTIC PHONON SCATTERING

We will consider FSQW of width a. The electron wavefunctions will be taken in the approximation of the infinitely deep quantum well. The electron states are characterized by the in-plane wavevector $\mathbf{k}_{||}$ and the subband number n. The acoustic phonon eigenmodes in FSQW and their dispersion relations were obtained in [3, 4, 5]. The acoustic phonons are characterized by the in-plane wave vector $\mathbf{q}_{||}$, the mode number m, and the symmetry α.

In accordance with the Fermi golden rule the probability density for the electron transition $(\mathbf{k}_{||}, n) \rightarrow (\mathbf{k}'_{||}, n')$ due to the confined phonon absorption (upper sign) or emission (lower sign) is given by the formula

$$W^{\left\{\substack{ab\\em}\right\}}_{\mathbf{k}_{\|},n\rightarrow\mathbf{k}'_{\|},n'} = \frac{\pi E_a^2 (n^{\alpha}_{\mathbf{q}_{\|},m} + \frac{1}{2} \mp \frac{1}{2}) \mid F_{\alpha,m} \mid^2 \left(q_{t,m}^2 - q_x^2\right)^2 \left(q_{l,m}^2 + q_x^2\right)^2}{\mathcal{A}\rho\omega_m^{(\alpha)}(\mathbf{q}_{||})} \times \qquad (1)$$

$$\times\, tsc_{\alpha}^2(\frac{a\, q_{t,m}}{2})\, \mathcal{G}(n', n, \alpha, q_{l,m})\, \delta_{\mathbf{k}_{\|}\pm\mathbf{q}_{\|},\mathbf{k}'_{\|}}\, \delta(\varepsilon \pm \hbar\omega_m^{(\alpha)}(\mathbf{q}_{||}) - \varepsilon')\,,$$

where we use the same notations as in Ref. [5, 6], the function $tsc_{\alpha} = \sin$, if $\alpha = dilatational$ and $tsc_{\alpha} = \cos$, if $\alpha = flexural$, $\mathcal{G}(n', n, \alpha, q)$ is the overlap integral.

To analyse the electron transport properties we will need scattering rates in the following form

$$\tau_G^{-1} = \sum_{n',\mathbf{k}'_\parallel,\alpha,m,\mathbf{q}_\parallel,\beta} W^{\beta}_{\mathbf{k}_\parallel,n\to\mathbf{k}'_\parallel,n'}\, G\,, \tag{2}$$

where β is either *absorption* or *emission*, G is some given function which may depend on all variables over which we take sum. We will also use $(\tau_G^{ab})^{-1}$ and $(\tau_G^{em})^{-1}$ which are defined in a similar way with the only distinction that we sum up either only absorption terms or only emission terms. There is an obvious relation between them: $\tau_G^{-1} = (\tau_G^{ab})^{-1} + (\tau_G^{em})^{-1}$. If we employ the formulae for transition probabilities (1) we may obtain the following relations for scattering rates

$$\left(\tau_G^{\left\{\substack{ab\\em}\right\}}\right)^{-1} = \frac{E_a^2\, m}{2\,\pi\,\hbar^2\,\rho\,k_\parallel} \sum_{n',\alpha,m} \sum_i \int_0^\infty dq_\parallel\, \mathcal{F}^{\left\{\substack{ab\\em}\right\}}(n',\, n,\, \alpha,\, m,\, q_\parallel)\, G\, \frac{1}{|\sin\Psi_i|}\,, \tag{3}$$

where

$$\mathcal{F}^{\left\{\substack{ab\\em}\right\}} = \frac{(n^{\alpha}_{\mathbf{q}_\parallel,m} + \frac{1}{2} \mp \frac{1}{2})\,|\,F_{\alpha,m}\,|^2 \left(q_{t,m}^2 - {q_x}^2\right)^2 \left(q_{t,m}^2 + {q_x}^2\right)^2}{\omega_m^{(\alpha)}(\mathbf{q}_\parallel)}\, tsc_\alpha^2\left(\frac{a\,q_{t,m}}{2}\right) \mathcal{G}(n', n, \alpha, q_{t,m})\,,$$

and angles $\Psi_i \in [0,\pi]$ are solutions of the transcendental equation

$$\cos\Psi = \frac{m\,\omega_m^{(\alpha)}(q_\parallel)}{\hbar\,k_\parallel\,q_\parallel} \pm \left(\frac{m\,(\varepsilon_n - \varepsilon_{n'})}{\hbar^2\,k_\parallel\,q_\parallel} - \frac{q_\parallel}{2\,k_\parallel}\right).$$

III. RELAXATION TIMES IN THE TEST PARTICLE APPROXIMATION

In the test particle approximation, the scattering rate, τ, the momentum relaxation rate, τ_p, and the energy relaxation rate, τ_ε are given by the formulae

$$\tau^{-1} = \sum_{\mathbf{p}'} W_{\mathbf{p}\to\mathbf{p}'}\, \frac{1 - f^0_{\mathbf{p}'}}{1 - f^0_{\mathbf{p}}}\,, \tag{4}$$

$$\tau_p(\mathbf{p})^{-1} = \sum_{\mathbf{p}'} W_{\mathbf{p}\to\mathbf{p}'} \left[1 - \frac{p'\cos\psi}{p}\right] \frac{1 - f^0_{\mathbf{p}'}}{1 - f^0_{\mathbf{p}}}\,, \tag{5}$$

$$\tau_\varepsilon(\mathbf{p})^{-1} = \sum_{\mathbf{p}'} W_{\mathbf{p}\to\mathbf{p}'} \left[1 - \frac{\varepsilon'}{\varepsilon}\right] \frac{1 - f^0_{\mathbf{p}'}}{1 - f^0_{\mathbf{p}}}\,, \tag{6}$$

where $f^0_{\mathbf{p}}$ is the Fermi distribution function, ψ is the angle between $\mathbf{p}$ and $\mathbf{p}'$.

We have computed integrals in (4), (5), and (6) for the electron scattering by the dilatational phonons in the lowest electron subband numerically and obtained the scattering rates τ^{-1}, τ_p^{-1}, and τ_ε^{-1} as functions of energy. The calculations were made for $GaAs$ QW of width $a = 100\mathring{A}$, for temperatures $T = 300K$, $T = 77K$, and $T = 4.2K$, for both degenerate and nondegenerate electron gas. The degenerate electron gas was characterized by the Fermi energy $\mu = 50\,meV$ and corresponding to it the electron concentration $n_s = 1.4 \times 10^{12}\,cm^{-2}$. The most interesting dependences are depicted in the Fig. 1 through 4. Fig. 1 and 2 correspond to the case of nondegenerate electron gas at temperatures $T = 300K$ and $T = 4.2K$ accordingly and Fig. 3 and 4 correspond to the case of degenerate electron gas at temperatures $T = 77K$ and $T = 4.2K$

accordingly. The solid lines in the Fig. 1 through 4 correspond to the acoustic phonon emission and the dotted lines correspond to the acoustic phonon absorption. It should be noted that the relaxation times τ and τ_p are very similar, therefore we will provide the graphs of only τ_p. The quantities τ_ε^{-1} for phonon absorption are obviously negative due to the factor $[1 - \varepsilon'/\varepsilon]$, however we use the same axes as for the energy relaxation rate corresponding to the phonon emission and plot them as positive functions.

IV. THE KINETIC RELAXATION TIME

The kinetic equation for the electron distribution function, f, may be solved in the case of small deviation from equilibrium. In this case $f = f_{\mathbf{p}}^0 + \mathrm{f_1}\,\mathrm{p}/p$, where the nonequilibrium part of the distribution function may be represented in the form $\mathrm{f_{1p}} = -\tau_1(p)\,\mathbf{F}\,\frac{\partial f_{\mathbf{p}}^0}{\partial_p}$. The momentum relaxation time, τ_1 satisfies the Fredholm equation of the second kind

$$\tau_1(p) = \tau(p) + \tau(p) \sum_{\mathbf{p}'} W_{\mathbf{p}\to\mathbf{p}'}\, \frac{p' \cos\psi}{p}\, \tau_1(p')\, \frac{1 - f_{\mathbf{p}'}^0}{1 - f_{\mathbf{p}}^0}. \qquad (7)$$

We solved Eq. (7) by iterations. The electron conductivity may be expressed through τ_1 in the following way

$$\sigma = \frac{e^2}{2\pi\hbar^2 m^2 T} \int_0^\infty dp\, p^3 \tau_1(p)\, f_{\mathbf{p}}^0 (1 - f_{\mathbf{p}}^0).$$

Acknowledgement: The work was supported by ARO and NSF.

References

[1] J. C. Nabity, M. N. Wybourne. Phys. Rev. B., 44, 8990, (1992).
[2] Y.F.Chen, J.L.Shen, L.Y.Lin. J.Appl.Phys., 73, 4555, (1993).
[3] Physical Acoustics, Ed. W. Mason, Vol.1, Part A, Academic Press, New York, 1964.
[4] B. A. Auld, Acoustic Fields and Waves, Wiley, New York, 1973.
[5] N. Bannov, V. Mitin, and M. A. Stroscio, Phys. Stat. Sol. (b), 183, no 1, 1994.
[6] V. Aristov, N. Bannov, V. Mitin, and M. A. Stroscio, Int. Workshop on Comp. Electr., Portland, Oregon, May 1994.

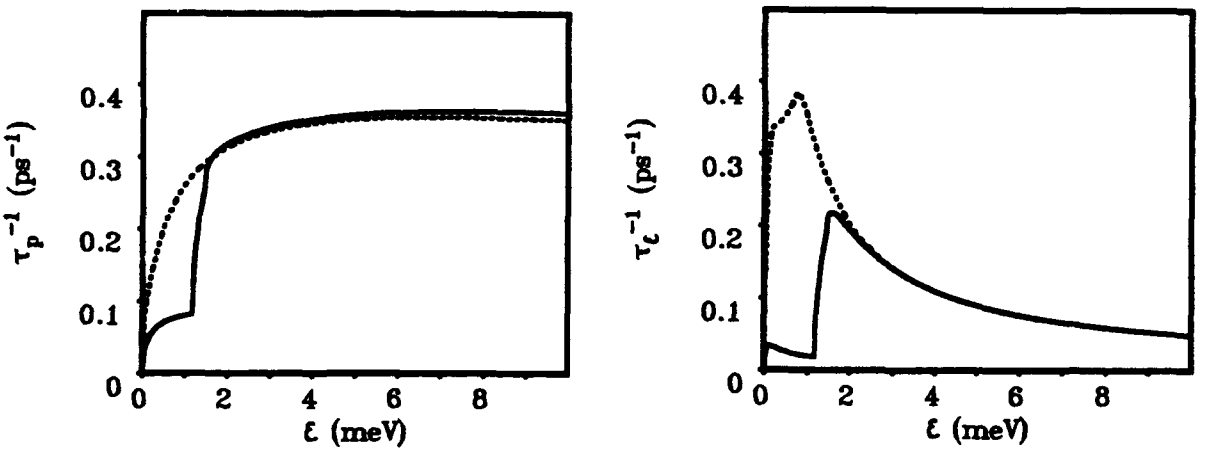


Figure 1. The momentum relaxation rate, τ_p^{-1}, (left) and the energy relaxation rate, τ_ε^{-1}, (right) in *GaAs* FSQW of width $a = 100\mathring{A}$. Nondegenerate case, $T = 300K$, solid line corresponds to the phonon emission, dotted line corresponds to the phonon absorption.

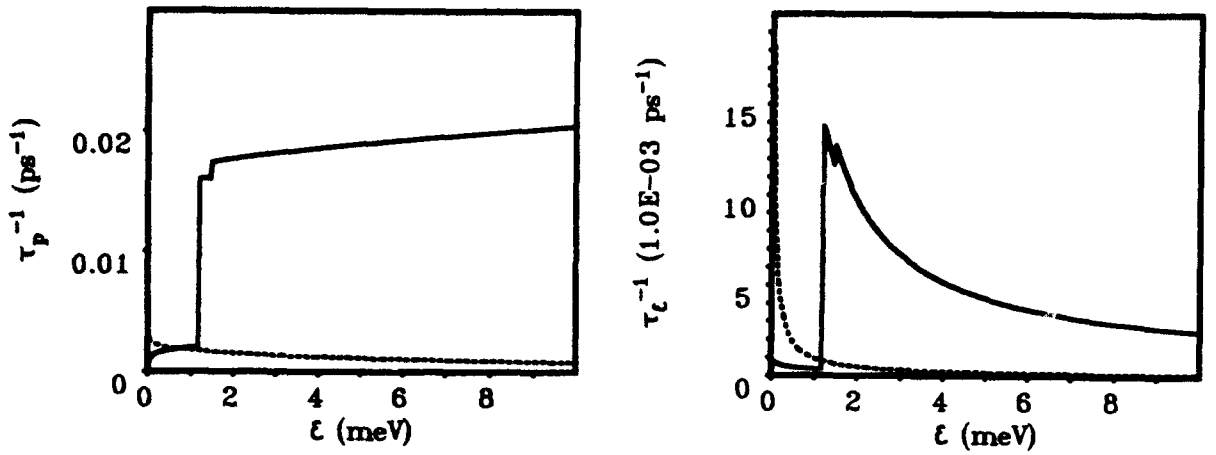


Figure 2. The momentum relaxation rate, τ_p^{-1}, (left) and the energy relaxation rate, τ_ε^{-1}, (right) in *GaAs* FSQW of width $a = 100\mathring{A}$. Nondegenerate case, $T = 4.2K$, solid line corresponds to the phonon emission, dotted line corresponds to the phonon absorption.

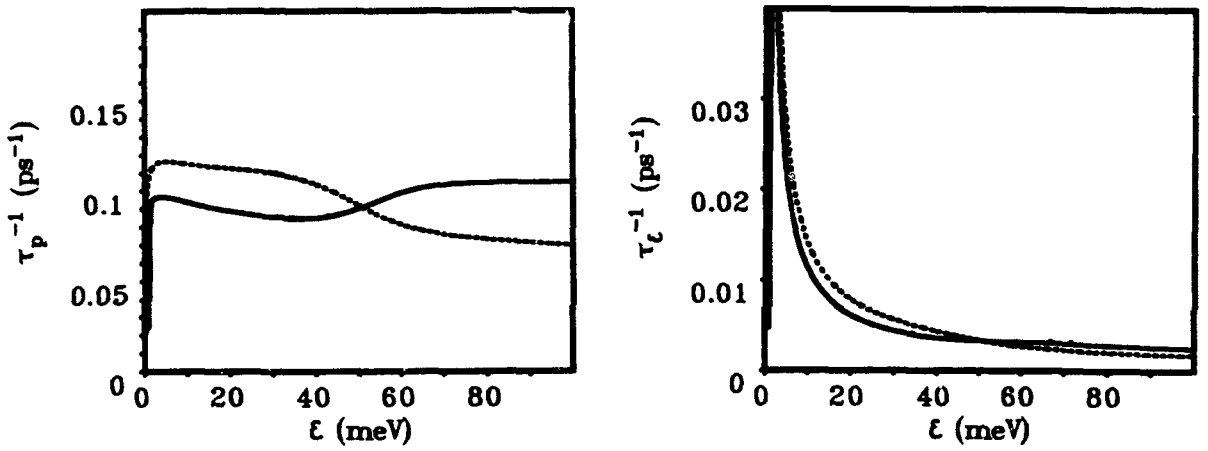


Figure 3. The momentum relaxation rate, τ_p^{-1}, (left) and the energy relaxation rate, τ_ε^{-1}, (right) in *GaAs* FSQW of width $a = 100\mathring{A}$. Degenerate case, $T = 77K$, solid line corresponds to the phonon emission, dotted line corresponds to the phonon absorption.

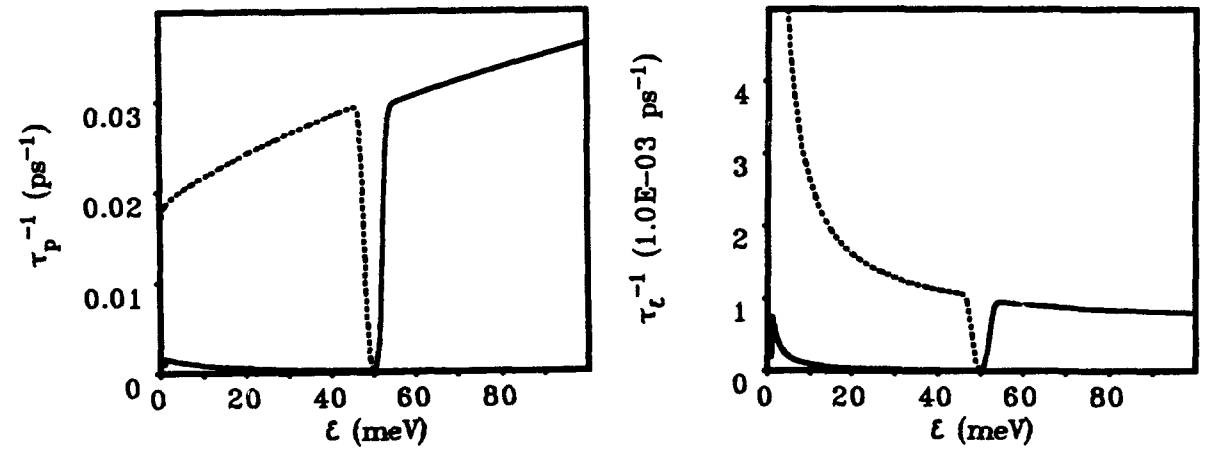


Figure 4. The momentum relaxation rate, τ_p^{-1}, (left) and the energy relaxation rate, τ_ε^{-1}, (right) in *GaAs* FSQW of width $a = 100\mathring{A}$. Degenerate case, $T = 4.2K$, solid line corresponds to the phonon emission, dotted line corresponds to the phonon absorption.

MONTE CARLO SIMULATION OF ELECTRON STREAMING CAUSED BY INELASTIC ACOUSTIC-PHONON SCATTERING IN QUANTUM WIRES

R. Mickevičius and V. Mitin

Department of Electrical and Computer Engineering
Wayne State University, Detroit, MI 48202

Abstract

We have simulated by the Monte Carlo technique a ***qualitatively new*** regime of electron transport in quantum wires, which resembles electron streaming. Unlike conventional streaming caused by optical-phonon scattering, the streaming reported here is due to inelastic ***acoustic-phonon*** scattering. Both the analytical model and the Monte Carlo simulations yield $E^{1/5}$ field dependence of the drift velocity as a function of electric field E in the streaming regime. We demonstrate that this regime of electron transport is accompanied by strong radiation of nonequilibrium acoustic phonons from a quantum wire.

I. INTRODUCTION

Qualitatively new regime of electron transport in quasi-one-dimensional (1D) quantum wires (QWIs) has been predicted recently [1]. This regime resembles electron streaming and originates from strongly inelastic acoustic phonon scattering in QWIs [2]. Unlike conventional streaming due to electron scattering by optical phonons [3-5] the streaming-like electron behavior reported in [1] is caused by strongly inelastic acoustic-phonon scattering. The streaming due to acoustic-phonon emission leads to non-linear velocity-field dependence and to oscillating electron velocity autocorrelation function.

In this paper we present the results of Monte Carlo simulation of electron streaming due to acoustic-phonon emission in QWIs.

II. ANALYTICAL APPROACH

Let us first consider idealized model in order to obtain simple analytical expressions. We neglect acoustic-phonon absorption and assume that electrons are scattered exactly to the subband bottom after emission of acoustic phonon. Let us define ϵ_c as the characteristic acoustic-phonon energy determined by uncertainty of momentum conservation in 1D structures. It is given [2] by $\epsilon_c \approx 2\pi\hbar u/L$ where u is the sound velocity in the material of a QWI and L is the effective thickness of the structure $L^{-2} = L_y^{-2} + L_z^{-2}$. At low energies ($\epsilon < \epsilon_c$) the acoustic-phonon emission rate can be approximated by:

$$\lambda(\epsilon) = \Lambda\,\epsilon^2\,, \tag{1}$$

where Λ is a constant independent of cross-section of a QWI. The mean free flight time $< \tau >$ generally reads as,

$$< \tau > = \int_0^\infty d\tau\ \tau\ \lambda[\epsilon(\tau)]\ \exp\left(-\int_0^\tau dt\ \lambda[\epsilon(t)]\right)\,, \tag{2}$$

where λ is the total scattering rate, in our case equal to acoustic-phonon emission rate given by Eq. (1). In electric field electron momentum during free flight is governed by: $dp/dt = eE$. Substituting energy expressed through momentum $\epsilon = p^2/2m^*$ and assuming that electron after a free flight is scattered exactly to the subband bottom, we get the solution in the form of $\epsilon = (eEt)^2/2m^*$. Substituting it into Eq. (1), then Eq. (1) into Eq. (3), and performing integration we get,

$$< \tau > = \left(\frac{5}{\Lambda C^4 E^4}\right)^{1/5}\ \Gamma\,, \tag{3}$$

where $C = e/\sqrt{2m^*}$ and constant $\Gamma = \Gamma(6/5) \approx 0.9182$ is the value of the Gamma function. Then averaging energy $\epsilon(t)$ over the mean free flight $0 - < \tau >$ we find the mean electron energy,

$$< \epsilon > = \frac{\hbar\omega}{3}, \tag{4}$$

where $\hbar\omega$ is the average acoustic-phonon energy emitted by electrons,

$$\hbar\omega = \left(\frac{5C}{\Lambda}\right)^{2/5} \Gamma^2 \, E^{2/5} . \tag{5}$$

Similarly averaging instant electron velocity over $< \tau >$ the drift velocity is obtained,

$$v_d = \sqrt{\frac{\hbar\omega}{2m^*}} . \tag{6}$$

Hence, the mean electron energy is a $E^{2/5}$ function and the drift velocity is $E^{1/5}$ function of the electric field. The drift velocity and the mean electron energy are simply related to each other: $< \epsilon >= 2/3 \; m^* \; v_d^2$. The relationships (4) and (6) are the same as for conventional streaming due to optical phonon emission [5], but the characteristic acoustic phonon energy (5), unlike optical phonon energy, depends on electric field. Therefore, in contrast to conventional streaming where v_d and $< \epsilon >$ saturate [5], the streaming due to acoustic-phonon emission leads to field-dependent v_d and $< \epsilon >$.

The conventional streaming due to optical phonons can be realized if certain conditions are met [4]: (i) the temperature must be low enough, generally $k_B T \ll \hbar\omega$, where $\hbar\omega$ is phonon energy, (ii) the phonon emission rate must exceed all other scattering rates near the emission threshold, (iii) the electric field should be strong enough to accelerate electron up to the phonon emission threshold without scattering, but weak enough to avoid deep electron penetration beyond the emission threshold and thus to assure scattering by phonon emission down to the conduction band bottom.

In the case of electron streaming due to periodic *acoustic*-phonon emission the first two conditions, however, are generally fulfilled for acoustic-phonon scattering if $\epsilon_c \gg k_B T$. Let us estimate the range of electric fields $E_{min} \ll E < E_{max}$, where the streaming due to acoustic-phonon emission occurs. First, we define the "passive region" as the energy range where acoustic-phonon emission rate is less than absorption rate. By requiring that electron acceleration time through the "passive region" be much less than the absorption time, we get the lower field limit E_{min}. The condition $\hbar\omega < \epsilon_c$ sets the upper field limit of E_{max}. The lower field limit E_{min} weakly depends on the cross-section and is approximately equal to 1 V/cm. The upper limit is around 200 V/cm for $40 \times 40 \mathring{A}^2$ QWI, 35 V/cm for $80 \times 80 \mathring{A}^2$ QWI, and 4 V/cm for $250 \times 150 \mathring{A}^2$.

III. MONTE CARLO SIMULATIONS

1. Model

We have carried out Monte Carlo simulations of electron transport in a wide range of electric fields. We have considered rectangular GaAs QWIs embedded in AlAs with infinitely deep potential well for electrons. We have chosen several different cross-sections of a QWI, from rather thick $250 \times 150 \mathring{A}^2$ QWI, where separation between two lowest subbands is less that optical phonon energy, to unrealistically thin $40 \times 40 \mathring{A}^2$ QWI, which represents the extreme limit. Simulations have been performed for low temperature $T = 4 \; K$ and non-degenerate electron gas. Electron scattering by confined longitudinal optical (LO) phonons and localized interface (surface) SO phonons [6,7] as well as by bulk-like acoustic phonons [2] has been taken into account in our model. Our model incorporates as many subbands as there are actually occupied by electrons. Ionized impurities are

assumed to be located sufficiently far from the QWI so that their influence on the electron motion inside the wire is negligible.

2. Results

Fig. 1 demonstrates drift velocity as a function of electric field calculated by the Monte Carlo technique. There are four distinguishable regions on velocity-field dependence. The near-ohmic velocity-field dependence in a field range below 1 V/cm turns into a sub-linear dependence. Then the slope again increases and decreases approaching saturation at high electric fields. The first sub-linear region extends through the two orders of magnitude in electric fields in $40 \times 40\mathring{A}^2$ QWI (2 V/cm to 200 V/cm) and appears just as a small kink in $250 \times 150\mathring{A}^2$ QWI at around 2 V/cm. In the field range of 5 V/cm to 200 V/cm in $40 \times 40\mathring{A}^2$ electron drift velocity increases near as $E^{1/5}$ function of electric field as is predicted by Eqs. (5)–(6). Note that this field range coincides with the above estimated range 1 V/cm $\ll E <$ 200 V/cm for this QWI, where electron streaming due to acoustic-phonon emission occurs. At high electric fields the optical-phonon emission starts dominating and the drift velocity saturates. The transition from the acoustic-phonon controlled electron transport to the optical-phonon controlled transport occurs at lower electric fields in thick QWIs, where acoustic phonon scattering rate is lower [2].

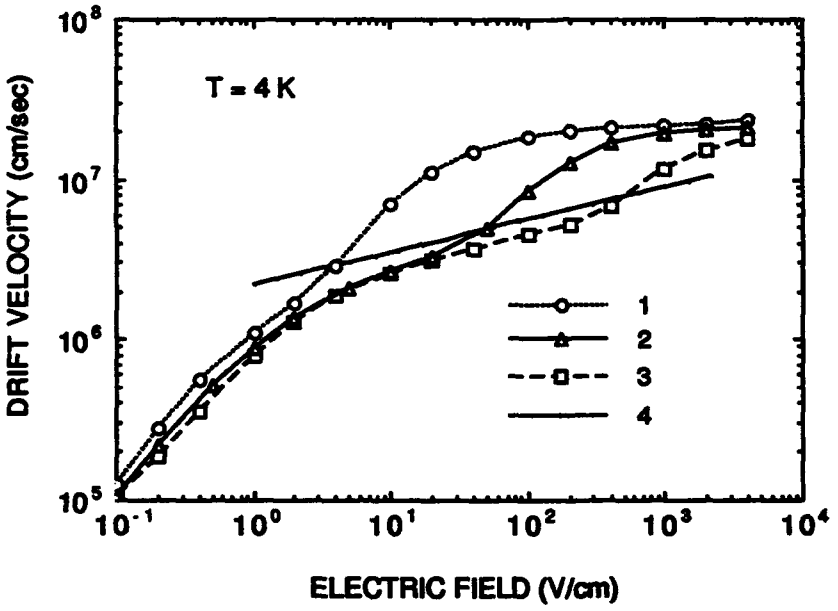


Fig. 1. Electron drift velocity versus applied electric field. Curve 1 represents velocity for $250 \times 150\mathring{A}^2$ QWI, curve 2 – $80 \times 80\mathring{A}^2$ QWI, and curve 3 – $40 \times 40\mathring{A}^2$ QWI; curve 4 represents analytical dependence given by Eqs. (6) and (5).

Fig. 2 demonstrates the relative scattering efficiency (ratio of the number of corresponding scattering events to the total number of real scattering events) versus electric field. One can see that in the field range below about 1 V/cm there is a balance between acoustic-phonon absorption and emission efficiency indicating the ohmic regime of electron transport. Then the emission efficiency gradually increases and the absorption efficiency decreases up to electric fields of about 200 V/cm reflecting the transition to the acoustic phonon controlled electron streaming. The emission of acoustic phonons remains the sole scattering mechanism in the field range of 10 V/cm to 200 V/cm. Strong emission of acoustic phonons suggests that QWIs should radiate nonequilibrium acoustic phonons in the streaming regime. The question is which part of this radiation is directed along the QWI and which part goes into surrounding material. The ratio of the transverse component of acoustic phonon wave vector to the total magnitude of the wave vector $\eta = q_T/\sqrt{q_z^2 + q_T^2}$ roughly defines the relative fraction of the radiation going outside a QWI in the total radiation of acoustic phonons. We have calculated η as a function of electric field by the Monte Carlo technique. Our

calculations show that $\eta \geq 0.98$ in the entire field range of 1 V/cm to 1000 V/cm. Consequently, a QWI radiates acoustic phonons predominantly in the perpendicular to a QWI direction. We believe that strong radiation of nonequilibrium acoustic phonons and their angular distribution could be experimentally measurable in the streaming regime.

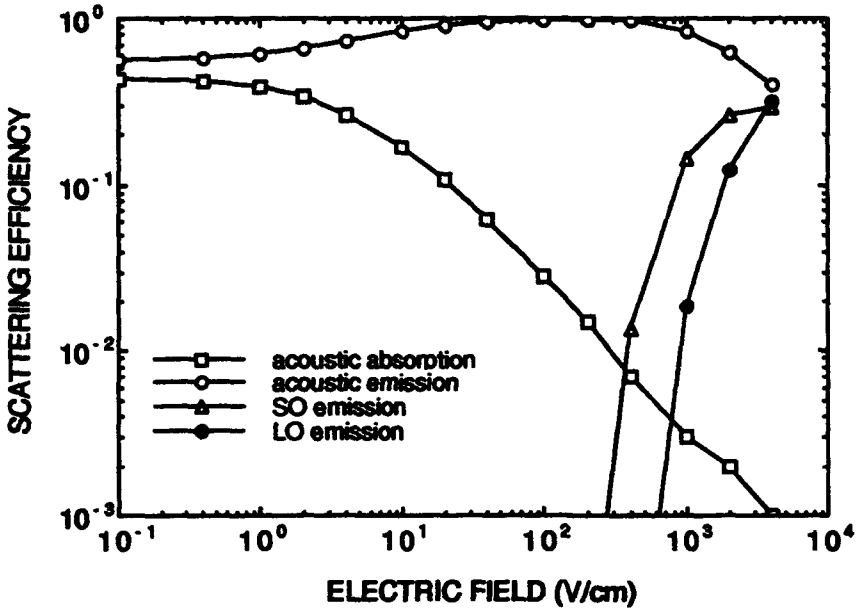


Fig. 6. Relative efficiency of various scattering mechanisms versus electric field for $40 \times 40\mathring{A}^2$ QWI.

The fields of about 300 V/cm and up are strong enough to heat the electrons up to the lowest optical phonon energy (in our case, GaAs-like interface mode energy equals 34.5 meV) and one can see the rapid onset of SO phonon scattering. With further increase of electric field the LO phonon scattering comes into play, and thus the electron transport in the fields exceeding 400 V/cm is primarily controlled by optical phonon scattering.

IV. SUMMARY

We have investigated qualitatively new regime of electron transport in QWIs, which closely resembles electron streaming due to *acoustic-phonon* emission. Both analytical and Monte Carlo calculations yield non-linear velocity-field relationships in the streaming regime. It is demonstrated that a QWI becomes effective radiator of nonequilibrium acoustic phonons in the streaming regime.

Acknowledgement: The work was supported by the U.S. Army Research Office and the National Science Foundation.

1. R. Mickevičius and V. Mitin, submitted to Phys. Rev. Lett.
2. R. Mickevičius and V. Mitin, Phys. Rev. **B48**, 17194 (1993).
3. W. Shockley, Bell Sys. Tech. J. **30**, 990 (1951).
4. S. Komiyama, Adv. Phys. **31**, 255 (1982).
5. D. Jovanovic and J.P. Leburton, Superlattices and Microstructures **11**, 141 (1992).
6. M. A. Stroscio, G. J. Iafrate, K. W. Kim, M. A. Littlejohn, H. Goronkin, and G. N. Maracas, Appl. Phys. Lett. **59**, 1093 (1991).
7. R. Mickevičius, V. V. Mitin, K. W. Kim, and M. A. Stroscio. Semicond. Sci. Technol. **7**, B299 (1992).

ACOUSTIC PHONON SPECTRUM AND DENSITY OF STATES IN FREE STANDING QUANTUM WELLS

V. Aristov [a], N. Bannov [a], V. Mitin [a], and M. A. Stroscio [b]
[a] *Department of Electrical and Computer Engineering*
Wayne State University, Detroit, MI 48202
[b] *U.S. Army Research Office, P.O.Box 12211*
Research Triangle Park, NC 27709-2211

Abstract

The confined acoustic phonons in free-standing quantum wells are considered. Their spectrum may determined from the dispersion equations. We have developed a special stable algorithm to obtain numerical solutions of these equations. We have calculated the acoustic phonon density of states in a free-standing quantum well. The density of states is , on the average, a quadratic function of energy, however it has singularities corresponding to the extrema in the dispersion relations.

I. INTRODUCTION

In low dimensional microstructures acoustic phonon states may undergo significant modifications due to the quantization in one, two, or three directions. Acoustic phonon confinement will strongly affect the electron and photon interactions with acoustic phonons resulting in peculiarities of electron transport properties and light scattering. Therefore it is necessary to develop an adequate model of acoustic phonon states in low dimensional structures and their interactions with electrons and photons.

In this paper we consider confined acoustic modes in a thin solid slab of isotropic material. We have calculated the confined phonon spectrum and the corresponding density of states (DOS). The equation governing the elastic vibrations in our system is the Navier equation for a relative displacement vector and it is supplemented by appropriate boundary conditions which are the conditions of free (unstressed) surfaces [1, 2, 3]. We have transformed the problem at hand to an eigenvalue equation with a Hermitian matrix differential operator. The solutions of this eigenvalue problem are three different types of modes with different symmetries: shear waves, dilatational waves and flexural waves. Although the general form of the solution may be obtained analytically, it includes several parameters (phonon quantum numbers) which should be determined by numerically solving the system of nonlinear dispersion equations. These phonon quantum numbers are complex valued functions of the in-plane phonon wave vector and they may approach each other so closely for some values of the in-plane wave vector, that the numerical solution of the dispersion equations leaps from one branch to another. We have developed a special stable algorithm to obtain these solutions.

II. CONFINED ACOUSTIC PHONON SPECTRUM

Shear waves have the simplest quantization rules. A vector of relative displacement in shear waves has only one nonzero component in the direction perpendicular to both the direction of propagation and the direction perpendicular to the slab . The dispersion relation for shear waves is

$$\omega_n = s_t \sqrt{q_n^2 + q_{||}^2}\,, \tag{1}$$

where s_t is the transverse sound velocity in the bulk material, $q_{||}$ is an in-plane wave vector, $q_n = (\pi n/a)$, $n = 0, 1, 2, ...$

Dilatational waves and flexural waves have two nonzero components of the vector of relative displacement – in the direction of wave propagation and in the perpendicular to the slab direction. The pattern of the vector of relative displacement is symmetric in respect to the slab midplane for dilatational waves and antisymmetric for flexural waves. The dispersion relations for dilatational waves are given implicitly by the system of equations

$$\omega_n^2 = s_l^2\,(q_\parallel^2 + l_n^2) = s_t^2\,(q_\parallel^2 + t_n^2)\,, \tag{2}$$

$$\frac{\tan(t_n a/2)}{\tan(l_n a/2)} = -\frac{4q_\parallel^2 l_n t_n}{(q_\parallel^2 - t_n^2)^2}\,, \tag{3}$$

where s_l is the longitudinal sound velocity in the bulk material, parameters l_n and t_n are determined from equations (2) and (3), which have many solutions as denoted by the index n. The dispersion relations for flexural waves are given implicitly by eq. (2) and the equation

$$\frac{\tan(l_n a/2)}{\tan(t_n a/2)} = -\frac{4q_\parallel^2 l_n t_n}{(q_\parallel^2 - t_n^2)^2}\,. \tag{4}$$

The graphs of functions $\omega_n(q_\parallel)$ obtained by numerical solutions of the system of eqs. (1)-(4) is shown in the Fig. 1a, 2a and 3a for shear waves, dilatational waves and flexural waves, respectively. We used elastic constants of $GaAs$ and took the slab width as $a = 100\mathring{A}$. These graps are plotted for the 12 lowest modes.

III. ACOUSTIC PHONON DENSITY OF STATES

The peculiarities of the acoustic phonon spectrum will be markedly pronounced in the their density of states (DOS). The DOS of confined phonons is defined by the formula

$$\mathcal{N} = \frac{\mathcal{A}}{(2\pi)^2}\sum_n \int_{\omega_n = const} \frac{dq_\parallel}{\left|d\omega_n/dq_\parallel\right|}\,, \tag{5}$$

where $\mathcal{A}$ is the area of the slab, and the sum is taken over phonon modes; integral in (5) is taken over the curve of constant energy and $\mathcal{N}$ is a function of the energy.

We have to specify the Brillouin zone to calculate the DOS over a wide range of energy. For a model estimation we accepted a simple square Brillouin zone. So we take into account only those acoustic phonons in integral (5) which have wavevectors inside the first Brillouin zone. The lattice constant is taken equal $5.65\mathring{A}$ which corresponds to the case of $GaAs$. The graph of the DOS obtained by numerical calculation of the integral of (5) for shear, dilatational and flexural phonons is depicted in Fig. 1b, 2b, and 3b, respectivelyly. At energies lower than some critical energy (corresponding to the edge of the Brillouin zone) the DOS is, on the average, a quadratic function of energy. This functional dependence occurs when many phonon branches contribute to the DOS and it corresponds to the case of bulk acoustic phonons. It is obscured in Fig. 1b, 2b, and 3b because the graphs are plotted in the semilogarithmic scale to emphasize the singularities of the DOS. These singularities correspond to the extrema in the dispersion relation; formally the DOS goes to infinity in such points. In Fig. 1b, 2b, and 3b, the DOS is plotted for energies up to $10\,meV$. At higher energies the finiteness of the Brillouin zone becomes important and the function $\mathcal{N}$ saturates in the average.

The DOS may be determined experimentally from neutron scattering spectra [4, 5] or from Brillouin light scattering spectra [6, 7]. It is a very important function characterizing the acoustic phonon subsystem and determining peculiarities of phonon interactions with phonons, photons and electrons. The singular points of the DOS make the observing conditions for the neutron scattering spectra and the Brillouin light scattering spectra more favorable, because the intensity of the scattered (reflected) radiation is proportional to the DOS of acoustic phonons.

IV. CONCLUSIONS

We have calculated the acoustic phonon modes and their density of states in free-standing quantum wells. The density of states has singularities related to the extrema of the acoustic phonon dispersion law. In these singular points the DOS formally goes to infinity. It makes the observing conditions for light and neutron scattering spectra more favorable.

Acknowledgments: The work was supported by ARO and NSF.

References

[1] Physical Acoustics, Ed. W. Mason, Vol.1, Part A, Academic Press, New York, 1964.
[2] B. A. Auld, Acoustic Fields and Waves, Wiley, New York, 1973.
[3] N. Bannov, V. Mitin, and M. A. Stroscio, Phys. Stat. Sol. (b), **183**, no 1, 1994.
[4] P. Bruesch, Phonons: Theory and experiments II, Springer-Verlag, 1986.
[5] B. Di Bartolo, R. Powell, Phonons and resonances in solids, John Wiley and Sons, 1976.
[6] M. Grimsditch, R. Bhadra, and I. Schuller, Phys. Rev. Lett., **58**, 1216 (1987).
[7] B. Bhadra, M. Grimsditch, I. Schuller, F. Nizzoli, Phys. Rev. B, **39**, 12456 (1989).

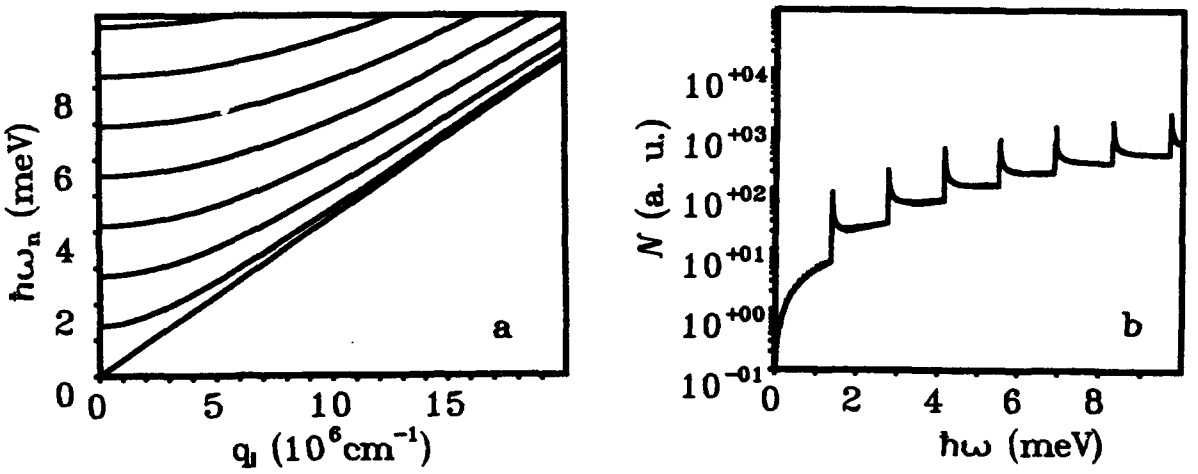


Figure 1. The dependences of the phonon energy, $\hbar\omega_n$, on the in-plane wavevector, $q_{||}$, (a) and the density of states, $\mathcal{N}$, on the phonon energy, $\hbar\omega$, (b) for shear phonons in a free-standing *GaAs* quantum well of width 100Å.

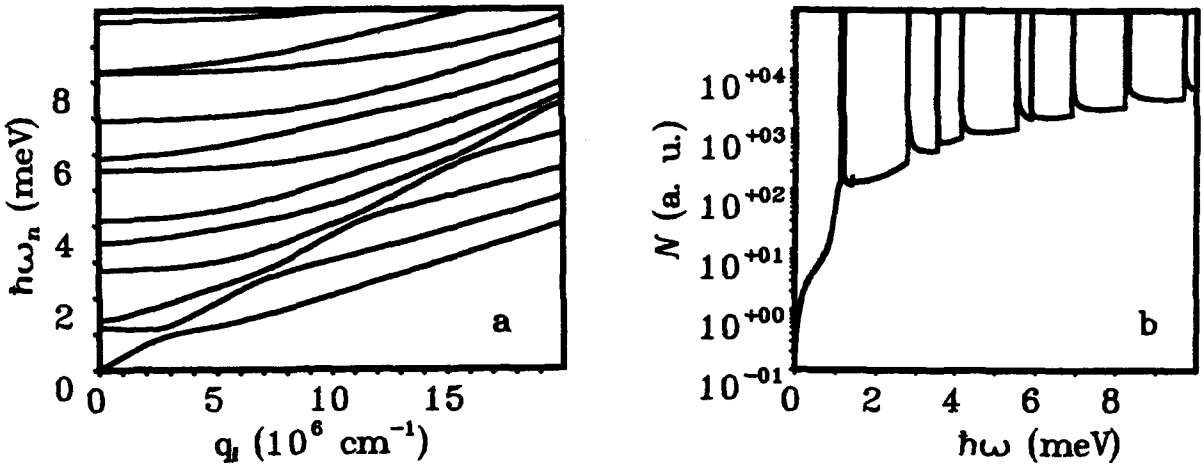


Figure 2. The dependences of the phonon energy, $\hbar\omega_n$, on the in-plane wavevector, $q_{||}$, (a) and the density of states, $\mathcal{N}$, on the phonon energy, $\hbar\omega$, (b) for dilatational phonons in a free-standing *GaAs* quantum well of width 100Å.

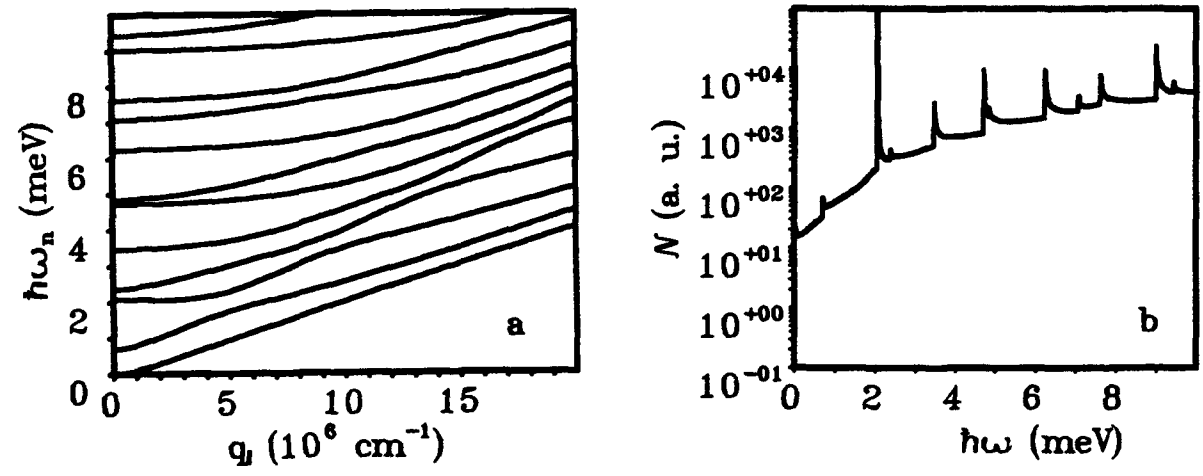


Figure 3. The dependences of the phonon energy, $\hbar\omega_n$, on the in-plane wavevector, $q_{||}$, (a) and the density of states, $\mathcal{N}$, on the phonon energy, $\hbar\omega$, (b) for flexural phonons in a free-standing *GaAs* quantum well of width 100Å.

DYNAMICAL TRANSPORT MODEL OF A COUPLED ELECTRON-PHONON SYSTEM IN A QUANTUM WIRE

Julie A. Kenrow and T. Kenneth Gustafson
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720

Abstract

A model has been developed for studying the effect of the electron-polar optical phonon interaction on electron transport in a GaAs/$Al_xGa_{1-x}As$ rectangular quantum wire system. The objective of this work is to investigate electron transport in ultra-submicron devices for which the inverse scattering rate can be on the order of the transit time. We solve the 3D, time-dependent Schrödinger equation non-perturbatively for the time evolution of the wavefunction of the coupled electron-LO and -SO phonon sytem. The time evolution of the mean electron momentum and directed energy are presented as a function of initial directed electron energy and applied voltage between the ends of the wire. Comparisons are made to semi-classical Monte Carlo results which are shown to significantly overestimate the amount of scattering on 100 fs time scales.

I. INTRODUCTION

The emerging capabilities in nanoscale electronic device fabrication have spurred the development of fully quantum mechanical electron transport models in semiconductors. Two important aspects of electron transport in nanoscale devices that necessitate a quantum treatment are confinement in one or more dimensions and ultra-short (i.e., sub-picosecond) transit times. To correctly model electron transport under conditions where transit times are comparable to mean free collision timess (~ 100 fs), a dynamical treatment of the electron-phonon interaction should be incorporporated. This precludes the common treatment of electron-phonon scattering based upon the use of predetermined Fermi's Golden Rule (FGR) obtained scattering rates. A dynamical treatment requires a quantum mechanical (coherent state) description for both the electron and the lattice phonons. In this paper we develop such a quantum transport model (QTM) for a quantum wire (QWI), which describes the time evolution of an electron wavepacket coupled to the non-equilibrium phonon mode spectrum. As a model problem we consider placing an electron wavepacket in a rectangular, finite length GaAs QWI surrounded by $Al_xGa_{1-x}As$, and model its time evolution as it simultaneously accelerates in response to an applied electric field, and interacts with either the confined LO or interfacial SO phonon modes.

II. QUANTUM TRANSPORT MODEL

We seek the time-dependent wavefunction $\Psi(\vec{r},t)$ of the coupled system, consisting of a single electron, and either the LO or SO polar optical lattice modes with which the electron interacts. The lattice temperature is taken to be 0 K, in order for $\Psi(\vec{r},t)$ to describe a coherent state with a well-defined initial condition. We solve the 3D time-dependent Schrödinger equation in the effective mass approximation,

$$i\hbar\frac{\partial\Psi}{\partial t} = (H_e + H_{ph} + H_{e\text{-}p})\Psi(\vec{r},t) \qquad (1)$$

where

$$H_e = -\frac{\hbar^2\nabla^2}{2m^*} + V(\vec{r}) \tag{2}$$

$$H_{ph} = \sum_q \hbar\omega_q (a_q^\dagger a_q + \tfrac{1}{2}) \,, \tag{3}$$

$V(\vec{r})$ is the potential energy for both the conduction band minimum and an externally applied voltage, and $\hbar\omega$ is the phonon energy. In (1), the interaction Hamiltonians for LO and SO phonons, appropriate for an embedded rectangular QWI, are [1]

$$H_{e\text{-}p}^{LO} = 2i\zeta \cos\left(\frac{\pi x}{\ell_x}\right)\cos\left(\frac{\pi y}{\ell_y}\right)\sum_q \frac{1}{Q}\left[a_q e^{iqz} - a_q^\dagger e^{-iqz}\right] \tag{4}$$

$$H_{e\text{-}p}^{SO} = \sum_q (-e)\,\Phi(q,x,y)\left(\frac{\hbar}{2\omega}\right)^{\frac{1}{2}}\left[a_q e^{iqz} - a_q^\dagger e^{-iqz}\right] \tag{5}$$

where $q \equiv q_z$ is the LO or SO phonon wave number corresponding to the direction of free propagation, along the wire. In this study, since we limit the transport time to 150 fs, we neglect inter-subband transitions, and assume that the electron remains in the transverse (x,y) ground state. Hence, in Eq (4), only the dominant, lowest order transverse LO phonon mode has been retained.

We write the solution $\Psi(\vec{r},t)$ as a superposition over the orthogonal basis of LO phonon number states,

$$\Psi(\vec{r},t) = \alpha(\vec{r},t)\,e^{-i\omega(N/2)t}\,|\,0\rangle + \sum_{\ell=1}^{N}\beta_\ell(\vec{r},t)\,e^{-i\omega(1+N/2)t}\,|\,1\rangle_\ell + \sum_{m=1}^{N_2}\gamma_m(\vec{r},t)\,e^{-i\omega(2+N/2)t}\,|\,2\rangle_m + \cdots \,, \tag{6}$$

where $|\,i\rangle_j$ represents the ith order phonon number state for the jth unique combination of phonon mode occupancies in the quantum wire, e.g.,

$|\,0\rangle = |0000...0\rangle$ — 0th order (vacuum state)

$|\,1\rangle_1 = |1000...0\rangle,\ |\,1\rangle_2 = |0100...0\rangle,\ \ldots$ — 1st order

$|\,2\rangle_1 = |1100...0\rangle,\ |\,2\rangle_2 = |1010...0\rangle,\ \ldots\ ;$ — 2nd order

$|\,2\rangle_{N_2-N+1} = |2000...0\rangle,\ \ldots,\ |\,2\rangle_{N_2} = |0000...2\rangle$

and where the modes span the spectrum of wavenumbers q. It follows that for a system of N modes, there are N combinations of 1st order number states and $N_2 = N(N+1)/2$ combinations of 2nd order number states. For this study, we retain terms only up through 2nd order in Eq (6). From Eqs (1)-(3), (4) or (5), and (6), the required set of coupled differential equations for the coefficients $\alpha(z,t)$, $\beta_\ell(z,t)$, $\gamma_m(z,t)$ is obtained by directly applying the Hamiltonian operators to $\Psi(\vec{r},t)$ in (6) and then equating coefficients associated with each unique number state. Previously [2], we have reported an approximate 1D solution to this problem for the LO interaction, in which the replacement $\cos(\pi x/\ell_x)\cos(\pi y/\ell_y)\rightarrow 1$ in Eq (4) was made to enable a separable solution

to Eq (1). Results for both the 1D and 3D models are presented here.

III. RESULTS AND DISCUSSION

The initial condition for $\Psi(\vec{r},0)$ is $\beta_\ell(z,0) = \gamma_m(z,0) = 0$ for all ℓ and m (corresponding to the lattice vacuum state at $T = 0\,\mathrm{K}$), and

$$\alpha(\vec{r},0) = \exp[-(z-z_o)^2/2(\Delta z_i)^2 + ik_i z]\,\phi_o(x,y), \tag{7}$$

where $\phi_o(x,y)$ is the 2D ground state eigenfunction for the GaAs/Al_xGa_{1-x}As system ($x = .45$ or 1.0). The effective strength of electron-LO phonon scattering (using the 1D model and $x = .45$) as a function of electron energy is shown in Fig 1a, in which $\langle k_z\rangle_t$ is plotted for several initial directed electron energies, E_i with initial spread $\Delta k_i = 5 \times 10^7$ m^{-1}, and no applied field. The initial flatness of the curves in Fig 1a is a result of the time ($\sim 1/\omega_{LO}$) required for conservation of energy. In Fig 1b, plots of $\langle k_z\rangle_t$, calculated from a semiclassical Monte Carlo Model (MCM) are shown. The MCM consists of an ensemble average over 10,000 electrons, with identical initial conditions, propagating along the wire, and interacting with the lattice by means of forward and backward scattering rates obtained from the approximate 1D LO Hamiltonian. The MCM predicts a much faster momentum relaxation since, over 150 fs, virtually all the electrons in the ensemble scatter one or more times, whereas the QTM results predict that a significant fraction of the initial electron wavepacket remains unscattered during this time period.

The effect of LO phonon scattering (using the 1D model and $x = .45$) on $\langle k_z\rangle_t$ has been studied for various electric field strengths with $E_i = 0$ and $\Delta k_i = 5 \times 10^7\ \mathrm{m}^{-1}$, as shown in Fig 2a. For each field strength, the $\langle k_z\rangle_t$ curves increase linearly with time until $\langle k_z\rangle_t = k_{ph} - \Delta k_i = 2.0 \times 10^8\ \mathrm{m}^{-1}$, at which time some of the electron wavenumber components begin to emit LO phonons. The corresponding results predicted by the MCM are shown in Fig 2b. Due to the singular scattering rate at the LO phonon energy of 36.2 meV, all electrons in the ensemble scatter when they reach an energy of 36.2 meV. In contrast, the QTM predicts only partial scattering of the electron wavepacket, so that the unscattered portion of the wavepacket can be accelerated (for the larger field strengths) to energies well in excess of 36.2 meV.

In Fig 3, $\langle k_z\rangle_t$ is plotted for $E_i = 50$ meV, $\Delta k_i = 5 \times 10^7\ \mathrm{m}^{-1}$, and no applied field, for both $x = .45$ and 1.0, using either the 3D electron-LO or -SO Hamiltonians

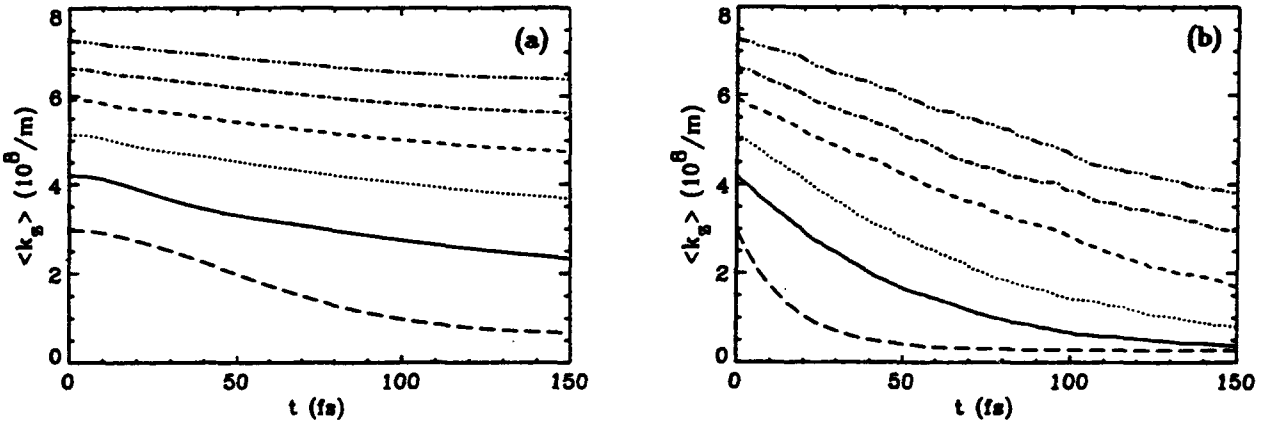


Fig. 1. $\langle k_z\rangle_t$ vs. t (a) QTM and (b) MCM for $E_i = 0.05$, 0.1, 0.15, 0.20, 0.25, 0.30 eV.

[Eqs (4)-(5)]. In agreement with published scattering rates for GaAs/AlAs quantum wires [1], the electron-LO interaction is more effective than the electron-SO interaction at dissipating the directed electron momentum. In addition to a greater interation strength for H_{LO} versus H_{SO}, the electron-LO interaction is more effective at randomizing the scattered wavevectors. This can be seen in Fig 3b, in which the electron probability density is plotted along the k_z axis at 150 fs for the cases from Fig 3a corresponding to $x = 1.0$, for each type of interaction. A large unscattered component is evident for the H_{SO} curve, and it can be seen that the forward- and back-scattered wavevector components (centered about $k_z = 0$) are much more nearly equal in magnitude for the H_{LO} curve compared to the H_{SO} curve, also in agreement with scattering rate predictions [3].

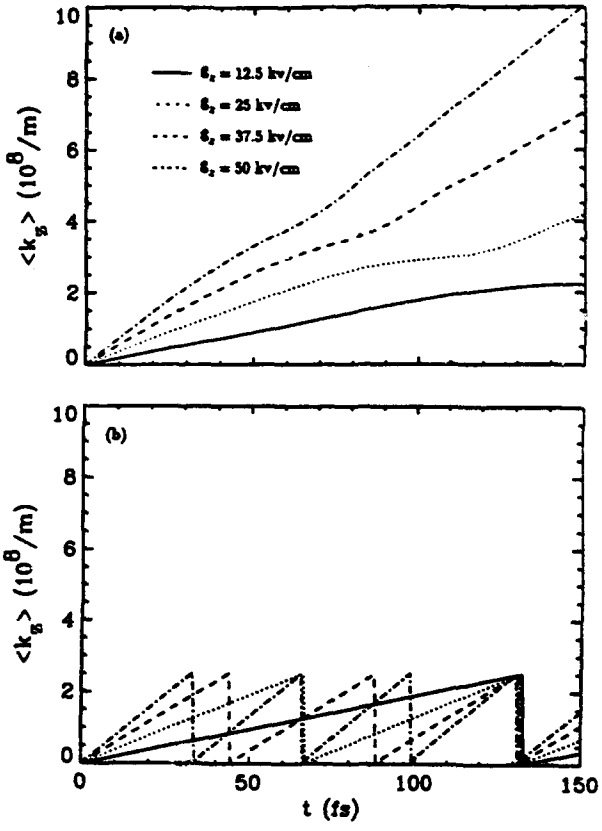


Fig. 2. $\langle k_z \rangle_t$ vs. t for $E_i = 0.0$ eV and various applied fields (a) QTM and (b) MCM.

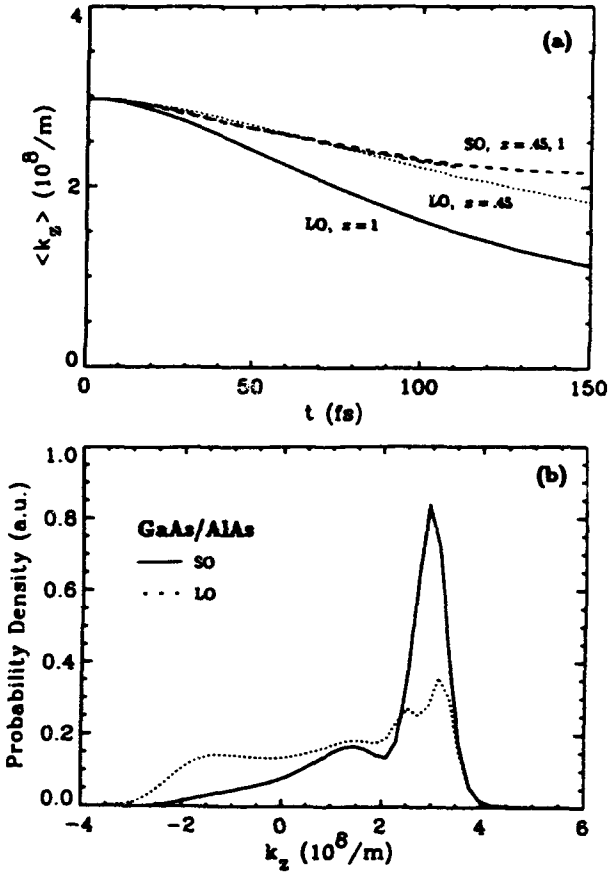


Fig. 3. (a) $\langle k_z \rangle_t$ vs. t for $E_i = 0.05$ eV. (b) Electron probability density $\rho_e(k_z, t)$ at 150 fs.

References

[1] K. W. Kim, M. A. Stroscio, A. Bhatt, R. Mickevicius, and V. V. Mitin, *J. Appl. Phys.* 70, 319 (1991).

[2] J. A. Kenrow and T. K. Gustafson, submitted for publication to *Phys. Rev. B.*

[3] R. Mickevicius, V. V. Mitin, K. W. Kim, M. A. Stroscio, and G. J. Iafrate, *J. Phys.: Condens. Matter* 4, 4959 (1992).

THE INCLUSION OF A FINITE CAPTURE TIME IN THE NUMERICAL SIMULATION OF QUANTUM EFFECT DEVICES.

M. Gault, H. Matsuura, K. Furuya*, P.Mawby and M.S. Towers*

Department of Electronic Engineering
University College of Swansea
Singleton Park Swansea
SA2 8PP
U.K.

**Department of Electronic Engineering*
Tokyo Institute of Technology
O-Okayama Meguro-Ku
Tokyo 2-12-1
Japan

Abstract

A simple numerical method is presented to include the phenomenon of a finite quantum well capture time in the numerical simulation of quantum effect devices. If the time taken to leave the quantum well (through tunnelling) is sufficiently short the electron states in the quantum well will remain relatively unoccupied due to the finite scattering time from the three dimensional continuum into the two dimensional states. A numerical formulation is presented which models this phenomenon by using an effective Fermi function for the occupancy of the two dimensional states, enabling the method to be used in conjunction with general purpose device simulators. The method is applied to a simple tunnel barrier to show the generality of the model.

I. INTRODUCTION

In the operation of single or double heterobarriers the accumulation of electrons at the heterointerface can have significant effects on the performance of the device. The principle of operation is shown in figure 1. As the bias is increased across the heterobarrier a potential induced accumulation layer forms at the heterointerface. Electrons in the accumulation layer have their energies quantised in the direction of propagation.

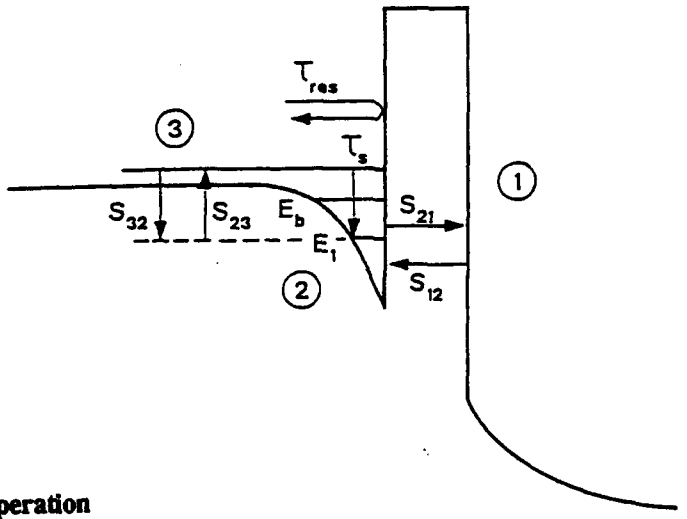


Figure 1: Principle of operation

To model this phenomenon requires accurate knowledge of the electron concentration in the accumulation layer. As the bias is increased above zero volts it is no longer valid to assume that the electrons in the accumulation layer are in thermal equilibrium. If the time required to enter the 2D states is much longer than the time required to leave the 2D states then the 2D states will remain relatively unoccupied. A numerical model for such structures is presented next.

II. NUMERICAL MODEL.

In order to simulate the steady state characteristics of the device the following equations must be solved self consistently. To obtain the electrostatic potential throughout the device Poisson's equation must be solved (assuming majority carrier operation)

$$\nabla.(\varepsilon\nabla\varphi) = -q(-n + N_D) \tag{1}$$

where ε is the composition dependent permittivity, q is the electronic charge, φ is the electrostatic potential, n is the electron concentration, and N_D is the concentration of ionised donors. The electron density is calculated from the addition of 2D electrons in bound energy states and 3D thermal electrons.

$$n = \underbrace{\sum_{i=1}^{n} \frac{m^* k_B T}{\hbar^2} f_i^* \ln\left[1 + \exp\left(\frac{E_f - E_i}{k_B T}\right)\right] |\varphi_i|^2}_{2D} + \underbrace{N_c F_{1/2}\left(\frac{E_f - E_b}{k_B T}\right)}_{3D} \tag{2}$$

where f is the non-equilibrium modification factor, k_B is Boltzmann's constant, E_f is the quasi-Fermi energy, N_C is the effective density of states, E_b represents the energy above which electrons are unbound (usually the top of the quantum well), $F_{1/2}$ represents the Fermi integral, m^* is the electron effective mass and $\hbar$ is Planck's constant $h/2\pi$. E_i and φ_i are the quantised energy levels and wavefunctions respectively, calculated from the effective mass Schrödinger equation,

$$\left(\frac{-\hbar^2}{2}\nabla\left(\frac{1}{m^*}\nabla\right) + V\right)\varphi_i = E_i\varphi_i \tag{3}$$

where V is the potential defining the quantum well (the conduction band edge).

To calculate the tunnelling current through the barrier a modified Tsu-Esaki formula [1] is used. In the case of an accumulation layer there are 2 components to the current density, a two dimensional component J_{2D} and a three dimensional component J_{3D}.

$$J_{2D} = \sum_{i=1} \frac{4\pi q m^* k_B T}{h^3} \int_0^{E_0} T(E).\ln\left(\frac{\left(1+\exp\left(\frac{E_f - E}{k_B T}\right)\right)^{f_1^*}}{\left(1+\exp\left(\frac{E_f - E - qV_{bi}}{k_B T}\right)\right)}\right) dE \qquad (4)$$

$$J_{3D} = \frac{4\pi q m^* k_B T}{h^3} \int_{E_0}^{\infty} T(E).\ln\left(\frac{\left(1+\exp\left(\frac{E_f - E}{k_B T}\right)\right)}{\left(1+\exp\left(\frac{E_f - E - qV_{bi}}{k_B T}\right)\right)}\right) dE \qquad (5)$$

where T(E) is the energy dependent tunnelling probability through the barrier (obtained from a solution to (3)) and V_{bi} is the voltage drop across the heterobarrier. Note that due to coupling between the quantum well and the collector the integral for the 2D current cannot be replaced by a summation.

The non-equilibrium factor f is used to cope with the non-equilibrium of electrons and is derived as [2].

$$f^* = \frac{\frac{\tau_{res}}{\tau_s} f_3^o g_3 + T_{21} f_1 g_1}{\frac{\tau_{res}}{\tau_s} f_3^o g_3 + T_{21} f_2^0 g_1} \qquad (6)$$

T_{21} represents the peak tunnelling probability through the barrier, τ_{res} represents the resident time of an electron above the quantum well, τ_s represents the scattering time into the well, f represents the occupancy probability and g represents the density of states. The subscripts refer to figure 1. The only unknowns are f_1 and g_1 and these can easily be calculated.

To model the effects of scattering on energy spectrum a simple method proposed by [3] is used. Lorentzian broadening of the tunnelling probability spectrum is assumed, the magnitude of which is calculated from the phase relaxation time, the time between phase altering events. In this way the tunnelling probability T(E) for any particular energy E is modified according to

$$T^*(E) = T(E)\left(\frac{\Gamma_c}{\Gamma}\right)\frac{1}{1+\left[(E-E_r)/(\Gamma/2)\right]^2} \qquad (7)$$

where $T^*(E)$ is the modified transmission probability of the barrier at energy E, E_T is the resonance energy of the well, Γ is the total energy width due to coupling outside the quantum well and phase breaking inside the well, $\Gamma = \Gamma_C + \Gamma_p$ where Γ_C and Γ_p are the coupling width and the width caused by phase breaking respectively. The relationship between the width caused by phase breaking and the phase breaking time τ_p is given by the uncertainty principle.

III. APPLICATIONS

Figure 2a shows how the occupancy of the well decreases with decreasing width of barrier. In this example the scattering time into the well is taken to be 1ps in agreement with recent experimental results[4]. The

result clearly indicates the importance of considering a finite scattering time in the case of narrow barriers. At barrier of widths of 8nm and above the quantum well may be taken as fully occupied (within the bounds of the underlying Fermi-Dirac statistics) but below 8nm this is clearly not so. At barrier widths of 2nm and below the occupancy of the two dimensional states in the quantum well drops rapidly after a small applied voltage and plays only a small part in the current operation of the device.

Figure 2b shows how the occupancy of the well varies for various scattering times into the quantum well. The barrier width in this exampie is taken to be 4nm. At a scattering time of 1ps the occupancy of the well drops linearly with applied voltage. As the scattering time into the quantum well increases so the occupancy of the well decreases, as would be expected.

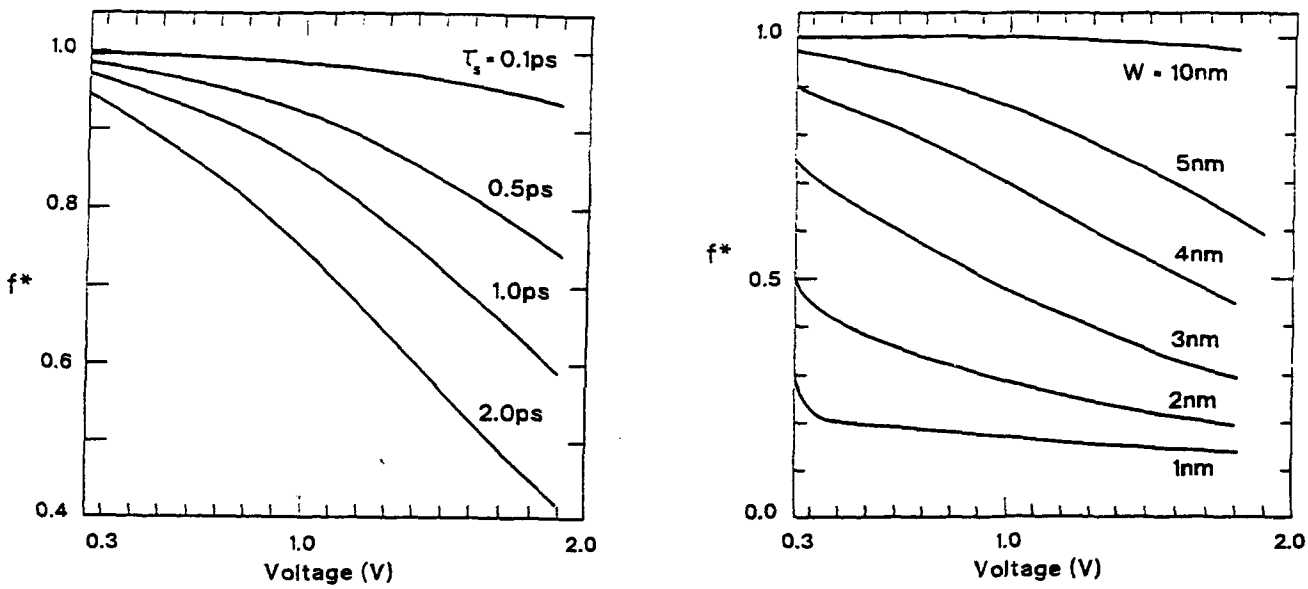


Figure 2 Occupation probability vs. a) scattering time and b) barrier thickness

IV. CONCLUSIONS

In summary a simple formalism has been derived to calculate the reduced occupancy of a quantum well due to a finite capture time which is both easy to calculate and to implement within existing device simulators. The results show the importance of including the finite capture time in tunnel barrier simulation for narrow barriers where the occupancy can be reduced to around 10% of its equilibrium value at a moderately high applied voltage.

V. REFERENCES

1. R. Tsu and L. Esaki, *Appl. Phys. Lett.*, 22, 562 (1973).
2. M. Gault, H. Matsuura, K. Furuya, P.Mawby and M.S. Towers, To be published in solid-state electronics (1994).
3. M. Jonson and A. Grincwajg, *Superlattices and Microstructures*, 12, 453 (1992).
4. M.R.X. Barros, P.C. Becker, D. Morris, B. Deveaud, A. Regreny, F. Beisser, *Physical Review B*, 47, 10951 (1993).

A SELF-CONSISTENT SOLUTION OF ONE-DIMENSIONAL SCHRODINGER-POISSON EQUATIONS BY NEWTON-RAPHSON ITERATION TECHNIQUE

S. Subramanian*
Department of Electrical and Computer Engineering, Oregon State University, Corvallis, OR 97331
* On leave of absence from the Tata Institute of Fundamental Research, Bombay, India.

ABSTRACT

A self-consistent, one-dimensional solution of Schrodinger-Poisson equations is obtained by Newton-Raphson iteration technique using a finite-difference method with a non-uniform mesh. The method is applied to the simulation of a Schottky barrier placed on the surface of a single quantum well structure. The capacitance of the structure is calculated directly from the solution of the a.c. Poisson equation. The simulated apparent carrier profile is shown to be in excellent agreement with the experimental carrier profile obtained by C-V measurements on a GaAs/InGaAs/GaAs strained layer single quantum well structure.

I INTRODUCTION

A number of papers in the literature deal with the self-consistent solution of one-dimensional Schrodinger-Poisson equations [1-4]. The most commonly used technique to obtain the required self-consistent solution is the so-called relaxation method. In this method, a trial potential distribution $V_{in}(x)$ is used to solve the Schrodinger equation and the resulting wavefunctions and the energy eigenvalues are used to calculate the charge density in the quantum well regions. Poisson equation is then solved using this charge density to get the new potential distribution $V_{out}(x)$. For the solution of Schrodinger equation in the next iteration, a linear combination of $V_{in}(x)$ and $V_{out}(x)$ given by

$$V'_{in}(x) = V_{in}(x) + f(V_{out}(x)-V_{in}(x)) \qquad ...(1.1)$$

is used as the input potential, where f is known as the relaxation factor. The whole procedure is repeated until convergence, viz. the input and the output potentials $V_{in}(x)$ and $V_{out}(x)$ are equal within the desired accuracy. A small value of f usually gives a safe but slow convergence. A number of methods have been described for the choice of f which give faster convergence [1]. Yet, the convergence of relaxation method is generally slow. In this paper, we describe an alternative method of obtaining the self-consistent solution by solving the Poisson equation by Newton-Raphson (NR) iteration technique. The application of NR method for the solution of non-linear Poisson equation when the charge density can be expressed explicitly as a function of the potential is well known. This paper basically extends this method to the case when the charge density calculation requires the solution of Schrodinger equation.

II NEWTON-RAPHSON METHOD

Newton-Raphson method in N dimension deals with the solution of N non-linear equations

$$F_i(x_1,x_2,.....x_N) = 0, \quad i=1,2,.....N. \qquad ...(2.1)$$

In the vector notation, eq.(2.1) is written as

$$\mathbf{F(x)} = 0 \qquad ...(2.1a)$$

where x and F stand for the vectors $(x_1,x_2,....x_N)$ and $(F_1,F_2,....F_N)$, respectively. Eq. 2.1a is solved by starting with a trial vector x_0, and adding to it a correction vector δx given by

$$\mathbf{F(x_0+\delta x) = F(x_0) + J.\delta x = 0} \qquad ...(2.2)$$

or,

$$\mathbf{J.\delta x = -F(x_0)} \qquad ...(2.2a)$$

where the elements of the Jacobian matrix **J** are given by

$$J_{ij} = \partial F_i / \partial x_j. \qquad ...(2.3)$$

The matrix equation 2.2a is solved by standard LU decomposition and the solution vector is found by

$$x_{new} = x_{old} + \delta x. \qquad ...(2.4)$$

III BASIC EQUATIONS AND THE ITERATIVE PROCEDURE

In this section, we summarize the basic equations and the iterative procedure that we used for the self-consistent solution. The one-dimensional Schrodinger equation is :

$$-\hbar^2/2m^*(x)\, d^2y(x)/dx^2 + V(x)y(x) = E\, y(x) \qquad ...(3.1)$$

where y(x) is the wavefunction, E is the energy eigen value, m^* is the effective mass and V(x) is the potential energy. With a given potential energy distribution V(x), eq. 3.1 is solved numerically to get the energy eigenvalues E_k and the corresponding wavefunctions $y^k(x)$. The electron density distribution in the quantum well region is given by

$$n(x) = m^*/\pi\hbar^2 \, \Sigma_k \, kT \ln(1+\exp((E_F-E_k)/kT)) \mid y^k(x) \mid^2 \qquad ...(3.2)$$

where the summation (Σ_k) is carried out over all the allowed bound states in the well and E_F is the Fermi energy. The electron density distribution in the regions outside the quantum well is given by the conventional 3-D Fermi-Dirac statistics, viz.

$$n(x) = 2/\sqrt{\pi}\, N_c\, F_{1/2}\, [(E_F-V(x))/kT] \qquad ...(3.3)$$

where N_c is the effective density of states in the conduction band, and $F_{1/2}$ is the Fermi-Dirac integral. The Poisson equation in one-dimension is

$$d^2\Phi(x)/dx^2 = -\, \rho(x)/\varepsilon(x) \qquad ...(3.4)$$

where Φ is the electrostatic potential, $\varepsilon(x)$ is the dielectric constant, and the charge density $\rho(x)$ is given by

$$\rho(x) = q(N_d^+(x) - n(x)). \qquad ...(3.5)$$

In eq. 3.5, $N_d^+(x)$ and n(x) are the ionized donor density and the electron density distribution, respectively. The Poisson equation 3.4 is solved numerically by Newton-Raphson method as described in section **IV**. The potential energy V(x) for the solution of Schrodinger equation in the next iteration is calculated from the electrostatic potential $\Phi(x)$ through the relation

$$V(x) = -q\Phi(x) + E_g(x) \qquad ...(3.6)$$

where E_g is the bandgap of the semiconductor. In eq. 3.6, the exchange-correlation term is neglected and the zero of energy is taken to be the valence band maximum. The whole procedure is repeated until the potential energy distribution V(x) is equal for successive iterations, within a desired accuracy.

IV NUMERICAL SOLUTION OF POISSON EQUATION

In order to apply Newton-Raphson (NR) method for the solution of Posisson equation, we first rewrite eq. 3.4 in the form of eq.2.1 using a three point finite difference scheme and a non-uniform mesh:

$$F_i = R_i\Phi_{i-1} - (1+R_i)\Phi_i + \rho_i/\varepsilon_i h_i^2 R_i(1+R_i)/2 + \Phi_{i+1} = 0, \quad i=1...N, \qquad ...(4.1)$$

where h_i is the width of the i'th grid and $R_i=h_{i+1}/h_i$. As described in section **II**, eq. 4.1 can be solved by starting with a trial vector Φ^0, and adding a correction vector given by the solution of the matrix equation

$$\mathbf{J}.\delta\Phi = -\mathbf{F}. \qquad ...(4.2)$$

The Jacobian matrix **J** is seen to be of the familiar tridiagonal form with elements given by

$$J_{ij} = \partial F_i / \partial \Phi_j \qquad ...(4.3)$$

$$= R_i, \qquad j = i-1, \qquad ...(4.3a)$$

$$= 1, \qquad j = i+1, \qquad ...(4.3b)$$

$$= -(1+R_i)+h_i^2 Ri(1+R_i)/2\varepsilon_i \; d\rho_i/d\Phi_i, \qquad j=i, \qquad ...(4.3c)$$

$$= 0, \qquad \text{otherwise.} \qquad ...(4.3d)$$

The $d\rho_i/d\Phi_i$ term required in eq.4.3c is obtained by differentiating eq.3.5:

$$d\rho_i/d\Phi_i = q\{dN_d^+(\Phi_i)/d\Phi_i + dn(\Phi_i)/d\Phi_i\}. \quad ...(4.4)$$

The first term in eq.4.4 is easily evaluated all points. The second term in eq.4.4 is also easily evaulated at points outside the quantum well region. In order to evaluate the second term of eq.4.4 in the quantum well region, we make the approximation that the wavefunctions and the energies of the bound states with respect to the bottom of the well do not change for a small change in potential $d\Phi$. Our numerical experience shows that this approximation is generally quite valid. Thus, assuming

$$d|\, y_i^k\,|^2/d\Phi_i = 0, \text{ and } d(E_k - E_F) = -qd\Phi_i, \quad ...(4.5)$$

eq.3.2 can be differentiated to give

$$dn(x)/d\Phi_i = (m^*/\pi\hbar^2)\,\Sigma_k\, 1/(1+\exp((E_k-E_F)/kT))\,|\, y_i^k\,|^2 \quad ...(4.6)$$

in the quantum well region. Thus, $d\rho_i/d\Phi_i$ is evaluated at all points including the quantum well region, and hence all the elements of the Jacobian matrix can be calculated using eqs. 4.3, and the correction vector $\delta\Phi$ can be found by solving eq. 4.2. The new potential for the solution of Schrodinger equation in the next iteration is given by

$$\Phi_{new} = \Phi^0 + \delta\Phi, \quad ...(4.7)$$

and the whole procedure can be iterated.

V NUMERICAL SOLUTION OF SCHRODINGER EQUATION

Schrodinger equation 3.1 is discretized using a three point difference scheme and using a non-uniform mesh:

$$-\hbar^2/2m^*\,\{R_i y_{i-1} - (1+R_i)y_i + y_{i+1}\}/(h_i^2 R_i(1+R_i)/2) + V_i\, y_i = E\, y_i, \qquad i=1,2,.....N \quad ...(5.1)$$

where, once again, h_i is the width of the i' th grid and $R_i = h_{i+1}/h_i$. Eq. 5.1 can be written in the matrix form

$$\mathbf{H}.\mathbf{y} = E\,\mathbf{y} \quad ...(5.2)$$

where the Hamiltonian matrix $\mathbf{H}$ has the tridiagonal form given by

$$H_{ij} = -(\hbar^2/2m^*)/\{h_i^2 R_i(1+R_i)/2\}R_i, \quad j=i-1, \quad ...(5.3a)$$

$$= -(\hbar^2/2m^*)/\{h_i^2 R_i(1+R_i)/2\}, \quad j=i+1, \quad ...(5.3b)$$

$$= V_i + (1+R_i)\,(\hbar^2/2m^*)/\{h_i^2 R_i(1+R_i)/2\}, \quad j=i, \quad ...(5.3c)$$

$$= 0, \quad \text{otherwise}, \quad ...(5.3d)$$

and $y = (y_1, y_2, ...y_N)$ is a column vector. We used an inverse iteration procedure to solve eq. 5.2. Starting with a trial eigenvalue α_k and a corresponding normalized trial eigenvector $\mathbf{y}^k$, the solution of the following equation

$$(\mathbf{H} - \alpha_k \mathbf{I}).\mathbf{b}^k = \mathbf{y}^k \quad ...(5.4)$$

where $\mathbf{I}$ is the unit matrix, gives an improved eigenvector $\mathbf{b}^k$, and

$$E_k = \alpha_k + 1/(\mathbf{b}^k.\mathbf{y}^k) \quad ...(5.5)$$

gives an improved estimate of the energy eigenvalue. The procedure is iterated by normalizing $\mathbf{b}^k$ and using in place of $\mathbf{y}^k$ until E_k and $\mathbf{y}^k$ are determined to the desired accuracy. The whole procedure is repeated until the eigenvalues and the eigenvectors of all the possible bound states in the quantum well are obtained.

VI CALCULATION OF CAPACITANCE

We calculate the capacitance of the structure by directly solving the a.c. Poisson equation [5]. The change in the electrostatic potential distribution for a small change δV in the bias across the device is given by the solution of the a.c. Poisson equation (obtained by differentiating eq. 3.4)

$$d^2u(x)/dx^2 = -(d\rho(x)/d\Phi)u(x)/\varepsilon(x) \quad ...(6.1)$$

where $u(x) \equiv d\Phi(x)/dV$. From Gauss's law, the total charge per unit area can be obtained from the surface

electric field:

$$Q = -\varepsilon \, d\Phi/dx \, |_{x=0} \qquad ...(6.2)$$

The small signal capacitance is given by

$$C = dQ/dV = -\varepsilon \, du(x)/dx \, |_{x=0} \qquad ...(6.3)$$

Thus, by solving eq. 6.1 using the values of $d\rho(x)/d\Phi$ obtained in section IV, and using eq. 6.3 the capacitance of the structure is easily determined.

VII RESULTS AND DISCUSSION

The method was applied to the simulation of a strained GaAs/InGaAs/GaAs single quantum well structure grown by metal-organic chemical vapor deposition (MOCVD). A Schottky barrier was assumed to be placed on the surface of the sample and the simulations were carried out to calculate the capcitance of the structure for different voltages applied to the Schottky contact with respect to the bottom GaAs layer. From the calculated capacitance-voltage (C-V) data, the apparent carrier concentration profile was calculated using the standard relations,

$$\hat{n}(x) = 2/(q\varepsilon A^2 dC^{-2}/dV), \text{ and } x = \varepsilon A/C, \qquad ...(7.1)$$

where x is the depth below the Schottky barrier plane, and A is the area of the device. The only fitting parameter used for the simulation was the band offset ΔE_c. All the other parameters required for the simulation, such as the quantum well width, the nominal Indium composition in the well, etc. were estimated from the growth conditions. Figure 1 shows a comparison between the simulated and the experimental carrier profiles. To demonstrate the sensitivity of the simulated profile to the value of the fitting parameter ΔE_c, the simulations are shown for two different values of ΔE_c. Excellent agreement between the simulated and the measured profiles is obtained for $\Delta E_c = 0.165$ eV. We have also developed a modified Kroemer's analysis for the determination of ΔE_c directly from the experimental carrier profile [6]. From this analysis we get $\Delta E_c = 0.162$ eV in excellent agreement with the value obtained from simulation.

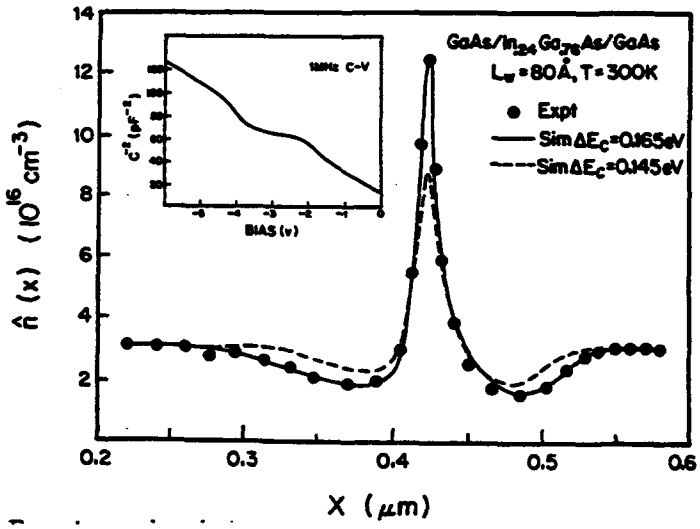


Fig1. Experimental and simulated carrier profiles of GaAs/InGaAs/GaAs strained SQW structure.

REFERENCES

1. F. Stern, J. Computational Physics **6**, 56 (1970).
2. M.B. Patil and H. Morkoc, Solid St. Electron. **33**, 99 (1990).
3. M.B. Patil and U. Ravaioli, Solid State Electron. **33**, 953 (1990).
4. A.M. Cruz Serra and H.A. Santos, J. Appl. Phys. **70**, 2734 (1991).
5. S. Subramanian, J. Appl. Phys. **64**, 1211 (1988).
6. S. Subramanian, et. al., J. Appl. Phys. **74**, 7618 (1993).

NUMERICAL SIMULATION OF THE EFFECT OF SURFACE CHARGES ON ELECTRON CONFINEMENT IN QUANTUM DOT STRUCTURES *

Minhan Chen and Wolfgang Porod,
Department of Electrical Engineering,
University of Notre Dame,
Notre Dame, IN 46556

Abstract

We present numerical simulations of electron confinement in gated AlGaAs/GaAs quantum dot structures. The confining quantum dot potentials are obtained from solutions of the axisymmetric Poisson equation. Our model takes into account the effect of surface states by viewing the exposed surface as the interface between the semiconductor and the dielectric. We investigate the confining potentials and the dot occupation as a function of different physical models for surface states at the exposed semiconductor surface.

I. INTRODUCTION

In recent years, fabrication techniques have made possible confinement of a two-dimensional electron gas into wires or dots where quantum effects are significant. Typically, such device structures are defined by metallic gates. In order to model electronic confinement in these structures, one has to solve the Poisson and the Schrödinger equations. For solving the Poisson equation, one needs to specify boundary conditions for the potential and/or flux at the exposed semiconductor surface. This is a crucial problem, especially in quantum devices where the confined electrons reside close to the surface. In previous work [1], we have shown that different choices for the boundary conditions at the exposed surface result in noticeable differences for the confining potentials. Highly accurate models of the potential or dielectric flux variation on the exposed semiconductor surface will be needed to realize recently proposed computing architectures for quantum devices, so called Quantum Cellular Automata, which consist of cells of coupled quantum dots in the few electron regime [2].

In our formulation [1], we view as the natural problem domain both the semiconductor and the dielectric, as schematically shown in Fig. 1(a). Thus the usual Dirichlet or Neumann boundary conditions at the exposed semiconductor surface are replaced by more physical matching conditions at the interface between the semiconductor and the dielectric. We assume that the potential is continuous across this interface and that the jump in the normal dielectric flux density is equal to the surface/interface charge density, Q_{int}, which is determined by microscopic models for surface/interface states. We apply our coupled finite-element/boundary-element (FBEM) algorithm to quantum dot structures with axisymmetry. The numerical formulation of the problem is developed in Sec. II and numerical results are presented in Sec. III.

II. PROBLEM FORMULATION

1. Problem Statement

A model quantum dot structure with axisymmetry is shown in Fig. 1. In the semiconductor domain, a quantum dot is realized at the AlGaAs/GaAs heterojunction and is defined by applying a sufficiently nega-

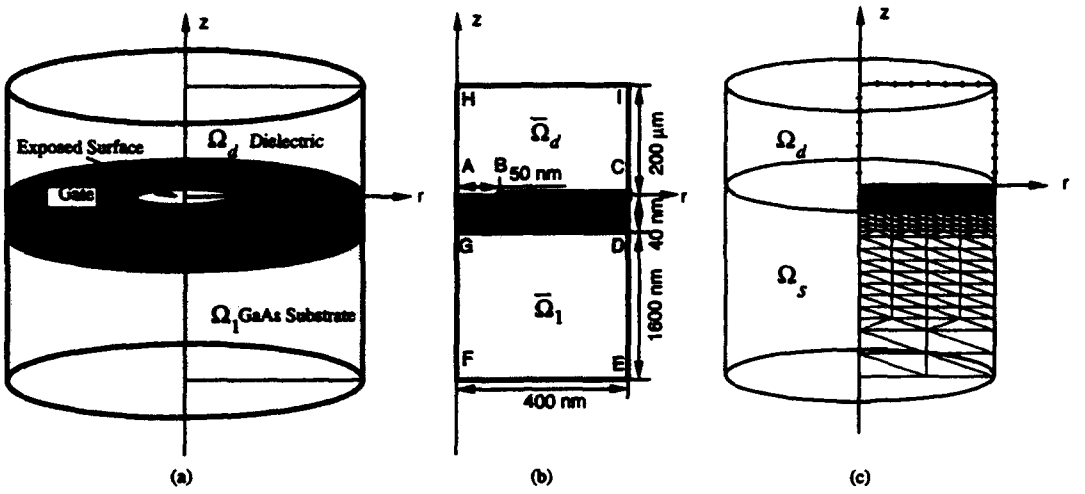


Figure 1: The model quantum dot heterostructure. Fig. 1(a) shows the whole problem domain with axial symmetry which consists of both the semiconductor (Ω_1 and Ω_2) and the dielectric (Ω_d) regions. Fig. 1(b) indicates the two dimensional generating areas and boundaries with typical dimensions. Fig. 1(c) shows the FBEM mesh, which is dense inside the semiconductor region and only consists of the discretized boundary surrounding the dielectric region.

tive gate voltage V_G to the metal gate on the top surface, which contains a circular opening thus exposing the semiconductor surface. For axial symmetry, Poisson's equation can be written in cylindrical coordinates (r,θ,z) as,

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\varepsilon\frac{\partial u(r,z)}{\partial r}\right)+\frac{\partial}{\partial z}\left(\varepsilon\frac{\partial u(r,z)}{\partial z}\right) = -f[u(r,z)], \tag{1}$$

where $u = (E_C(\phi) - E_F)/kT$ is a reduced variable which measures the separation between the conduction band edge and the Fermi level in units of the thermal energy kT, $f = e\rho/kT$ is the charge density term in the semiconductor, ε is the dielectric constant, and m^* is the effective mass. The generating domains and boundaries are shown in Fig. 1(b).

Equation (1) is a boundary value problem. We solve it by our FBEM algorithm [1], which is a combined finite element method (FEM) for the semiconductor domain and a boundary element method (BEM) for the dielectric region. For the semiconductor domain Ω_s, with $\Omega_s = \Omega_1 \cup \Omega_2$, the standard FEM discretization of equation (1) results in the following non-linear system of equations,

$$\begin{aligned} \boldsymbol{K}_{11}\boldsymbol{u}_o^s + \boldsymbol{K}_{12}\boldsymbol{u}_{BA}^s &= \boldsymbol{P}_f^s, \\ \boldsymbol{K}_{12}^T\boldsymbol{u}_o^s + \boldsymbol{K}_{22}\boldsymbol{u}_{BA}^s &= \boldsymbol{P}_{BA}^s, \end{aligned} \tag{2}$$

where $\boldsymbol{u}_{BA}^s$ and $\boldsymbol{P}_{BA}^s$ contain the potentials and nodal forces at the nodes on the interface $\partial\Omega_{BA}$ between the semiconductor and the dielectric, whereas $\boldsymbol{u}_o^s$ and $\boldsymbol{P}_f^s$ contain the potentials and nodal forces at all other nodes inside the semiconductor domain, and $\boldsymbol{K}$ is the stiffness matrix.

The dielectric domain, Ω_d, is a charge free region. The governing equation is Laplace's equation. Since the fundamental solution of Laplace's equation is known, a boundary integral equation technique can be employed. With the known three dimensional fundamental solution of Laplace's equation in cylindrical coordinates and its associated dielectric flux density [3], the boundary contour $\partial\bar{\Omega}_d$ can be calculated explicitly in terms of complete elliptic integral of the first and second kind, $K(m)$ and $E(m)$, respectively. The resultant system of equations can be expressed as,

$$\begin{aligned} \boldsymbol{S}_{11}\boldsymbol{u}_o^d + \boldsymbol{S}_{12}\boldsymbol{u}_{BA}^d &= \boldsymbol{P}_o^d, \\ \boldsymbol{S}_{21}\boldsymbol{u}_o^d + \boldsymbol{S}_{22}\boldsymbol{u}_{BA}^d &= \boldsymbol{P}_{BA}^d, \end{aligned} \tag{3}$$

where S is the equivalent stiffness matrix, and P^d is the equivalent nodal force vector.

The matching conditions at the exposed surface [1] are given in discretized form by,

$$u^s_{BA} = u^d_{BA} = u_{BA} \quad \text{and} \quad P^s_{BA} + P^d_{BA} = \frac{e}{kT} Q_{int}, \tag{4}$$

where $Q_{int} = Q_{int}(u_{BA})$ is the nodal charge density on the exposed semiconductor surface. A global system of equations is formed by coupling the semiconductor, equation (2), to the dielectric, equation (3), while enforcing the matching conditions (4),

$$\begin{bmatrix} S_{11} & S_{12} & 0 & 0 \\ S_{21} & S_{22} & 0 & I \\ 0 & K_{12} & K_{11} & 0 \\ 0 & K_{22} & K_{12}^T & -I \end{bmatrix} \begin{bmatrix} u^d_o \\ u_{BA} \\ u^s_o \\ P^s_{BA} \end{bmatrix} = \begin{bmatrix} P^d_o \\ Q_{int} \\ P^s_f \\ 0 \end{bmatrix}. \tag{5}$$

Solution of this set yields the potential distribution in the semiconductor and dielectric domains, including the interface $\partial\Omega_{BA}$, and the nodal flux on $\partial\Omega_{BA}$.

2. Interface Charge Density on the Exposed Semiconductor Surface

In order to solve the above problem, the surface charge density, Q_{int}, must be given to specify the matching condition (4). It is known that surface states lying within the energy band gap play a dominant role in the surface charge Q_{int} on exposed semiconductor surfaces [4-6].

In this model, the characters of the surface states are assumed to be acceptor-, D^s_A, and donor-, D^s_D, like. Typical energy distributions of surface states are shown in Fig. 2 [4,5]. The semiconductor surface charge density, as a function of surface potential u_{BA}, is given by Fermi-Dirac statistics with an appropriate quasi-Fermi level for cases of applied bias,

$$Q_{int}(u_{BA}) = -e\int_{E_V}^{E_C} \frac{D^s_A}{1+exp\,[\,(E-E^s_F)/kT]}\,dE + e\int_{E_V}^{E_C} \frac{D^s_D exp\,[\,(E-E^s_F)/kT]}{1+exp\,[\,(E-E^s_F)/kT]}\,dE, \tag{6}$$

The strong non-linearity of the surface charge as a function of the potential may cause numerical convergence problems, particularly for low temperatures. We implemented a modified Bank-Rose damping scheme [7] to stabilize the convergence by adaptive underrelaxation and to accelerate the convergence speed of Newton's method. The combination of our coupled finite-element/boundary-element algorithm and the adaptive damping scheme performs quite satisfactorily in our numerical examples.

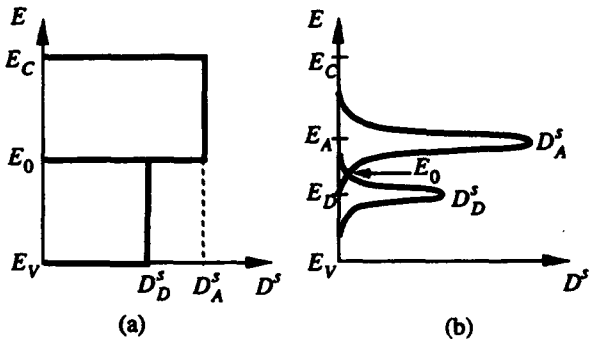


Figure 2: Typical energy distributions of interface states across the semiconductor bandgap; (a) uniform, and (b) localized distributions.

III. NUMERICAL RESULTS

An example heterostructure is shown in Fig. 1 with its relevant physical dimensions indicated. In the

semiconductor domain, a quantum dot is realized at the AlGaAs/GaAs heterointerface by applying a sufficiently negative gate voltage V_G to the patterned metal gate on the top surface. The n-type doping density is assumed to be $10^{18} cm^{-3}$ in the AlGaAs layer and $10^{15} cm^{3}$ in the GaAs substrate. We assume both uniform and *gaussian* localized energy distributions of surface states across the semiconductor bandgap, as shown in Fig. 2. The characters of the surface states are assumed to be acceptor-, D_A^s, and donor-,D_D^s, like. A semi-classical Thomas-Fermi charge model is assumed in the semiconductor domain.

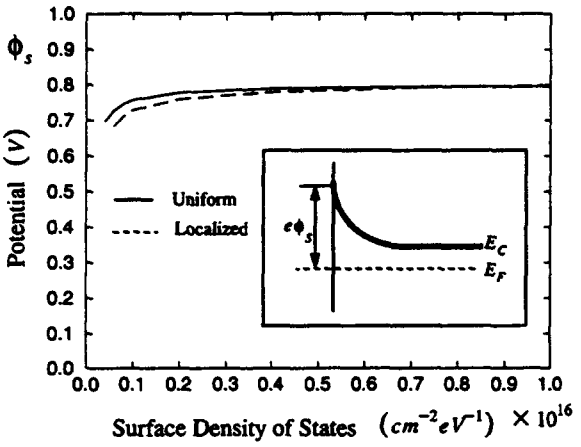


Figure 3: Surface potential ϕ_s on the surface of AlGaAs as a function of the surface density of states for both uniform and gaussian localized energy distributions. Here, $D_A^s = D_D^s$ is assumed.

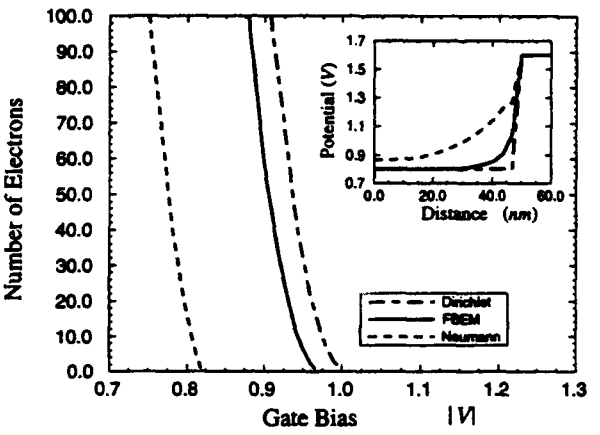


Figure 4: Comparison of the number of confined electrons as a function of gate bias for the three types of boundary conditions on the exposed semiconductor surface. The inset shows surface potential profiles on the semiconductor surface for -0.8 *V* gate bias.

As shown in Fig. 3, the semiconductor surface potential, ϕ_s, varies with the surface density of states for low defect densities on the surface. For higher surface density of states, however, the surface potential ϕ_s saturates, and the surface Fermi level is then pinned at or near to the energetic position of the neutral level, E_0, shown in Fig. 2. This pinning behavior is observed for both uniform and localized energy distributions of surface states.

Figure 4 presents a comparison of the number of confined electrons as a function of gate bias for different boundary conditions at the exposed semiconductor surface, namely the more conventional Dirichlet and Neumann boundary conditions, and our FBEM matching technique. The parameters and dimensions of the quantum dot structure are the same as those given in Fig. 1. We see that the different formulations produce significantly different results. Specifically, the Dirichlet boundary condition produces a significantly higher number of electrons than the FBEM algorithm, and the Neumann boundary condition produces a much lower quantum dot occupation.

*Supported in part by ARPA/ONR and AFOSR.

[1] M. Chen, W. Porod, and D. J. Kirkner, *J. Appl. Phys.* **75**, 2545 (1994).
[2] C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, *Nanotechnology* **4**, 59 (1993).
[3] C. A. Brebbia, J. C. F.Telles, and L. C. Wrobel, *Boundary Element Techniques* (Springer, Berlin, 1984).
[4] A. M. Cowley and S. M. Sze, *J. Appl. Phys.* **36**, 3212 (1965).
[5] W. E. Spicer, I. Lindau, P. Skeath, and Y. Su, *J. Vac. Sci. Technol.* **17**, 1019 (1980).
[6] W. Mönch, Surf. Sci. **132**, 92 (1983).
[7] R. E. Bank and D. J. Rose, *SIAM J. Numer. Anal.* **17**, 806 (1980).

A LINEAR EIGENVALUE METHOD FOR CALCULATING THE POSITIONS OF TRANSMISSION POLES AND ZEROS IN RESONATOR STRUCTURES

Zhi-an Shao, Wolfgang Porod, and Craig S. Lent
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

Abstract

We present a numerical technique which yields, as the solutions of a linear eigenvalue problem, the positions of transmission poles and zeros in resonator structures with arbitrary potential profiles. We present several examples to demonstrate the utility of this numerical technique.

I. INTRODUCTION

A common computational problem is to find the quasi-bound states of resonant transmitting systems. For an isolated bound system, because of the zero wavefunction boundary conditions, the Hamiltonian of the system is Hermitian, hence the system has only bound states. However, for an open unbound system, because the wavefunctions at the boundary are non-zero, the complex boundary condition may lead the Hamiltonian of the system non-Hermitian, hence the system possesses quasi-bound states for resonant transmission [1]. In general, to find the quasi-bound states of a given system with scattering boundary conditions requires to search for the zeros of an energy-dependent matrix determinant [2, 3].

In this paper, we use another approach to solve this problem. Based on the quantum transmitting boundary method (QTBM) and a finite element discretization [4], we present an eigenvalue algorithm which yields the positions of the transmission poles. We can also use this algorithm to calculate the positions of transmission zeros in quantum waveguide systems [5].

II. APPROACH

In general, a transmission problem shown in Figs. 1(a) and 2(a) may be formulated as an inhomogeneous problem, $Au=\alpha P$. Here, A is an energy-dependent coefficient matrix, u is the unknown wavefunction, and αP is the source flux. Specifically, α can be either the incoming amplitude, i(E) in figure 1(a), or the transmission amplitude, t(E) in figure 2(a), and P is an energy-dependent vector. For a given source flux αP, the solution of the inhomogeneous system is uniquely determined. We can also force the source flux $\alpha P=0$, as shown in figures 1(b) and 2(b), which results in a homogeneous problem, $Au=0$. This is, in general, a nonlinear eigenvalue problem. Using the finite element discretization, furthermore, results in a linear eigenvalue problem.

For the transmission problem, shown in Fig. 1(a), Schrödinger's equation can be written as the following inhomogeneous system, where all matrices are constant and the energy dependence is shown explicitly,

$$(\mathbf{H} - E\,\mathbf{Q} + k_L \mathbf{B_L} + k_R \mathbf{B_R})\, \psi = i(E)\, k_L\, \mathbf{p} \ . \qquad (1)$$

Here, i(E) is the amplitude of the forcing incoming flux at energy E. The wavenumbers at the left and right boundaries of the system are k_L and k_R, respectively, which are related through the external bias V_{bias} by, $k_R^2 - k_L^2 = (2m^* eV_{bias})/\hbar^2$; all symbols have their usual meaning. The bound state problem is contained in the above system as $(\mathbf{H} - E\,\mathbf{Q})\,\psi = 0$, and the matrices $\mathbf{B_L}$, $\mathbf{B_R}$, and $\mathbf{p}$ arise due to the open boundaries.

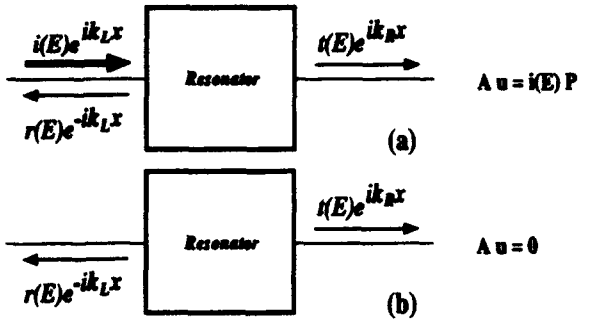


Figure 1. Schematic diagram of a resonant structure with a forcing incoming flux (thick arrow). (a) shows an incident wave from the left (source) with its transmitted and reflected components, which results in an inhomogeneous problem; (b) setting the incident wave (source) to zero, leads to an eigenvalue problem. Its solutions give us the quasi-bound states of the system, or the positions of the transmission poles.

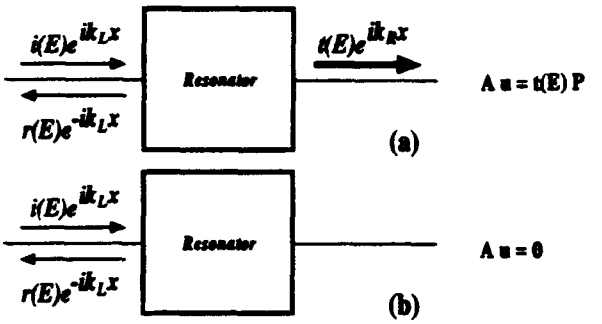


Figure 2. Schematic diagram of a resonant structure with a forcing transmitted flux (thick arrow). (a) shows an incident wave from the left with its transmitted (source) and reflected components, which results in an inhomogeneous problem; (b) setting the transmitted wave (source) to zero, leads to an eigenvalue problem. Its solutions give us the positions of the transmission zeros.

Forcing the incoming flux to zero, i(E)=0 as shown in Fig. 1(b), produces the decaying quasi-bound states of the system. Equation (1) becomes a polynomial eigenvalue problem of degree p=2 for an unbiased system ($V_{bias} = 0$ and $k_L=k_R$) and of degree p=4 for a biased system. In the latter case, we perform the following transformations, $k_R=\bar{k}+\Delta$ and $k_L=\bar{k}-\Delta$, with $\Delta=(m^* eV_{bias})/(2\hbar^2\bar{k})$. This leads to a fourth-order polynomial eigenvalue problem in $\bar{k}$,

$$(A_0 + \bar{k} A_1 + \bar{k}^2 A_2 + \bar{k}^3 A_3 + \bar{k}^4 A_4) \psi = 0 \ , \qquad (2)$$

where the above A's are related to the matrices in equation (1). The polynomial eigenvalue problems of degree p can be rearranged into linear eigenvalue problems with p times the original matrix size. Since the resulting matrix is not Hermitian in this case, the eigenvalues are located in the complex-energy plane. The real and imaginary parts of these eigenvalues correspond to the energies and lifetimes of the quasi-bound states of the resonant transmission system.

The transmission problem may also be viewed as one in which the resonant structure is forced to yield a certain transmitted amplitude t(E), as schematically shown in Fig. 2(a). In this case, the required incident and reflected amplitudes are the unknowns. Using the boundary condition $\psi(x_R) = t(E) \exp(ik_R x_R)$ at the right edge x_R of the system, we may re-write equation (1) in a form where only terms proportional to t(E) appear on the right-hand-side. Terms proportional to the incident amplitude i(E) appear on the left-hand-side, and i(E) now is part of the solution vector ψ which contains the unknowns.

Forcing the transmitted flux to zero, t(E)=0 as shown in figure 2(b), produces the transmission zeros. It can be shown that the corresponding eigenvalue problem is linear in the energy, and has the form,

$$(H' - E Q') \psi' = 0 \ , \qquad (3)$$

where the matrices H' and Q' are related to the corresponding ones in (1). For t-stub systems, furthermore, it can be shown that H' is also Hermitian. As a consequence, the eigenvalues in this case, which are the energies of the transmission zeros, always occur on the real-energy axis. This result is consistent with our previous scattering matrix investigations, where we also proved that transmission zeros always occur on the real-energy axis [5].

III. EXAMPLES

We now present several examples to demonstrate the utility of our approach. First, we apply our method to a multi-barrier resonant-tunneling structure with applied external bias. Next, we locate the positions of transmission poles and zeros in quantum waveguide systems, which include t-stub and loop structures. We compare the results of our direct eigenvalue method to the more conventional method of searching in the complex-energy plane for the zero of the system determinant.

1. Multi-Barrier Resonant-Tunneling Structure with Applied Bias

As our model system, we consider a 10-barrier resonant-tunneling structure in a uniform electric field of $\mathcal{E}$=150 kV/cm. The barrier width and height are 1.4 nm and 5.0 eV, respectively, and the well width is 4.9 nm. For the finite element discretization, we use an average mesh size of 0.7 nm for the numerical calculation, which yields matrices of dimension 92 in equation (1). We choose the middle of the structure as the zero point of the potential.

Applying our eigenvalue method to this structure, we obtained the energies of the quasi-bound states, which are the poles of the transmission amplitude in the complex-energy plane. It is well known that no transmission zeros exist in this case. It is an easy matter to numerically obtain the eigenvalues of the linear system (2) with dimension 368. The results are plotted in Fig. 3, and the numerical values for the real and imaginary parts of the poles are given in tabular form. The horizontal lines indicate the computed spatial electron densities in each quasi-bound state. The formation of minibands is evident, which are derived from the individual states in each well. The imaginary part of each pole gives the inverse of the lifetime for the corresponding quasi-bound state. As one would expect, the longest-lived states are concentrated in the middle of the structure, and states toward the edges are more "leaky." Note that the imaginary parts vary by many orders of magnitude. This makes a direct search for the locations of the poles in the complex-energy plane very costly since a very fine mesh has to be used in order to avoid missing poles. In contrast, our direct method yields the energies of all poles, without any search, as the solutions of a linear eigenvalue problem.

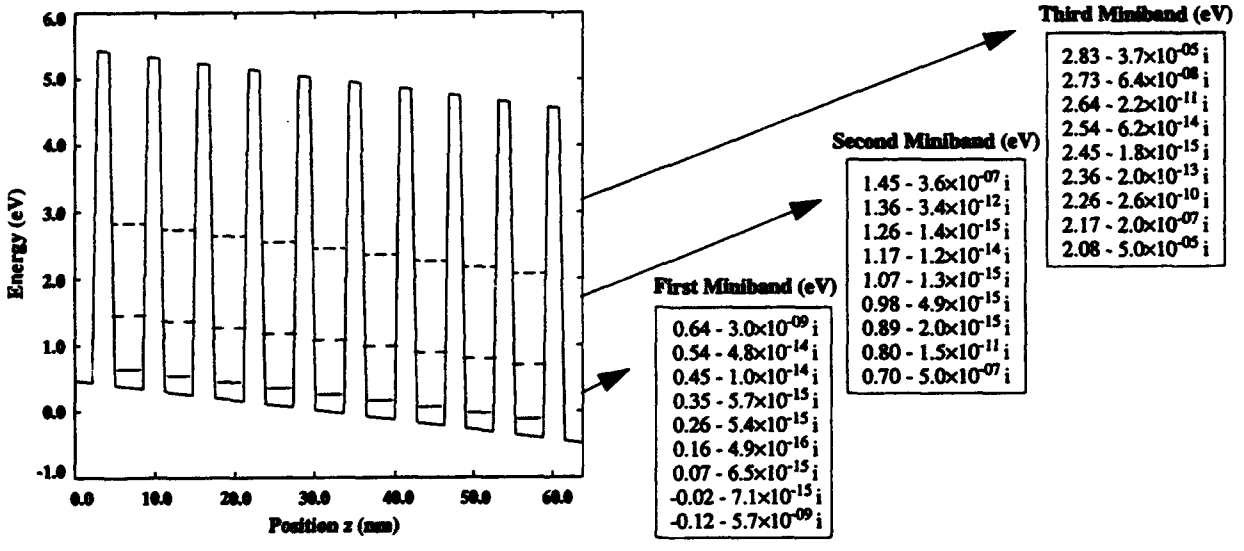


Figure 3. The quasi-bound states of a multi-barrier resonant tunneling structure in a uniform electric field. The states are plotted as horizontal lines at the real energy of the resonance, and the lines are drawn for those positions at which the absolute value of the wavefunction is larger than a threshold value. The real- and imaginary-parts of the resonances in each miniband are also given.

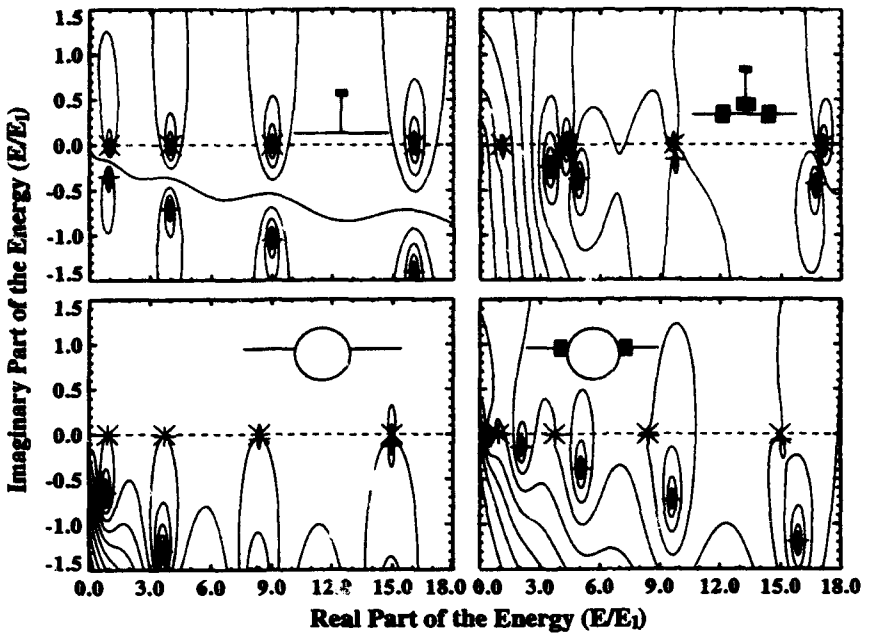


Figure 4. Shown are contour plots of the absolute value of the transmission amplitude for t-stub and loop structures, which are schematically shown in the insets. The '+' and 'x' symbols represent the positions of transmission poles and zeros, respectively, which were calculated by our direct eigenvalue method. The energy of the first standing wave in the stub (E_1=56.2 meV) is used as the unit of energy. The results obtained by both methods agree very well.

2. Quantum Waveguide Structures

We choose t-stub and loop structures as our model systems, which are schematically shown in the insets of Fig. 4. The solid lines represent the waveguides which are transmission channels. The shaded boxes represent tunneling barriers (0.5 eV high and 1 nm thick) and the full filled box terminates the stub. For the t-stub structures, the length of the stub is 10 nm and the distance between two tunneling barriers on the main transmission channel is 4 nm. For the asymmetrical loops shown here, the lengths of the two arms are 10 and 11 nm, respectively. Spatial mesh dimensions of 0.2 nm are used in the numerical calculations.

It is well known that these systems possess both transmission poles and zeros [5]. The contour lines in Fig. 4 represent the absolute value of the transmission amplitude in the complex-energy plane, which is obtained from a solution of the inhomogenoues problem (1). Poles and zeros, which occur on the real-energy axis, are easily discerned. Using the appropriate eigenvalue problem, we also show the directly calculated locations of the transmission poles and zeros which are indicated by the symbols '+' and 'x', respectively. Note the perfect agreement between the two methods. Again, our technique directly yields poles and zeros without a need to search for them in the complex-energy plane.

IV. SUMMARY

We presented a new approach for directly calculating the positions of transmission poles and zeros in resonant transmission structures. In general, a transmission problem is an inhomogeneous problem. Forcing the source flux to zero, for either the incoming wave or the transmitted wave, results in a non-linear eigenvalue problem. Using the finite element method, furthermore, these eigenvalue problems become linear. It is then an easy matter to directly calculate the energies of the transmission poles and zeros.

Acknowledgments: This work was supported in part by ARPA/ONR and AFOSR.

[1] *Quantum Mechanics*, L. D. Landau and E. M. Lifshitz (Pergamon Press, 1962).
[2] W. R. Frensley, Superlatt. and Microstruct. **11**, 347 (1992).
[3] J. H. Luscombe, Nanotechnology **4**, 1 (1993).
[4] C. S. Lent and D. J. Kirkner, J. Appl. Phys. **67**, 6353 (1990).
[5] W. Porod, Z. Shao, and C. S. Lent, Appl. Phys. Lett. **61**, 1350 (1992).

MAGNETIC QUASI-BOUND-STATE INDUCED RESONANT COUPLING OF EDGE STATES

Manhua Leng and Craig S. Lent
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

Abstract

We numerically examine the ballistic transport properties of an electron channel with a single scatterer, an antidot, when a perpendicular magnetic field is present. Formation of magnetic quasi-bound-states (MQBS) is observed in such a structure. The MQBS's couple magnetic edge states, resulting in resonances. In the multiple mode regime, coupling can occur between opposite edge states, resulting in resonant reflection, or anti-resonance. An edge state can also tunnel through the scattering region via an MQBS, resulting in resonant transmission. These resonances are closely related to those observed in quasi-one-dimensional systems such as T structures where such resonances are associated with transmission poles in the complex energy plane.

I. INTRODUCTION

For a quantum channel with an impurity in a perpendicular magnetic field, there exist three types of electronic states: (1) Circulating Landau states in the bulk region of the channel, (2) localized states corresponding to the circulating orbits around the impurity, and (3) extended states corresponding to classical skipping orbits near the channel walls, illustrated in Figure 1. The first type is highly degenerate therefore does not carry net current; the second type forms magnetic bound-states, which do not carry net current either; the third type forms magnetic edge states, which move in opposite directions on opposite walls, carrying net current. For a wide channel in high magnetic field, the impurity potential does not couple edge states on opposite walls, therefore back-scattering is suppressed. This implies unity transmission probabilities for edge states, leading to the integer quantization of Hall resistance [1]. For a narrower channel where the extension of the circulating orbits around the impurity overlaps with the edge states, magnetic quasi-bound-states (MQBS) form. Through the interference of MQBS's, an edge state on the upper wall can be back-scattered to its counterpart on the lower wall, resulting in resonant reflection; or an edge state on the left side of the impurity can tunnel to the same edge state on the right side of the impurity on the same wall, resulting in resonant transmission. This resonance phenomenon was suggested by Jain and Kivelson in a semiclassical calculation [2]. In this paper we present a quantum mechanical calculation of an electron channel with an antidot by using the Quantum Transmitting Boundary Method (QTBM) in which magnetic field is taken into account in the whole device, including the lead regions [3]. The suggested resonances are confirmed in our results.

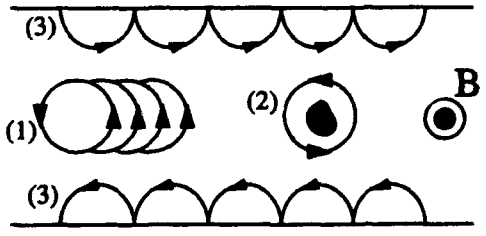


Figure 1. Electronic states in a quantum channel with an impurity. Circulating orbits around the impurity form magnetic bound-states or magnetic quasi-bound states; skipping orbits on the channel walls form magnetic edge states.

II. MODEL AND METHODS

Figure 2 shows the schematic geometry. The channel width is d, the radius of the antidot r. Device domain is marked by the dashed lines with an extension a in $\hat{x}$-direction. We present the results for the particular case where $r/d=1/20$. By including enough evanescent modes in the QTMB calculation, a can be chosen arbitrarily provided that the device region encloses the scatterer (antidot). We adopt the single-band, effective mass model with $m/m^*=0.067$, appropriate for GaAs. Hard wall potentials are assumed to define the channel edges and the antidot area and zero potentials assumed elsewhere in the channel. We choose the vector potential in the Landau gauge, $\vec{A} = -By\hat{x}$. The two-dimensional Schrödinger equation becomes

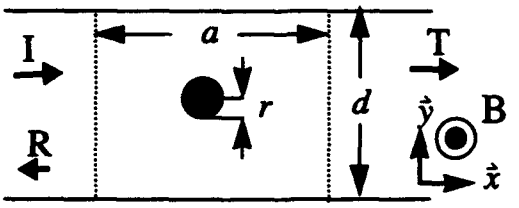


Figure 2. Schematic geometry of a quantum channel with one antidot as a single scatterer in a perpendicular magnetic field.

$$\left(\frac{-\hbar^2}{2m^*}\nabla^2 + \frac{i|e|\hbar By}{m^*}\frac{\partial}{\partial x} + \left(\frac{|e|^2B^2y^2}{2m^*}\right) + V_0(x,y)\right)\psi(x,y) = E\psi(x,y), \tag{1}$$

where $V_0(x,y)$ is the potential in the device region. We employ the QTBM to solve Equation (1) for the scattering state. Boundary conditions are implemented in the QTBM by expanding the scattering state in a lead region as a superposition of the local transverse (including both traveling and evanescent) modes. Transverse modes $\{\varphi_n(y)\}$ are obtained by using the form $\psi_i(x,y) = exp(ik_n x)\varphi_n(y)$ for lead region i. The original Schrödinger equation becomes a quadratic eigenvalue problem for wavevector k_n at given energy E

$$\left(-\frac{\partial^2}{\partial y^2} + \left(\frac{eBy}{\hbar} + k_n\right)^2 + \frac{2m^*}{\hbar^2}V_i(y)\right)\varphi_n(y) = \frac{2m^*}{\hbar^2}E\varphi_n(y). \tag{2}$$

From the full wavefunction solution, complex energy-dependent transmission and reflection amplitudes for each transverse mode are obtained. Then we use the two-terminal Landauer formula to obtain the conductance in the linear response regime, $G = (2e^2/h)Tr(tt^\dagger)$. We also compute the particle current density in the device region from the following definition

$$\vec{j} = \frac{i\hbar}{m^*}(\psi\nabla\psi^* - \psi^*\nabla\psi) + \frac{|e|\vec{A}}{m^*}|\psi|^2. \tag{3}$$

III. RESULTS AND DISCUSSIONS

In Figure 3 we plot the conductance as a function of magnetic field and incident energy. The strength of the field is measured by the parameter $\beta = eBd^2/\hbar = d^2/l_H^2$ where l_H is the magnetic length. Energy is expressed in units of the first bulk Landau level $E_L(\beta) = \hbar\omega_c/2 = \hbar eB/2m^*$. Notice that the energy units $E_L(\beta)$ are different for different fields. In the inset we plot the individual transmission coefficients T_1 and T_2 at β=40 when the incident electrons are in the first edge state and second edge state respectively.

Should there be no scatterer (antidot), the conductance would be a series of smooth platforms with its height corresponding to the number of existing edge channels. In the presence of the antidot, however, scattering of edge states takes place and more structure develops in the conductance. Particularly, the spikes, as in a one-dimensional double barrier device, indicate certain resonant processes. They are evident at high fields in Figure 3, with dips indicating resonant reflection and peaks resonant transmission. When

the field is lowered, both the dips and the peaks are broadened and smoothed and they eventually disappear in very low fields. We now examine the wavefunctions of the electron scattering states for such resonant reflection state and resonant transmission state.

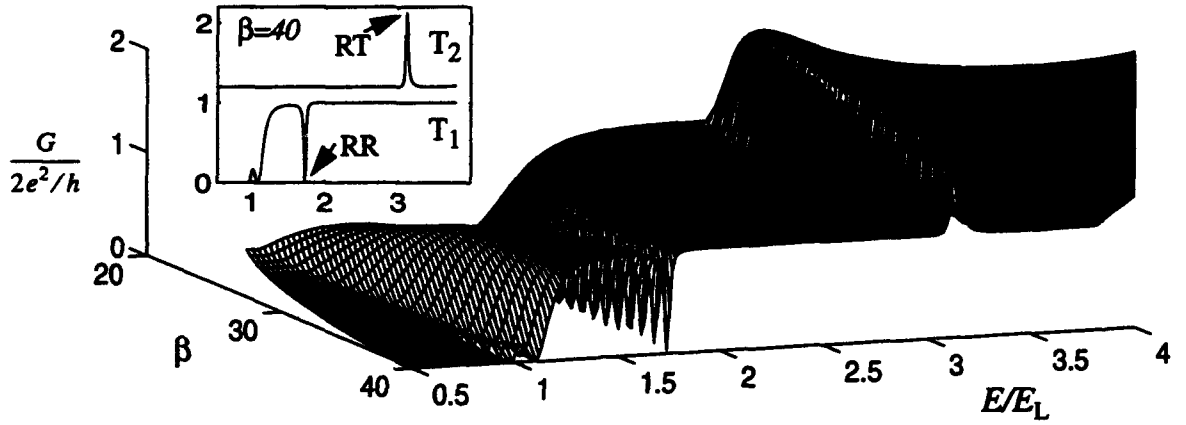


Figure 3. Conductance verses Energy E and magnetic field β. E is expressed in units of the first bulk Landau level $E_L(\beta)=\hbar\omega_c(\beta)/2$. Inset: modal transmission coefficients for the first edge state (T_1) and second edge states (T_2) at magnetic field $\beta=40$. RR is a resonantly reflected state, RT a resonantly transmitted state.

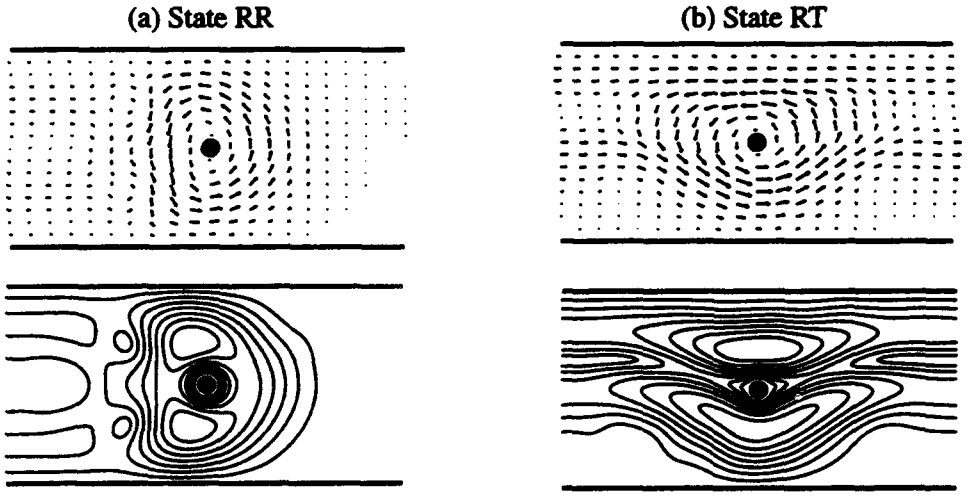


Figure 4. Particle current density (top) and probability density (bottom) distribution. (a) Resonant reflection state RR. (b) Resonant transmission state RT.

In Figures 4 we plot the distribution of the particle current density (vector field) and the probability density (contour lines) in a domain of $a=2d$ for states RR and RT indicated in the inset of Figure 3. In both cases strong circulating orbits are observed around the antidot, forming magnetic quasi-bound states. For state RR, the incident wave is in the first edge mode on the left upper wall; through the MQBS, it scatters to the first edge mode on the left lower wall, resulting in resonant reflection. For state RT, the incident wave is in the second edge mode on the left upper wall; through the MQBS, it tunnels to the second edge mode on the right upper wall, giving rise to resonant transmission. For a 1D double barrier (or T stub) structure, resonant tunneling (or reflection) occurs when the incident energy coincides with the energy levels of the quasi-bound states formed in the potential well (or in the stub); transmission probability on the real energy

axis can be deduced from the positions of reflection zeros (or transmission zeros for T stub) and transmission poles on the complex energy plane, both of which can be obtained by using an eigenvalue technique[4]. For our quasi 1D channel in a magnetic field, resonances are likewise induced by the quasi-bound states in the system, only now they are of magnetic in origin and formed around the antidot. The physical association between resonance and magnetic quasi-bound state is nevertheless evidently shown in our numerical calculation. So, when the magnetic field is decreased, the circulating orbit is held less tightly to the antidot and magnetic quasi-bound states eventually dissolves into the extended states. Interference between edge states due to the scattering potential of the antidot still occurs but resonances disappear.

We further introduce an additional stripe of potential barrier V_0 to the channel and the conductance results at $\beta=40$ are shown in Figure 5 for $V_0/E_L=0, 1, 3, 5$ and $r/d=1/20$ and $w/d=1/20$. In a wave packet calculation, a similar (but wider) structure was used by Müller to illustrate the lack of destructive interference of edge states where the potential barrier was mainly considered as to force the incident wave to split into a tunneling part directly through the potential stripe and a scattered part via the antidot [5]. Our geometry shows clear destructive interference manifested by the dips and peaks in conductance curves in Figure 5. However, the resonances, both reflection and transmission, evident at V_0=0.0, are weakened for non-zero V_0's. So, by introducing more overlaps between the circulating orbits and the edge states, one effect of the potential stripe is to lessen even eliminate the formation of MQBS's, hence broaden even lift the resonances.

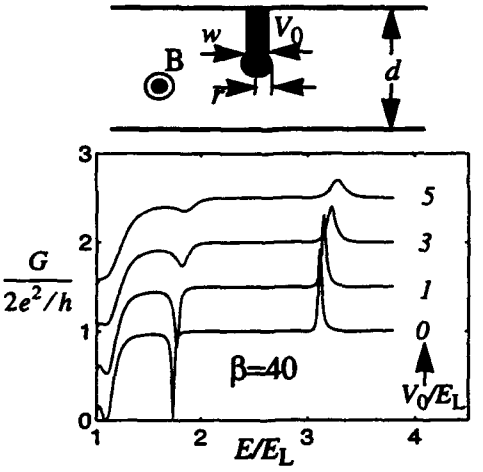


Figure 5. Conductance of a quantum channel with one antidot and a strip of potential barrier V_0. for $V_0/E_L=0, 1, 3$ and 5 at $\beta=40$.

IV. SUMMARY

We have calculated the magneto transport properties of a quantum channel with a single antidot. Resonant reflection and resonant transmission are observed at high magnetic fields. They are induced by the tunneling processes of the edge states, via the magnetic quasi-bound states formed around the antidot. The resonances are broadened in lower magnetic fields because the MQBS's are less localized. The smoothing of resonances is also observed when an additional potential barrier is introduced in the channel which also tend to hinder the formation of MQBS's.

Acknowledgment: This work was supported by the Air Force Office of Scientific Research.

REFERERENCES

[1] M. Büttiker, Physical Review B **38**, 9375 (1988).
[2] J.K. Jain and S.A. Kivelson, Physical Review Letters **60**, 1542 (1988).
[3] Manhua Leng and Craig S. Lent, to appear in Journal of Applied Physics. A similar quantum mechanical calculation was carried out by Takagaki and Ferry. However, in their calculation, the magnetic field is considered to be zero in the leads. See their work in Physical Review B **48**, 8152 (1993).
[4] W. Porod, Z. Shao, C. S. Lent, Physical Review B **48**, 9485 (1993).
[5] J.E. Müller, Physical Review Letters **72**, 2616 (1994).

DYNAMIC BEHAVIOR OF COUPLED QUANTUM DOT CELLS*

P. Douglas Tougaw and Craig S. Lent
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

We examine the dynamic behavior of a large group of coupled quantum dots responding to a changing electrostatic environment. The electrons occupying the quantum dots interact Coulombically and tunnel between neighboring dots. To model this system, we solve the time-dependent Schrödinger equation over finite and semi-infinite domains. These ideas are used to model the dynamic behavior of binary wires, the most fundamental elements of quantum cellular automata. The results of these simulations highlight the importance of kink propagation at polarization domain interfaces.

I. INTRODUCTION

The use of quantum mechanics to design and model computational elements has given rise to several new paradigms for computation. Among these is computing with the ground state, in which the time-independent behavior of a system can be used to perform useful logical functions. When an input is applied to such a system, it changes the boundary conditions of the quantum state so that the system is no longer in the ground state. Unavoidable dissipation then drives the system into the ground state corresponding to the new input. Mapping inputs to outputs enables one to perform useful computation using the dissipation inherent in the array. Since the devices use the ground state to perform calculations, a great deal of design work can be done without regard to the dynamic behavior of the devices. It is possible that some devices will be unable to reach the ground state due to the presence of metastable states, but such states should be quite rare. Information about response time and the possible existence of metastable states requires time-dependent modeling of the system. Such dynamic modeling is the topic of this research.

One example of a system that uses computing with the ground state, quantum cellular automata, is explained in section II. This is followed in section III by a description of the energy-absorbing boundary conditions recently introduced by Hellums and Frensley. Section IV presents results showing the dynamic behavior of two different binary wires. Finally, section V presents conclusions and areas for further work.

II. QUANTUM CELLULAR AUTOMATA

One new scheme that takes advantage of computing with the ground state is called quantum cellular automata, or QCA [1-3]. As shown in figure 1, the cells that compose a QCA consist of four quantum dots where tunneling is allowed between neighbors and with two electrons shared among the four dots. The model Hamiltonian for this system is:

$$H = \sum_{i,\sigma} E_{0,i}\ n_{i,\sigma} - \sum_{i>j,\sigma} t_{i,j}\left(a_{i,\sigma}^{\dagger} a_{j,\sigma} + a_{j,\sigma}^{\dagger} a_{i,\sigma}\right) + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{j,\sigma'}}{\left|\vec{R}_i - \vec{R}_j\right|} + \sum_{k\neq m,j} V_Q \frac{(n_{k,j} - \tilde{\rho})}{\left|\vec{R}_{k,j} - \vec{R}_{m,i}\right|} \quad (1)$$

This Hamiltonian contains second-quantized terms including on-site energies, tunneling between neighbors, and intracellular and intercellular Coulombic repulsion. Due to this repulsion between the two electrons, the charge density of the ground state of the cell is almost completely aligned in one of the two polarization states shown in figure 2. We can use this bistable saturation behavior to encode one bit of binary information in the quantum state of the cell.

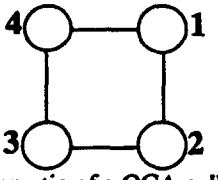


Figure 1. Schematic of a QCA cell. Two electrons are shared among the four sites, and tunneling is allowed between neighbors.

Figure 2. Antipodal alignment of electrons due to Coulombic repulsion.The ground state is highly bistable.

Figure 3 shows an arrangement of cells referred to as a binary wire [4]. In this arrangement, the polarization of each cell causes its neighbors to align in a similar direction, and the polarization information contained in the cell at one end of the wire is transmitted to the other end. The binary wire will be used to transmit polarization information from point to point in our scheme.

Figure 4 shows a group of cells which act as a majority logic gate. The cells on the top, bottom, and left sides have fixed polarizations, while the center cell and the output cell on the right are free to react to the polarizations of the other cells. When such a system is simulated, we find that the polarization of the free cells always aligns in the direction of a majority of the driving neighbors. Therefore, such an arrangement of cells performs majority logic on the three inputs.

If one of the three inputs is defined to be a program line, the device can be thought of as a programmable AND-OR gate. For example, if one of the inputs is held in the logical one state, the output of the majority device will also be one unless both of the other inputs are zero. The device therefore performs the OR operation on the two non-program inputs. Likewise, a zero on the program line causes the device to perform AND logic on the two non-program inputs.

Similar arrangements of QCA cells have been designed to perform inversion of a signal, coplanar crossing of wires, and dedicated AND and OR gates. Combinations of these devices have been used to synthesize more complex devices including exclusive-OR gates and full adders [5].

Figure 3. A binary wire. Dot radius on each site is proportional to the actual calculated charge density on that site. Data is transmitted from one end of the wire to the other.

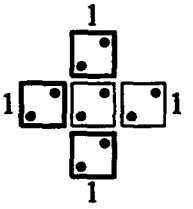


Figure 4. Majority logic gate. The state of the two free cells matches that of a majority of the driving cells.

III. ENERGY-ABSORBING BOUNDARY CONDITIONS

We model dissipation of kink energy in our system by the presence of a semi-infinite line of similar cells which can absorb kink energy. In this case, the time-dependent Schrödinger equation for the entire system (QCA device and reservoir) can be written as [6]:

$$ih\frac{\partial}{\partial t}\begin{bmatrix}\Psi_s\\ \Psi_r\end{bmatrix} = \begin{bmatrix}H_s & H_i^{\dagger}\\ H_i & H_r\end{bmatrix}\begin{bmatrix}\Psi_s\\ \Psi_r\end{bmatrix} \qquad (2)$$

where H_s is the Hamiltonian of the system given in equation (1), H_r is the Hamiltonian of the reservoir, and H_i is the interaction Hamiltonian between the two. As demonstrated recently by Hellums and Frensley, the

effect of H_r is included by using the Green function of a semi-infinite line, which is well known. H_i introduces a time-dependent convolution integral over the past history of the system, which makes the equation non-separable and causes the boundary condition to be non-Markovian.

These non-Markovian boundary conditions cause the dynamics of the QCA device to be irreversible, even though the dynamics of the combined system are reversible. Proper impedance matching of the system-reservoir interaction will prevent kink reflections at the system-reservoir interface, while poor impedance matching can increase the relaxation time for kinks to leave the system.

IV. KINK PROPAGATION IN THE BINARY WIRE

When these ideas are applied to the binary wire shown in figure 3, we can demonstrate the dynamic response of a kink traveling from end of the wire to the other. Figure 5 shows the dynamic response of a wire with relatively low tunneling barriers, while figure 6 shows the response of a wire with higher barriers. At t=0, the polarization of the cell at the left end of the wire is switched and held constant. This introduces a kink in the polarization profile of the wire and places the system in an excited state. The kink propagates away from the fixed end and eventually leaves the wire through the open boundary condition. This dynamic analysis shows that the response time of the faster wire is approximately 30 ps, that of the slower wire is approximately 150 ps, and there are no metastable states to prevent the systems from returning to the ground state. Higher tunneling barriers cause slower relaxation but more highly polarized cells, while lower barriers give a faster relaxation time with more weakly polarized cells.

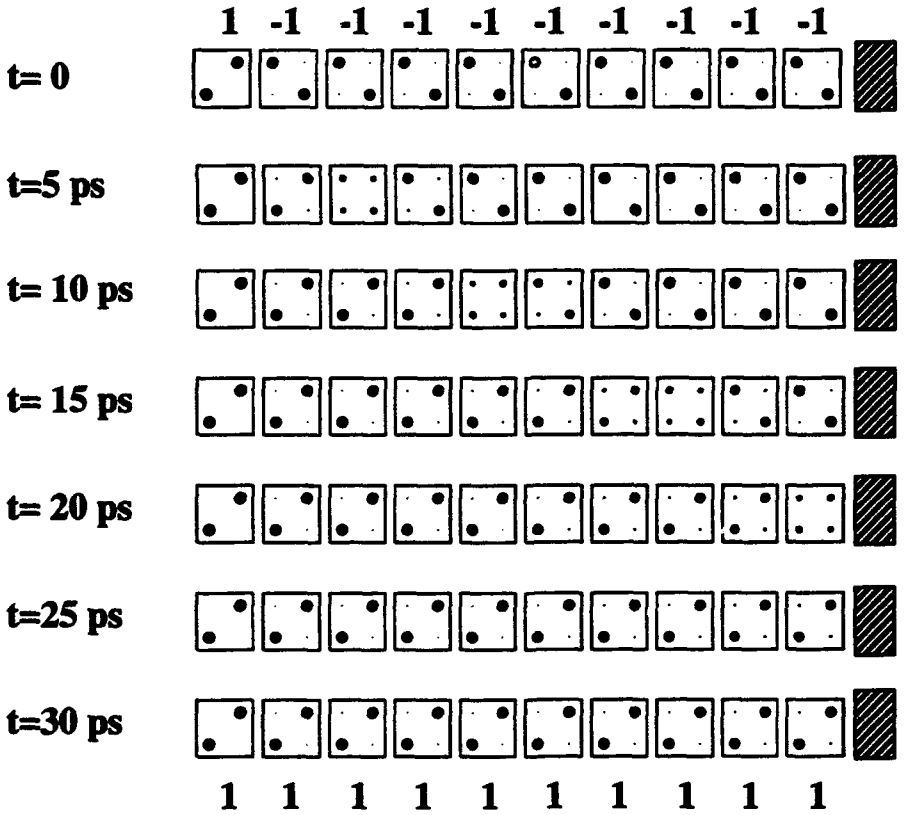


Figure 5. Kink propagation in a binary wire with a semi-infinite reservoir at the right end. These charge density plots show the state of the wire at different time steps. At t=0, a kink is introduced by flipping the left-most cell, and this kink has propagated out of the system into the semi-infinite reservoir by t=30 ps. Here, the tunneling coefficient is 3 meV, indicating relatively low tunneling barriers.

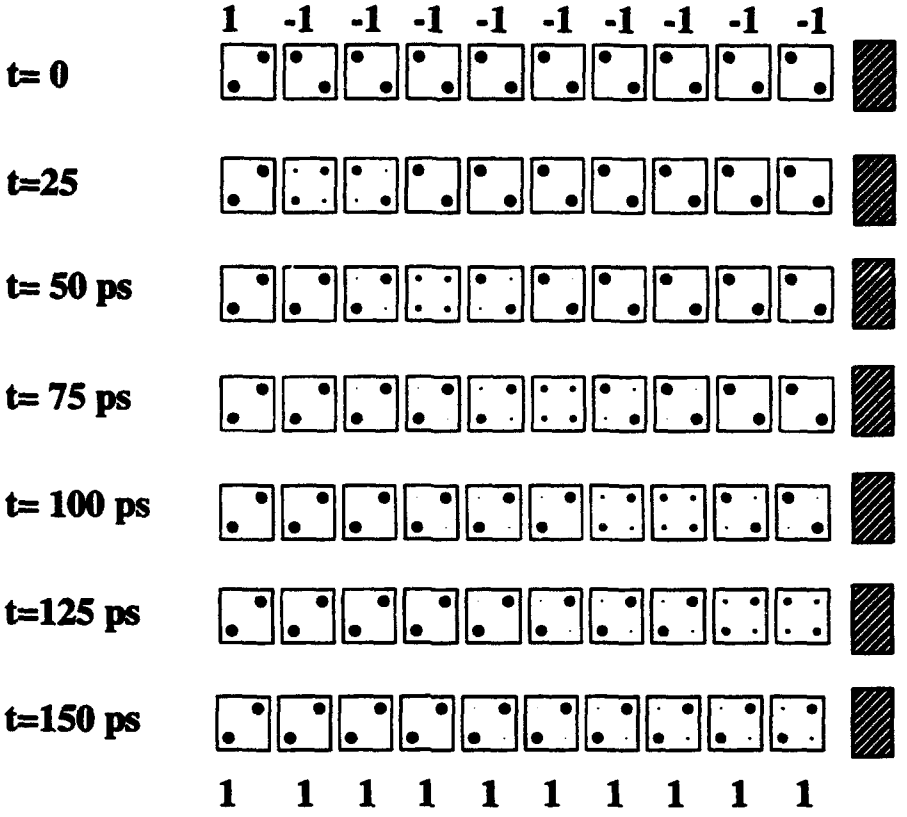


Figure 6. Kink propagation in a binary wire. The tunneling coefficient of this wire is 1.1 meV, indicating relatively high tunneling barriers. The response time of the wire is 150 ps, but the slower response time is accompanied by more highly polarized cells.

V. CONCLUSIONS

The introduction of open boundary conditions to the time-dependent model of QCA devices has allowed us to model the relaxation of the devices to their ground state. This dynamic analysis has provided useful information about relaxation times, as well as demonstrating the absence of metastable states. The importance of kink propagation in the relaxation of the binary wire was demonstrated.

Future directions for this research include more realistic models of dissipation throughout the system, dynamic modeling of more complex devices, and investigation of the impact of finite temperatures on device dynamics.

*This work was supported in part by ONR, AFOSR, and ARPA. This material is based in part upon work supported under a National Science Foundation Graduate Fellowship.

1. C. S. Lent, P. D. Tougaw, and W. Porod, *Appl. Phys. Lett.* **62**, 714 (1993).
2. C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, *Nanotechnology* **4**, 49 (1993).
3. P. D. Tougaw, C. S. Lent, and W. Porod, *J. Appl. Phys.* **74**, 3558 (1993).
4. C. S. Lent and P. D. Tougaw, *J. Appl. Phys.* **74**, 6227 (1993).
5. P. D. Tougaw, C. S. Lent, and W. Porod, *J. Appl. Phys.* **75**,1818 (1994).
6. J. R. Hellums and W. R. Frensley, Phys. Rev. B **49**, 2904 (1994).

MONTE CARLO STUDY OF IONIZED IMPURITY SCATTERING IN QUANTUM WIRES

Lucio Rota
Clarendon Laboratory, Department of Physics
University of Oxford, Parks Road, Oxford OX1 3PU, UK

Stephen M. Goodnick
Department of Electric and Computer Engineering
Oregon State University, Corvallis, Oregon, 97331

Abstract

We have developed a multi-subband Monte Carlo simulation of nonequilibrium transport in quasi-one-dimensional AlGaAs/GaAS quantum wire systems. The simulation includes scattering due to confined polar optical phonons, carrier-carrier scattering, and scattering due to elastic mechanisms such as boundary roughness and impurity scattering. In the present work, we present a detailed Monte Carlo investigation of the effect of ionized impurity scattering in quantum wires under an applied electric field. Static screening in the long wavelength limit is assumed using the multisubband RPA dielectric function for the quasi 1D system. We studied separately the effect of uniformly distributed impurities in the wire and of remote impurities. The former strongly affects transport, particularly at low electric fields and at low temperature. The effect of remote doping is much weaker and practically negligible at room temperature.

I. INTRODUCTION

Recent advances in nanostructure semiconductor technology have allowed the fabrication of wire-like structures where quasi one-dimensional confinement is achieved in a semiconductor material surrounded by another semiconductor with larger band gap [1]. In particular, such wire-like structures have been fabricated with rectangular cross sections [2]. In general, carrier dynamics in wires may be expected to differ from the bulk case: the carrier-phonon scattering rate is affected by changes in the electronic [3] and vibrational [4] properties induced by reduced dimensionality. In the last few years a lot of attention has been dedicated to the study of the electron-phonon interaction in these systems [3-7]. This is certainly the more important interaction and almost the only one present if we are interested in the optical properties of wires. On the other hand, if we hope to be able to take advantage of the reduced 1D density of states to produce electronic devices with higher mobility, the electron-impurity interaction must be seriously considered. The aim of this work is to accurately study the effect of this interaction in quantum wires through a Monte Carlo simulation [8].

II. THEORY

We start by considering the bare unscreened interaction with an impurity located in the point (X, Y, Z), the matrix element is given by:

$$H(q_z) = \langle \Psi_i(x,y,z) \mid H \mid \Psi_j(x,y,z) \rangle = \frac{e^2}{4\pi\epsilon} \int dx \int dy \phi_i(y,z)\phi_j^*(y,z) \cdot$$

$$\frac{1}{L} \int dx \frac{e^{-i(k_x - k_x')x}}{\sqrt{(x-X)^2 + (y-X)^2 + (z-Z)^2}} =$$

$$\frac{e^2 e^{-iq_x X}}{2\pi\epsilon L}\int dz \int dy \phi_i(y,z)\phi_j^*(y,z) K_0\left(\mid q_x \mid \sqrt{(y-Y)^2+(z-Z)^2}\right) , \qquad (1)$$

where $q_x = k_x - k_x'$. The total scattering rate for a carrier in a given subband i and with a given wave vector k_x is the sum over all available final states:

$$\Gamma(i,k_x) = \frac{e^4}{4\pi^2\epsilon^2\hbar^3 L}\sum_j \left[\frac{m_j}{\mid k_x' \mid}\left(\mid H(k_x - k_x')\mid^2 + \mid H(k_x + k_x')\mid^2\right)\right] . \qquad (2)$$

This is the probability for an electron to interact with one impurity. If n_I is the linear density of impurities in the wire, we generate in our simulator their position randomly and then we sum over all the impurities to compute the total scattering rate. In the present work, we used a randomly uniform distribution in the three spatial directions, although the model can handle any profile distribution in the same way. The sum over j in Eq. 2 represent the sum over all possible final subbands. For each final subband there are only two contributions, forward $(k_x - k_x')$ and backward $(k_x + k_x')$ scattering. When the initial and the final subbands are the same, only the backward scattering remains. Looking at Eq. 2 we can also draw some initial conclusions: As in all many-subband systems this scattering is mainly effective through the intrasubband scattering (due to the larger matrix element). In this case there is only one relevant final state available, i.e. $k_x' = -k_x$. If the scattering rate is sufficiently high, the effect of this interaction will be to balance the number of carriers in k_x and $-k_x$ and doing so will drastically reduce the drift velocity when an external electric field is applied. On the other hand, the coulomb interaction is strongly dependent on q_x and we can expect it to decrease strongly for large values of q_x, i.e. for large values of k_x. When an intersubband scattering takes place, we have two different final states and the scattering probability will favor the one which involves the smaller change in momentum. So, when an electric field is applied and we have a population with a given (say positive) average momentum, the preferred scattering will be the forward one and the reduction in the average velocity of the system will not be as dramatic as for the intrasubband scattering. Nevertheless, due to the divergent density of states at the bottom of each subband, most of the scattering will take place from an high energy states in one subband to a low energy states (near to the bottom of the final subband) in higher energy subband. This will result in an almost zero final velocity for the carrier.

Impurities, both inside the wire and in the confining barriers, are screened by the free carriers, as the interaction with these impurities is elastic we can introduce a screening model based on the static limit approximation. In this case the total scattering rate is simply obtained substituting the bare matrix element with the screened one. As a first approximation, this can be done with the substitution:

$$|H(q_x)| \rightarrow \frac{|H(q_x)|}{|\varepsilon(q_x)|} ,$$

where $\varepsilon(q_x)$ is the dielectric matrix obtained in the multisubband RPA approximation.

III. RESULTS AND CONCLUSION

We have investigated the effect of this interaction by varying the electric field, the lattice temperature and the doping concentration from 10^5 to 10^6 cm^{-1}. We also consider both remote impurities (in the AlGaAs layer) and uniform bulk doping. Two interesting effects are observed: At low temperature, and in particular at small electric field, most of the carriers occupy the bottom states in the band where the scattering rate is very high. Both the transient overshoot and the stationary value of the drift velocity are shown in Fig. 1 (a) and (b) for a 300 Å wide quantum wire and a doping concentration of 10^6 cm^{-1}. Here 12 subbands are included in the simulation, although only the first four are significantly populated for the fields considered. We can see that the drift veloci-

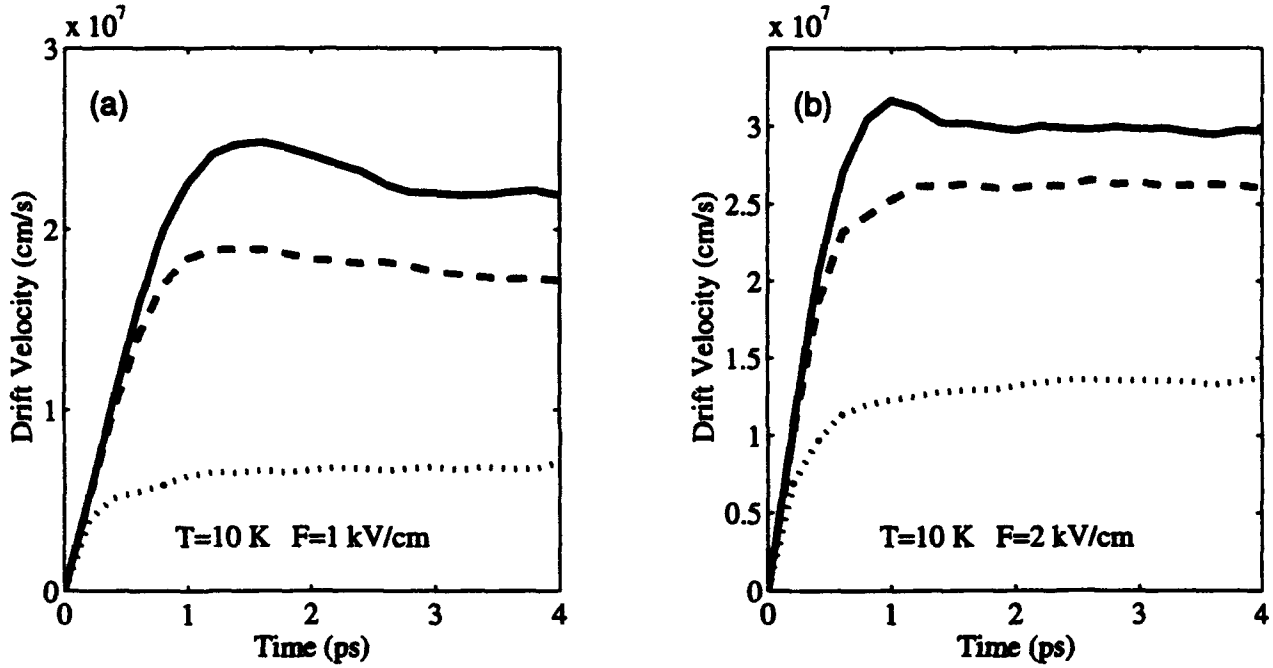


Fig. 1. Drift velocity for a wire without impurities (solid line), with remote (dashed line) and bulk (dotted line) impurities (10^6 cm^{-1}).

ty is strongly reduced by impurities in the wire (dotted line) and remote impurities (dashed line) play an important role as well (the solid line represent an ideal system with no impurities).

In Fig. 2 we plot the room temperature transient velocity under the same conditions. In this case, the remote impurities do not affect the drift velocity at all and even the effect played by the bulk impurities is strongly reduced. This is mainly due to the fact that at room temperature, most of the carriers are in higher energy states where the scattering rate is significantly smaller.

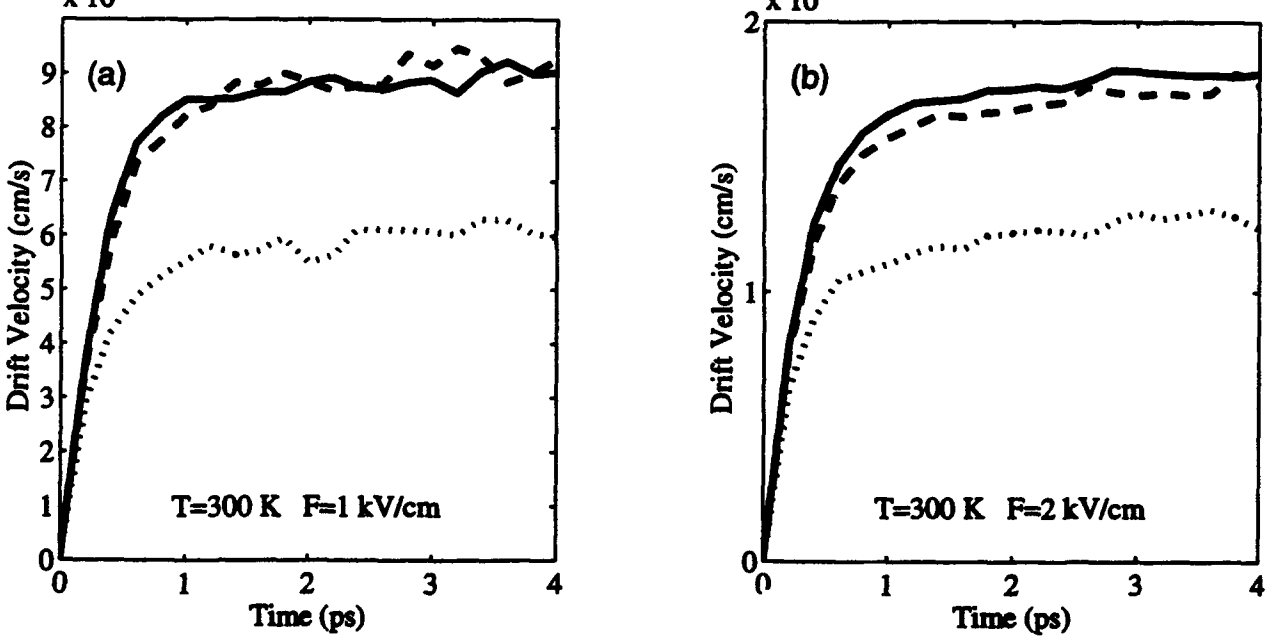


Fig. 2. Drift velocity for a wire without impurities (solid line), with remote (dashed line) and bulk (dotted line) impurities (10^6 cm^{-1}).

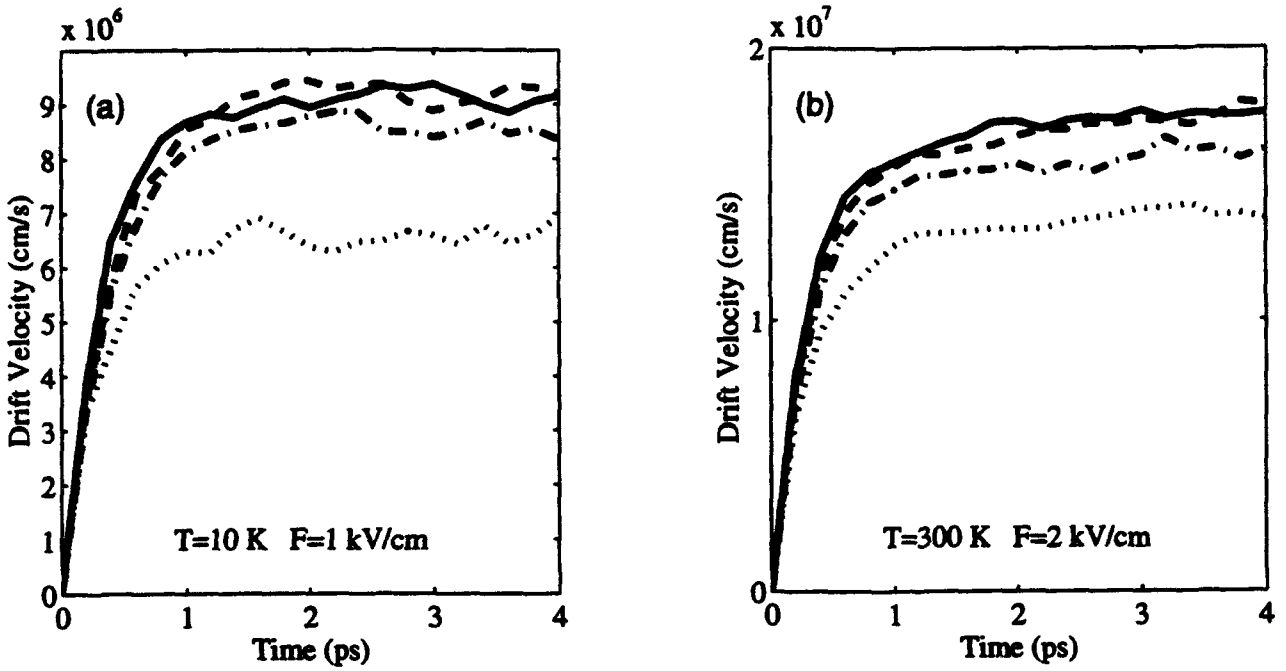


Fig. 3. Drift velocity for a wire without impurities (solid line), with remote (dashed line) and bulk (dotted line) impurities (5×10^5 cm^{-1}). The dash-dotted line represents a system with a bulk doping of 10^5 cm^{-1}.

Fig. 3 shows the same results but with a doping concentration of 5×10^5 cm^{-1}. As in the previous case, the effect of remote impurity is totally negligible and, as expected, the reduction of the drift velocity caused by bulk impurity is smaller. In the same figure the dash-dot line represents a system with a bulk doping of 10^5 cm^{-1}. Even with this low doping the drift velocity is slightly reduced.

In conclusion, we have presented a Monte Carlo investigation of impurity scattering in multi-subband quantum wires. Our results show that remote impurity do not affect significantly the transport properties of the system while direct doping in the wire can reduce the drift velocity significantly. In both cases the effect is weaker at room temperature than at low ones.

IIII. REFERENCES

[1] See e.g. *Nanostructure Physics and Fabrication*, edited by M. A. Reed and P. Kirk, Academic Press, Boston (1989).

[2] T. Demel, D. Heitmann, P. Granbow, and K. Ploog, Phys. Rev. B **38**, 12732 (1988).

[3] see e.g. U. Bockelmann and G. Bastard, Phys. Rev. B **42**, 8947 (1990), and references therein.

[4] see e.g. P. A. Knipp and T. L. Reinecke, Phys. Rev. B **48** 5700 (1993), and references therein.

[5] S. Briggs and J. L. Leburton, Phys. Rev. B **38**, 8163 (1988).

[6] K. W. Kim, M. A. Stroscio, A. Bhatt, R. Mickevicius, and V. V. Mitin, J. Appl. Phys. **70**, 319 (1991).

[7] L. Rota, F. Rossi, S. M. Goodnick, P. Lugli, E. Molinari, and W. Porod, Phys. Rev. B **47**, 1632 (1993).

[8] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer-Verlag, Berlin, 1989), and references therein.

INTRINSIC HIGH FREQUENCY CHARACTERISTICS OF TUNNELING HETEROSTRUCTURE DEVICES

Chenjing L. Fernando, William R. Frensley
Program in Electrical Engineering
Erik Jonsson School of Engineering and Computer Science
The University of Texas at Dallas
Richardson, TX 75083

Abstract

We have developed a general numerical method to solve the periodic time-dependent Schrödinger equation where Quantum Transmitting Boundary Method (QTBM) is used to formulate the boundary conditions of the far-from-equilibrium open systems. The approach is applied to the resonant tunneling diode (RTD) with a superposition of a dc and sufficiently small ac bias. Results of the linear admittance, rectification coefficient and second harmonic generation coefficient are presented as a function of frequency and bias. The calculation has shown that at high frequency (several THz), the intrinsic linear response of RTD becomes capacitive in the NDR region and the rectification coefficient and second harmonic generation coefficient show a resonant enhancement. It indicates that the intrinsic high frequency limit (f_{max}) is influenced more by the electron exchange between the reservoir and the resonant state in the well than by the resonant width. Our results are consistent with those obtained by Wigner function, but in disagreement with most of the results obtained by Schrödinger equation and Green's function. This contradiction is solely due to the problems of definition of reactive current component in the literature.

I INTRODUCTION

High speed device and circuit applications generate considerable interest for the study of the tunneling heterostructure devices. Since the first detection of resonant tunneling diode (RTD) at 2.5 GHz by Sollner and co-workers [1] both experimental and theoretical research work have been widely carried out [2-6].

The existing theoretical results unfortunately conflict with one another. For example, those obtained by conventional tunneling theory based on Schrödinger equation predict inductive behavior at high frequency while Wigner function gives capacitive results. To settle the contradiction, we have developed a systematical numerical method based on single-particle Schrödinger equation with boundary conditions set up by Quantum Transmitting Boundary Method (QTBM)[7]. This method performs task under any bias condition. The application to RTD shows a consistent characteristics with that of Wigner function. The electron exchange between the reservoir and the resonant state in the well plays more important role in the high frequency response than we expected. Two definitions of the current reactive component are to be discussed and compared, which will unify the conflicting results by different approaches.

II THEORETICAL MODEL

In this section, an one-dimensional numerical model of periodic time-dependent Schrödinger equation is to be presented. We consider an open system with two boundary regions (reservoirs): left

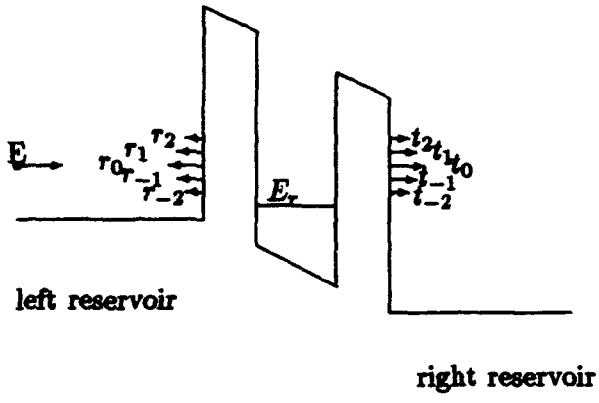


Figure 1: The numerical model of periodic time-dependent Schrödinger equation.

and right. The left voltage is $V_{left} = V_L + \tilde{v}_L \cos\omega t$, the right $V_{right} = V_R + \tilde{v}_R \cos\omega t$, and inside the system $v(x,t) = v_{dc}(x) + \tilde{v}(x)\cos\omega t$. All the incoherent processes are ignored and the flat band potential distribution is assumed, as illustrated in Fig. 1.

Within the open system the wavefunction has the form:

$$\psi(x,t) = \sum_{m=-\infty}^{\infty} \psi_m(x) e^{-im\omega t - i\omega_0 t}. \tag{2.1}$$

Inserting (2.1) into the time dependent Schrödinger equation $i\hbar\partial\psi/\partial t = H\psi$ and collecting terms of equal frequency leads to these equations, in discretized form [8] :

$$\begin{aligned}
-s_j\psi_{0,j-1} + (d_j - E)\psi_{0,j} - s_{j+1}\psi_{0,j+1} &= 0, \\
-s_j\psi_{\mp1,j-1} + (d_j - E \pm \hbar\omega)\psi_{\mp1,j} - s_{j+1}\psi_{\mp1,j+1} - \left(\frac{\tilde{v}_j}{2}\right)\psi_{0,j} &= 0, \\
&\vdots \\
-s_j\psi_{\mp m,j-1} + (d_j - E \pm m\hbar\omega)\psi_{\mp m,j} - s_{j+1}\psi_{\mp m,j+1} - \left(\frac{\tilde{v}_j}{2}\right)\psi_{\mp(m-1),j} &= 0,
\end{aligned} \tag{2.2}$$

where d_j and s_j are the diagonal and off-diagonal elements of the Hamiltonian matrix defined in [8], respectively.

Since we are now dealing with the open system, the appropriate boundary conditions should be applied to (2.2). Fig. 1 shows the physical picture of a single electron with energy E incident from the left reservoir. At the two boundary regions, the wavefunction can be written as [9] :

$$\psi_B = \sum_{m=-\infty}^{\infty} \left(t_{m,B} e^{ik_{mB}x} + r_{m,B} e^{-ik_{mB}x}\right) e^{-i(\omega_o + m\omega)t} \sum_{l=-\infty}^{\infty} J_l\left(\frac{\tilde{v}_B}{\hbar\omega}\right) e^{-il\omega t}. \tag{2.3}$$

where $B = L$ (left reservoir) or $B = R$ (right reservoir) and $J_l(x)$ is the Bessel function.

We formulate the boundary conditions by implementing Quantum Transmitting Boundary Method (QTBM) to the boundary wavefunctions (2.3). Incorporate these boundary conditions with (2.2), a set of linear equations are readily formed, which compose a block tridiagonal matrix. This final matrix is the system to be solved.

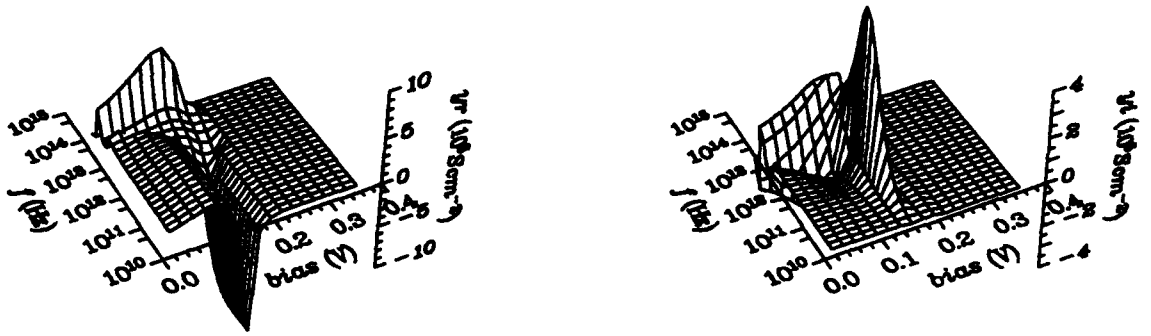


Figure 2: Linear response of a RTD structure.

III AC SMALL SIGNAL RESPONSE

The total current can be represented as:

$$I = -q\hbar/m^* \sum_k P_k Im < \psi^* \frac{\partial \psi}{\partial x} > . \tag{3.1}$$

where P_k is the probability for the wave vector k.

The current components are defined as ($v = \tilde{v}_R - \tilde{v}_L$):

$$I = I_0 + \frac{1}{2}(yve^{i\omega t} + y^*ve^{-i\omega t}) + \frac{1}{4}a_{rect}v^2 + \frac{1}{8}(a_{2\omega}v^2e^{2i\omega t} + a^*_{2\omega}v^2e^{-2i\omega t}) \tag{3.2}$$

or one can rewrite the above definition in a sinusoidal form:

$$\begin{aligned} I &= I_0 + Re(y)v\cos(\omega t) - Im(y)v\sin(\omega t) \\ &\quad + \frac{1}{4}a_{rect}v^2 + \frac{1}{4}Re(a_{2\omega})v^2\cos(2\omega t) - \frac{1}{4}Im(a_{2\omega})v^2\sin(2\omega t) \end{aligned} \tag{3.3}$$

Because these authors neglected *the minus sign in the definition of the sinusiodal form*, the inductive results [3,4] claimed by them are essentially capacitive which is in agreement with our calculation as well as with that of Wigner function.

IV RESULTS

We apply our method to a GaAs/AlGaAs RTD structure with barrier width 28.25 A and well width 45.2 A. The results of the linear response ($yr = Re(y)$, $yi = Im(y)$) and the nonlinear reponse of the second order($a_{2\omega}$, a_{rect}) are demonstrated in Fig. 2 and Fig. 3. Our calculations confirm that the linear response of RTD is capacitive at high frequencies and the nonlinear responses ($a_{2\omega}$ and a_{rect}) do show enhancement peaks. According to our calculation the f_{max} derived from the linear responses and the peak positions of the nonlinear responses are less influenced by the width of the resonant state E_r. They are more closely related with the energy difference between the resonant state E_r and the reservoir.

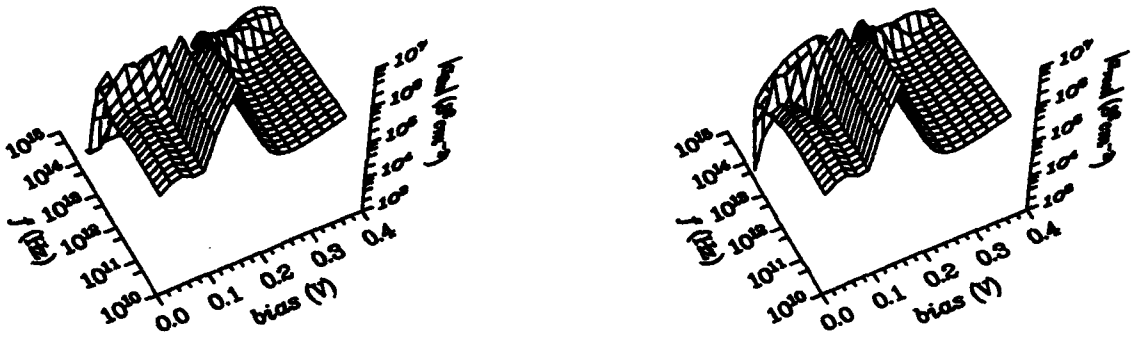


Figure 3: Non-linear response of a RTD structure.

V CONCLUSIONS

We have demonstrated for the first time the three dimensional plots of the linear admittance and the nonlinear responses of the second order for a RTD structure as a function of bias and frequency. The results obtained from the periodic-time dependent Schrödinger equation and Wigner function are characteristically agreeable. The numerical method presented here is able to be applied to any tunneling heterostructure with arbitrary potential distribution within the device.

References

[1] Sollner, T. C. L. G., W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, 1983, Appl. Phys. Lett. **43**, 588.

[2] Frensley, W. R., 1990, Rev. Mod. Phys.**62**, 745.

[3] Mains, R. K., and G. I. Haddad, 1988, J. Appl. Phys. **64**, 5041.

[4] Liu, H. C., 1991, Phys. Rev. B **43**, 12538.

[5] Kislov, V. and Kamenev, A., 1991, Appl. Phys. Lett. **59**, 1500.

[6] Wingreen, N. S., 1990, App. Phys. Lett. **56**, 253

[7] Lent, C. S. and Kirkner, D. J., 1990, J. Appl. Phys. **67**, 6353.

[8] Frensley, W. R. in Chapter 9, *Heterostructures and Quantum Devices* (W. R. frensley and N. G. Einspruch, eds), Academic Press, San Diego, 1994.

[9] Coon, D. D. and Liu, H. C., 1985, J. Appl. Phys. **58**, 2230.

QUANTUM DEVICE MODELING WITH NON-EQUILIBRIUM GREEN FUNCTIONS

R. Lake
Corporate R&D, Texas Instruments, Inc., Dallas, TX 75265

Abstract

I describe the non-equilibrium Green function approach to modeling wide-cross-sectional area quantum devices such as resonant tunneling diodes. Several approximations for treating scattering are discussed: Born, self-consistent Born, and a single-electron multiple sequential scattering algorithm. A generalized treatment of open-system boundaries is presented.

I HAMILTONIAN AND BASIS

The general form of the Hamiltonian is $H = H_o + \underbrace{H_{pop} + H_{ac} + H_{IR}}_{\Sigma}$ where H_o contains the kinetic energy and the effects of the band structure, the applied potential, and the Hartree potential. The three terms to the right represent the potential felt by the electrons due to polar optical phonons, acoustic phonons, and interface roughness, respectively. The underbrace and Σ indicate that these terms are included through self-energies. For numerical reasons, H_o is broken down into three terms, $H_o = H_o^D + \underbrace{H_o^L + H_o^R}_{\Sigma_B}$, which represent the Hamiltonian of the device, the left contact, and the right contact, respectively. The underbrace indicates that the effect of the contacts on the device is also taken into account through a self-energy, Σ_B.

The Hamiltonian matrix will explicitly be written in terms of the basis $\langle \mathbf{r}|\mathbf{k},n\rangle = e^{i\mathbf{k}\cdot\mathbf{r}_t}\phi_i(z)/\sqrt{A}$ where $\mathbf{k}$ is the transverse wavevector and $\phi_i(z)$ is a localized (Wannier) function localized around site 'i'. Keeping only nearest neighbor matrix elements, the matrix elements of H_o are $\langle \mathbf{k},i|H_o|\mathbf{k},j\rangle = (\epsilon_{\mathbf{k}i} + \epsilon_i)\delta_{i,j} - t_{i,j}\delta_{i,j\pm1}$. The site energies and hopping elements can be related to the discretized effective mass Hamiltonian, $H_o = \frac{-\hbar^2}{2}\frac{d}{dz}\frac{1}{m^*(z)}\frac{d}{dz} + V(z)$ in the usual way.

II EQUATION OF MOTION FOR G^R and $G^<$

The non-equilibrium Green function (NEGF) formalism provides a method for calculating the non-equilibrium statistical ensemble average of the single particle correlation operator, ($\hbar = 1$), $G^<_{i,j}(\mathbf{k};t,t') = i\langle c^\dagger_{j,\mathbf{k}}(t')c_{i,\mathbf{k}}(t)\rangle$ where $c_{i,\mathbf{k}}$ is the electron annihilation operator for an electron in localized state 'i' with transverse momentum $\mathbf{k}$. Once the correlation function $G^<$ is known, it immediately provides the electron density, $n_i = \frac{-i}{Aa}\sum_{\mathbf{k}}\int\frac{dE}{2\pi}G^<_{i,i}(\mathbf{k},E)$ and the current density, $J_{i+1/2} = -\frac{1}{A}\sum_{\mathbf{k}}\int\frac{dE}{2\pi}\left[t_{i,i+1}G^<_{i+1,i}(\mathbf{k},E) - t_{i+1,i}G^<_{i,i+1}(\mathbf{k},E)\right]$.

The two equations of motion that must be solved are (in matrix notation)

$$(E - H_o - \Sigma^R - \Sigma^{RB})G^R = 1 \tag{2.1}$$

$$(E - H_o - \Sigma^R - \Sigma^{RB})G^< = (\Sigma^< + \Sigma^{<B})G^{R\dagger} \tag{2.2}$$

In Eqs. (2.1) and (2.2), the effects of the scattering are contained in Σ^R and $\Sigma^<$ and the effect of the open boundaries are contained in Σ^{RB} and $\Sigma^{<B}$

III OPEN BOUNDARY SELF ENERGIES

For a device consisting of sites $1,\ldots,N$, I include the coupling to the leads, $t_{0,1}$ and $t_{N,N+1}$ exactly using Dyson's equation to obtain the boundary self energies: $\Sigma^{RB}_{1,1} = g^R_{0,0}|t_{0,1}|^2$, $\Sigma^{RB}_{N,N} = g^R_{N+1,N+1}|t_{N,N+1}|^2$, $\Gamma^B_{1,1} = a_{0,0}|t_{0,1}|^2$, $\Gamma^B_{N,N} = a_{N,N}|t_{N,N+1}|^2$, $\Sigma^{<B}_{1,1} = if_L\Gamma^B_{1,1}$, and $\Sigma^{<B}_{N,N} = if_R\Gamma^B_{N,N}$; where g^R is the Green function of the unconnected lead, a is the corresponding spectral function, and f is the Fermi-factor. The boundary self-energies Σ^{RB} and Γ^B are valid even if the leads have spatially varying potentials, and $\Sigma^{<B}$ is valid if the lead is in equilibrium.

An explicit representation of G^R for a device of 3 sites is

$$\left[G^R\right] = \begin{bmatrix} E-\epsilon_{k,1}-\epsilon_1-\Sigma^R_{1,1}-\Sigma^{RB}_{1,1} & t_{1,2}-\Sigma^R_{1,2} & -\Sigma^R_{1,3} \\ t_{1,2}-\Sigma^R_{1,2} & E-\epsilon_{k,2}-\epsilon_2-\Sigma^R_{2,2} & t_{2,3}-\Sigma^R_{2,3} \\ -\Sigma^R_{3,1} & t_{3,2}-\Sigma^R_{3,2} & E-\epsilon_{k,3}-\epsilon_3-\Sigma^R_{3,3}-\Sigma^{RB}_{3,3} \end{bmatrix}^{-1} \tag{3.3}$$

$\Sigma^B_{0,0}$ is obtained from $g^R_{0,0}$,

$$g^R_{0,0} = \begin{bmatrix} -t_{-2}e^{-i\gamma_{-2}a} & t_{-2,-1} & 0 \\ t_{-1,-2} & E-\epsilon_{k,-1}-\epsilon_{-1} & t_{-1,0} \\ 0 & t_{0,-1} & E-\epsilon_{k,0}-\epsilon_0 \end{bmatrix}^{-1}_{0,0} \tag{3.4}$$

In Eq. (3.4), I have again taken into account the semi-infinite *uniform potential* region to the left by a self energy which can be calculated analytically, $\Sigma^{RB}_{-2,-2} = g^R_{-3,-3}|t_{-3,-2}|^2 = -t_{-3,-2}e^{i\gamma_L a}$, and used the dispersion relation, $E = \epsilon_{k,L} + \epsilon_L - 2t\cos(\gamma_L a)$, to simplify the (-2,-2) element.

Substituting Dyson's equation, $G^<_{0,j} = g^R_{0,0}(-t_{0,1})G^<_{1,j} + g^<_{0,0}(-t_{0,1})G^A_{1,j}$, into the current expression for $J_{1/2}$, using $G^< = G^R(\Sigma^< + \Sigma^{<B})G^{R\dagger}$ and $A = G^R(\Gamma + \Gamma^B)G^{R\dagger}$ gives

$$\begin{aligned} J_{1/2}(k,E) &= ei\Gamma^B_{1,1} \sum_{n,m\in\{1,\ldots,N\}} G^R_{1,n}\left[f_L\Sigma^>_{n,m} + (1-f_L)\Sigma^<_{n,m}\right]G^{R\dagger}_{m,1} \\ &\quad + e\Gamma^B_{1,1}\Gamma^B_{N,N}|G^R_{1,N}|^2(f_L - f_R) \end{aligned} \tag{3.5}$$

The terms in (3.5) are functions of k and E. The first term is the contribution to the current from scattering and the second term is the contribution from direct transitions from the left to right contact. With no scattering, the first term is absent, and the second term is the usual tunneling formula using the Fisher-Lee form of the transmission coefficient [1]. The Γ^B's need not be simply factors of velocity but can account for leads with spatially varying potentials. The boundary conditions may provide a unified treatment within a Tsu-Esaki formulation of current from continuum states and emitter quasi-bound states [2].

IV COMBINING BORN AND SELF-CONSISTENT BORN SELF ENERGIES

The non-local self-energy due to polar-optical phonons couples Eqs. (2.1) and (2.2) in energy, E and momentum, k, necessitating the storage of four four-dimensional functions, $G^<_{i,j}(k,E)$, $G^R_{i,j}(k,E)$, $\Sigma^<_{i,j}(k,E)$, and $\Sigma^R_{i,j}(k,E)$. This is not feasible for any modestly sized device. Therefore, I combine a first Born treatment of the polar optical phonons and a self consistent Born treatment of the elastic scattering mechanisms: interface roughness and high temperature acoustic phonons. The new equations of motion are found from the Dyson equation for the path ordered Green function

$$G^p = g^p + g^p\Sigma^p G^p \tag{4.6}$$

where the matrix notation now denotes both summing over sites and integrating over the Keldysh contour. In a Born or self-consistent Born approximation, Σ has three contributions, one from each scattering process, acoustic phonon, polar optical phonon, and interface roughness: $\Sigma^p = \Sigma^p_{ap} + \Sigma^p_{pop} + \Sigma^p_{IR}$. I iterate once keeping only the first order term in Σ^p_{pop} to obtain $G^p = g^p + g^p[\Sigma^p_{ap} + \Sigma^p_{IR}]G^p + g^p\Sigma^p_{pop}g^p$. Since Σ^p_{pop} is being used in a first Born approximation, it must be calculated in the first Born approximation to conserve current. In keeping with the notation of big G's and little g's, little σ's will be used to denote the self energies due to the polar optical phonons. Breaking up G^p along the two branches of the time path [3], gives

$$G^R = g^R + g^R\Sigma^R G^R + g^R\sigma^R g^R \tag{4.7}$$

and

$$G^< = g^< + g^R\Sigma^R G^< + g^R\Sigma^< G^A + g^<\Sigma^A G^A + g^R\sigma^R g^< + g^R\sigma^< g^A + g^<\sigma^A g^A \tag{4.8}$$

where the large Σ is the sum of the self energies that I treat in the self-consistent Born approximation, ie. $\Sigma \equiv \Sigma_{ap} + \Sigma_{IR}$. Operating on (4.7) and (4.8) from the left with $E - \epsilon_k - H_o - \Sigma^{RB}$ and using the equations of motion for g^R and $g^<$, $(E - \epsilon_k - H_o - \Sigma^{RB})g^R = 1$ and $(E - \epsilon_k - H_o - \Sigma^{RB})g^< = \Sigma^{<B}g^A$ gives the final form of the equations that I need to solve.

$$(E - \epsilon_k - H_o - \Sigma^R - \Sigma^{RB})G^R = 1 + \sigma^R g^R \tag{4.9}$$

and

$$(E - \epsilon_k - H_o - \Sigma^R - \Sigma^{RB})G^< = (\Sigma^< + \Sigma^{<B})G^A + \sigma^R g^< + \sigma^< g^A \tag{4.10}$$

Equations (4.9) and (4.10) conserve current.

Treating the polar optical phonons in the first Born approximation leads to an immense simplification in the numerical solution. Since the self-energies due to high temperature acoustic phonons and interface roughness are elastic, equations (4.9) and (4.10) decouple in energy. Also, Σ^R becomes only a function of G^R so that Eq. (4.9) for G^R is a closed loop in an iterative solution of Σ^R and G^R. Furthermore, $\Sigma^R_{i,j}$ is diagonal $\propto \delta_{i,j}$. Therefore, for a given energy, in the iterative solution of (4.9), I only need to store the diagonal elements $G^R_{i,i}(k)$ and $\Sigma^R_{i,i}(k)$ and invert a tri-diagonal matrix.

V SINGLE-ELECTRON MULTIPLE SEQUENTIAL SCATTERING ALGORITHM

A self-consistent Born approximation (SCBA) requires a converged solution of the Σ's and the G's to conserve current. This can be numerically problematic. The following is an alternative.

The point-of-view which informs the work of Roblin and Liou [4] is used to create a multiple sequential scattering (MSS) algorithm for non-equilibrium Green functions. The point-of-view is as follows. A plane wave, ψ_0, injected from the contact propagates into the device and scatters due to the random potential of phonons and interface roughness. Flux is removed from the incident wave and fed into the scattered wave, ψ_1, which has no phase coherence with ψ_0. ψ_1 now scatters creating ψ_2, etc. The MSS approach is a single electron approach which does not account for the Pauli-exclusion principle, but it provides a means to truncate at any order the infinite expansion leading to the SCBA and still conserve current. In the limit of infinite sequential scattering, MSS is the SCBA for zero electron density.

N sequential scattering events give rise to $N+1$ retarded Green functions, G^R_n, and $N+1$ correlation functions, $G^<_n$, where $n = \{0, ..., N\}$. Here, I consider only elastic processes. For a given

total energy, E, the general form of the retarded self energy is $\Sigma^R(E_z) = \int_{-\infty}^{E} dE_z' D(E_z, E_z') G^R(E_z')$ where E is the total energy, the integral over E_z is the sum over transverse momentum, and $D(E_z, E_z')$ is a known function determined by the type of scattering considered. I have suppressed the position coordinates. The equations defining the propagators G_n^R for the N scattered waves are:

$$\begin{aligned}
\left[E_z^0 - H_o - \Sigma^{RB} - \int_{-\infty}^{E} dE_z^1 D(E_z^0, E_z^1) G_1^R(E_z^1)\right] G_0^R(E_z^0) &= 1 \\
\left[E_z^1 - H_o - \Sigma^{RB} - \int_{-\infty}^{E} dE_z^2 D(E_z^1, E_z^2) G_2^R(E_z^2)\right] G_1^R(E_z^1) &= 1 \\
&\vdots \\
\left[E_z^{N-1} - H_o - \Sigma^{RB} - \int_{-\infty}^{E} dE_z^N D(E_z^{N-1}, E_z^N) G_N^R(E_z^N)\right] G_{N-1}^R(E_z^{N-1}) &= 1 \\
\left[E_z^N - H_o - \Sigma^{RB}\right] G_N^R(E_z) &= 1 \qquad (5.11)
\end{aligned}$$

Since the last wave, by definition, does not scatter, the retarded Green function governing its propagation, G_N^R, is simply the bare Green function, g^R. Starting with the last propagator, $G_N^R = g^R$, and back-substituting into each higher equation, one calculates each G_n^R.

The equations defining $G_n^<$ are

$$\begin{aligned}
(E_z^0 - H_o - \Sigma^{RB} - \Sigma_0^R) G_0^<(E_z^0) &= \Sigma^{<B} G_0^{R\dagger}(E_z^0) \\
(E_z^1 - H_o - \Sigma^{RB} - \Sigma_1^R) G_1^<(E_z^1) &= \int_{-\infty}^{E} dE_z^0 \, D(E_z^1, E_z^0) \, G_0^<(E_z^0) \, G_1^{R\dagger}(E_z^1) \\
&\vdots \\
(E_z^N - H_o - \Sigma^{RB} - \Sigma_N^R) G_N^<(E_z^N) &= \int_{-\infty}^{E} dE_z^{N-1} \, D(E_z^N, E_z^{N-1}) G_{N-1}^<(E_z^{N-1}) G_N^{R\dagger}(E_z^N)
\end{aligned} \qquad (5.12)$$

One solves the set of equations (5.12) by starting at the top, solving for $G_0^<$, substituting that into the second equation, and so on working downward. Notice that the source term, $\Sigma^{<B} G^{R\dagger}$, for $G_0^<$ is due to injection from the contacts while the source terms for the scattered waves ($n = 1, ..., N$) come from the scattering of the previous waves.

The resulting electron density and current density per unit energy are $n_i(E) = -i \sum_{n=0}^{N} \int_{-\infty}^{E} dE_z \, \rho_{2D}(E_z) G_{n_{i,i}}^<(E_z)$ and $J_{i+1/2}(E) = \frac{1}{A} \sum_{n=0}^{N} \int_{-\infty}^{E} dE_z \rho_{2D}(E_z) |t_{i,i+1}|^2 \, 2\mathrm{Re} G_{n_{i,i+1}}^<(E_z)$ where the subscript 'i' is the site index that has been suppressed.

Acknowledgement: I acknowledge a very useful discussion with S. Hershfield.

References

[1] D.S. Fisher and P.A. Lee, *Phys. Rev. B*, **23**, 6851 (1981).

[2] T. Fiig and A.P. Jauho, *Surf. Sci.*, **267**, 392 (1992).

[3] D.C. Langreth in *1975 NATO Advanced Study Institute on Linear and Nonlinear Electron Transport in Solids*, (Plenum Press, New York, 1976), p. 3.

[4] P. Roblin and W. Liou, *Phys. Rev. B*, **40**, 2146 (1993).

AN ACCURATE, EFFICIENT ALGORITHM FOR CALCULATION OF QUANTUM TRANSPORT IN EXTENDED STRUCTURES

T. J. Godin[a] and Roger Haydock[b]
[a]*MSRC, Pacific Northwest Laboratory,* P.O. Box 999, Richland WA 99352*
[b]*Department of Physics and Materials Science Institute, University of Oregon, Eugene OR 97403*

Abstract

In device structures with dimensions comparable to carrier inelastic scattering lengths, the quantum nature of carriers will cause interference effects that cannot be modeled by conventional techniques. The basic equations that govern these "quantum" circuit elements present significant numerical challenges. We describe the block recursion method, an accurate, efficient method for solving the quantum circuit problem. We demonstrate this method by modeling dirty inversion layers.

I. INTRODUCTION

Numerical "pre-testing" of proposed integrated circuit elements can vastly reduce development time and expense [1]. "Conventional," semiclassical device modeling methods fail, however, for structures with dimensions comparable to the distance between carrier inelastic (*e.g.*, electron-phonon) collisions. On this submicron scale, without inelastic collisions to destroy electron wave coherence, the quantum wavelike nature of a carrier will produce interference effects as the carrier scatters from device walls, individual dopants, and other deviations from crystallinity [2]. Modeling such a "quantum circuit element" presents significant numerical challenges.

In this paper we describe the block recursion method, an efficient, accurate algorithm for solving the quantum circuit problem. We have presented thorough derivations and analyses of this method, and some applications, in specialized papers, [3,4,5] and block recursion software is publicly available [5]. Our purpose here is to present the salient points in a manner accessible to a broader audience. Readers interested in more detail should consult the references or authors.

II. THE BASIC EQUATIONS OF QUANTUM CIRCUIT THEORY

Conventional device modeling starts with a basic set of coupled equations (diffusion-drift, continuity, Poisson, etc.) that are discretized (*e.g.*, by finite difference or finite element schemes) and solved. Quantum effects are only considered implicitly (through effective masses, carrier/band energetics, etc.) [1]. At each node, Ohm's Law relates the local current to the local electric field using *conductivity*, an intensive, local property.

In quantum circuits, however, resistance cannot be locally defined. When electron waves scatter and interfere between two barriers (device walls, dopant atoms, etc.) total resistance depends not only on properties of each scatterer, but on electronic DeBroglie wavelengths and distances between scatterers. It is more useful, then, to calculate a device's total *conductance*, an extensive property. Without inelastic scattering, the conductance Σ of a device with carriers of energy E is given by the Landauer formula [6],

$$\Sigma(E) = (2e^2/\hbar)\, T(E)\,. \tag{1}$$

e is the carrier charge; $\hbar$ is Planck's constant divided by 2π. The transmittance T is the probability that an incident electron will be pass through a device. When transport occurs only near the Fermi Energy, we take this as E. T is calculated by solving the time independent Schrödinger Equation,

$$\{(-\hbar^2/2m^*)\nabla^2 + V(x,y,z)\}\,\psi = E\,\psi, \tag{2}$$

for an unknown quantum state ψ. $(-\hbar^2/2m^*)\nabla^2 + V(x,y,z)$ is H, the Hamiltonian operator. V, the electric potential within the device, arises from fields caused by device walls, dopants, gates, etc. m^* is the effective carrier mass. In a crystalline region, ψ is a traveling wave. If a device is connected to crystalline "leads," the transmittance is found by applying boundary conditions of incident and reflected waves in one lead and a transmitted wave in the other. The transmittance is the ratio of transmitted to incident wave amplitudes.

*Operated for the U.S. Dept. of Energy by Battelle Memorial Institute under contract DE-AC06-76RLO 1830.

transmitted wave in the other. The transmittance is the ratio of transmitted to incident wave amplitudes.

V must be specified on an atomic level, since electrons scatter from individual atoms. Moreover, Eq. (2) must be solved over an inelastic scattering volume (≈ 0.01 μ wide at room temperature [7]), even if this region extends beyond a single device (*i.e.*, an "extended structure"). Electrons can scatter and interfere at interfaces with leads and with other devices. Any model must preserve carrier phase relationships between these boundaries. (Larger volumes require a hybrid quantum/conventional model; this is beyond the scope of this paper.)

Eq. (2) may be discretized by a tight-binding method [8]. This is like a simple finite element model, with one node per atom. For example, consider a simple quantum circuit (Fig. 1): an (n-2) atom device connected to one-dimensional leads. First, we convert Eq. (2) to a matrix eigenvalue equation by defining a set of n×1 vectors $\{\phi_m: m= 1,2,...n\}$. The mth element of ϕ_m equals one, and all other elements equal zero. ϕ_m represents an electron localized at node m. In this basis, H is a matrix whose diagonal elements $H_{mm} = V_m$ equal the potential V at each node m. We set the off-diagonal elements of H equal to energy v for nearest neighbor atoms, and zero for all other atom pairs. v can be fit to the device band width. This model is now a discrete network (shown by the solid lines in Fig. 1) through which carriers move. Using this matrix form of H, Eq. (2) becomes a matrix eigenvalue equation whose solution is an n×1 vector Ψ. The n components of vector Ψ equal the function values of state ψ at each of the n nodes. The boundary condition[the values of ψ at nodes 0, 1, n-1 and n (Fig. 1)] must be satisfied; the solution algorithm presented in Sec. III does this automatically.)

This model is appropriate for simple potentials, and carriers confined to a single band. More complicated potentials can be modeled by varying the off-diagonal elements of H, and multiple bands by adding more basis vectors per node.

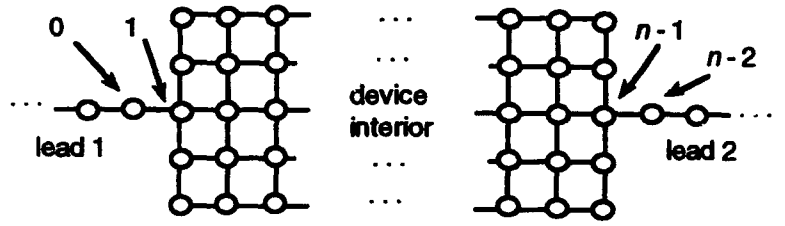


Figure 1. Schematic representation of a model of a simple two-lead quantum device with n atoms. Nodes (circles) are located at atoms. Lines connect nearest neighbors. Nodes (atoms) at the lead-device interfaces are numbered.

III. THE BLOCK RECURSION ALGORITHM

Eq. (2) presents significant numerical challenges. First, solids have on the order of 100 atoms per cubic nanometer. Models of even small device models may require millions of nodes. Moreover, this problem requires calculation of extremely sensitive coupling between leads. Many previous algorithms have worked only for small atomic clusters. For realistically sized devices, unstable matrix inversions or divergent recursions made these earlier algorithms impractical (for a review, see [5]).

We developed the Block Recursion Method for this problem. To illustrate this approach, consider the device in Fig. 1. Suppose we define an n×2 matrix, u_1, whose columns are the vectors representing ϕ_1 and ϕ_{n-2} respectively. Similarly, let the columns of u_0 be ϕ_0 and ϕ_{n-1}. We then use these two matrices to generate a set of n×2 matrices $u_2, u_3, \ldots u_{n/2}$. The (m + 1)th member of this set is generated from the mth and (m -1)th elements using the following recurrence:

$$u_{m+1} B_{m+1} = H u_m - u_m A_m - u_{m-1} B_m^{\dagger}$$
$$\text{where } A_{m-1} = u_m^{\dagger} H u_m \quad \text{and} \quad B_m^{\dagger} = u_{m-1}^{\dagger} H u_m \ . \qquad (3)$$

The A_m and B_m are 2×2 matrices. $u_{m+1}B_{m+1}$ is factored by requiring the columns of u_{m+1} to be mutually orthonormal. When this procedure is finished, the columns of the u_m form a new basis for the matrix eigenvalue problem of Eq. (2). The new basis has two important features. First, in this basis H is block tridiagonal: if H is divided into 2×2 blocks, the A_m form the diagonal blocks, the B_m and $B_m^{\dagger}$ form, respectively, the super- and sub-diagonal blocks, and all elements of all other blocks of H equal zero. Second, although most new basis elements no longer represent electrons localized on a single atom, the basis states in u_0 and u_1 still

represent electrons localized on atoms at the lead-device interface. This means that the boundary conditions can still be imposed on these components of the solution Ψ.

This transformation, in effect, replaces the network of Fig. 1 with a different network (Fig. 2), without altering the description of the leads. We call this new network an "effective quantum circuit," analogous to the effective circuit theory used to calculate the total impedance of a network of conventional resistors. This result is general—any device geometry can be transformed to a model like Fig. 2. More importantly, the transmittance of such an effective network is given by the following:

$$T(E) = 4 \sin^2 k \, | [e^{ik} I - v\, G(E)]^{-1}_{1,2} |^2, \quad \text{where} \tag{4a}$$

$$G(E) = [EI - A_1 - B_2^{\dagger}(EI - A_2 - \ldots)^{-1} B_2]^{-1} . \tag{4b}$$

k is the crystal momentum given by $E = \hbar^2 k^2/(2m^*)$. I is the 2×2 identity matrix. The subscript "1,2" signifies the off-diagonal element of the symmetric 2×2 matrix calculated in Eq. (4a).

This algorithm is descended from the Lanczos method of matrix diagonalization. The high stability and efficiency of this class of methods is well understood. [9] Briefly, a large rounding error is introduced in the transformation described by Eq. (3). This causes a rapid loss of orthogonality of the new basis vectors as the recursion progresses. These errors do not affect the calculated transmittance, however, because the new elements of H contained in the A_m and B_m are still accurate, to within a simple rounding error, for the new, nonorthogonal basis vectors.

In practice, the orthogonality loss prevents the recurrence relation Eq. (3) from terminating after n new vectors have been produced. That is, new vectors will continue to be generated, eventually leading to an overspecified basis. This presents no real problem; the matrix continued fraction (4b), which is highly stable, need only be continued until the transmittance converges. While this may take more than n levels; in practice it usually takes far less.

This method has impressive numerical performance. First, the number of operations required is proportional to n (most conventional schemes require n^3). Second, using Eqs. (3) and (4) provides transmittances correct to machine precision (that is, on random rounding architectures, the number of mantissa digits lost to rounding error will be approximately $\log_{16} \sqrt{n}$). Fig. 3 demonstrates the accuracy of this method by showing the calculated transmittance as a function of the number of levels used in the continued fraction (Eq. (4b)), for a model system whose transmittance is exactly known (dashed line). The calculation converges sharply to the correct value, once enough basis vectors are included.

We have tested the accuracy of this method for elastic scattering volumes containing up to 10^5 atoms (nodes). This exhausted our available computer memory. However, algorithmic advances and the availability of massively parallel architectures make much larger calculations possible. Multiple lead geometries can be accommodated by tridiagonalizing in larger blocks.

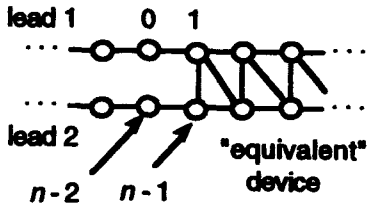


Figure 2. The block recursion method transforms the device model into an equivalent network without altering lead nodes.

IV. AN EXAMPLE: DIRTY INVERSION LAYERS

We have modeled inversion layers as clusters of atoms arranged in a two dimensional square lattice [3,4,5]. Fig. 4a shows conductance as a function of Fermi Energy (controlled by gate voltage) for a "clean" (pure crystalline) device. Fig. 4b represents a "dirty" layer (having a high concentration of defects, dopants, etc.) Both layers are 100 atoms square ($n = 10^5$), roughly equivalent to an elastic scattering area of a metallic layer at room temperature [7]. Here, we will only highlight some important results of this study.

In the clean sample, carriers with energies within the band conduct well because the electronic states of a crystal obey Bloch's theorem [7], and are thus extended through the layer and have high mobility. There is noticeable resistance, however; this comes from electron reflection and interference at the device walls and device-lead interfaces. At the band edge, conductance falls off sharply, since carriers above this energy must tunnel through the band gap.

Conduction in the disordered layer is weaker. Quantum states in a disordered system do not obey Bloch's theorem, and instead tend to destructively interfere, and hence decay with distance. This lowers carrier

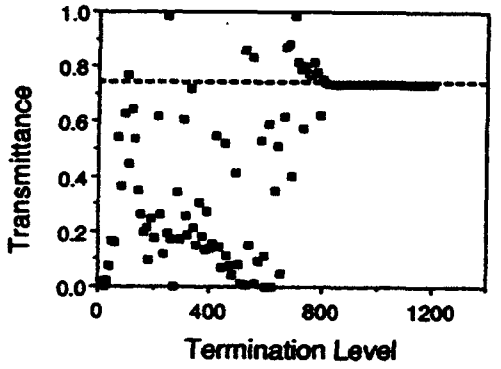


Figure 3. Convergence of the transmittance with dimension of the block-tridiagonal matrix.

mobility compared to a crystal [10]. Fig. 4b shows a "kink" in conductance, similar to the crystal band edge, at a "critical" energy E_C. However, Ec is not at the band edge ($E/v = 6.0$ for the disordered layer). We have shown that there are large numbers of allowed states for carriers with energies above E_C; however, their spatial decay length is much shorter than states with energies at or below E_C. That is, carrier mobilities change drastically when the Fermi Energy passes through E_C.

This result is surprising. Previously, it was widely believed that in disordered layers the strength of carrier state decay (and hence mobility) varied smoothly with carrier energy. Instead, we find a singularity at E_C. This mobility transition explains the sharp switching of even highly disordered FETs. If mobility varied smoothly with energy, current in such a device should switch gradually with gate voltage. Our results suggest, however, that the switching occurs when the Fermi energy sweeps past E_C, not the band edge. In other words, the current shuts off not because there are no carriers, but because carrier mobility has decreased sharply. This conclusion is supported by capacitance measurements in dirty inversion layers, which measure carrier densities as well as conductance [11]. Obtaining this result in a simulation, however, would be difficult without the sensitivity and resolution of the block recursion algorithm.

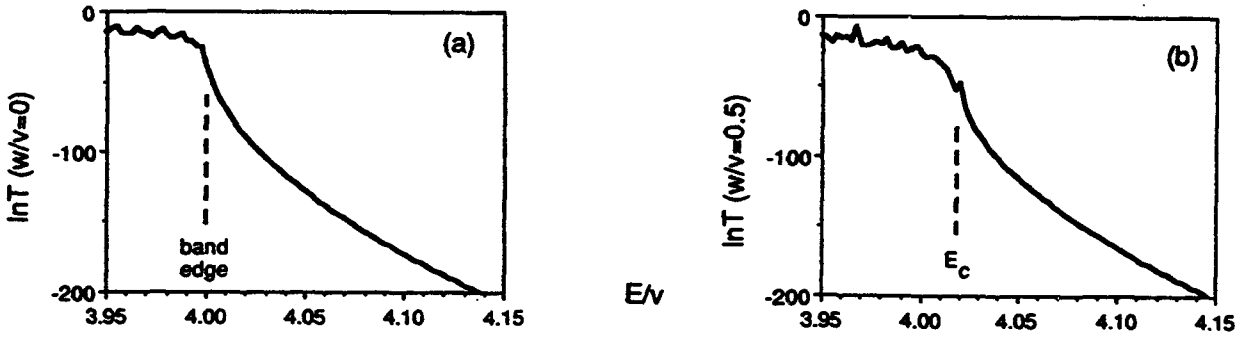


Figure 4. The variation of the logarithm of the transmittance with electron density for an ordered device (a), and a disordered device (b).

REFERENCES

1. S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer Verlag, New York, 1984).
2. J.C. Wu, M.N. Wybourne, A. Weisshaar, and S.M. Goodnick, J. Appl. Phys. 74 (7) 4590 (1993).
3. T.J. Godin and R. Haydock, Phys. Rev. B 46, 1528 (1992).
4. T.J. Godin and R. Haydock, Comput. Phys. Commun. 64, 123 (1991).
5. T.J. Godin and R. Haydock, Phys. Rev. B 38, 5237 (1988) and references therein.
6. R. Landauer, Phil. Mag. 21, 863 (1970).
7. N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (Holt Saunders, Philadelphia, 1976).
8. W.A. Harrison, *Electronic Structure and the Properties of Solids* (Dover, New York, 1989).
9. C.C. Paige, J. Inst. Math. Appl. 10, 372 (1972).
10. N.F. Mott and E.A. Davis, *Electronic Processes in Non-Crystalline Materials*, 2nd ed (Oxford University Press, London, 1979).
11. R.A. Davies and M. Pepper, J. Phys. C 15, L371.

Domain Decomposition Applied to the Drift-Diffusion Equations

W. M. COUGHRAN, JR.* Eric Grosse
AT&T Bell Laboratories
Murray Hill, NJ 07974, USA

For a number of years, AT&T Bell Laboratories has depended on Cray Research computers for semiconductor modeling. Recently, we started to explore a variety of domain decomposition methods employing a collection of high-performance workstations connected via 100Mb/s FDDI. We will describe our experience with such approaches for the drift-diffusion equations and our perceptions of the evolving workstation cluster alternatives.

*Speaker.

Adaptive Grid and Iterative Techniques for Submicron Device Simulation

Anand L. Pardhanani & Graham F. Carey
CFD Laboratory, WRW 301
The University of Texas, Austin, Texas 78712

Abstract

Numerical solution schemes based on adaptive grid redistribution and iterative techniques are investigated for submicron semiconductor devices. The grid adaptation strategy involves constructing a mathematical optimization problem to determine grid point locations. This is accomplished by defining an objective function which takes into account the numerical error and various geometric properties of the grid that affect accuracy. The iterative solution strategies considered here include multilevel nested iteration, and non-symmetric gradient-type iterative solvers for algebraic systems. Numerical tests are carried out using a non-parabolic hydrodynamic model for carrier transport.

Introduction

The use of adaptive gridding and iterative solution methods has proven very effective in many transport applications. Such techniques are particularly useful for problems involving multiple length scales or sharp solution gradients, which typically give rise to large algebraic systems. Semiconductor device applications are known to be grid-sensitive and they involve severe solution gradients, especially as device sizes shrink into the deep submicron regime. Thus, it is natural to consider iterative and adaptive grid approaches for device simulation.

Grid Adaptation

The present adaptive grid redistribution approach is developed using a discrete optimization model [2, 3]. This involves constructing a mathematical objective function that defines the desired grid properties, and then adjusting the grid to minimize this function. The optimization procedure is initiated by first computing an approximate solution using a simple initial grid, and then constructing an objective function based on a local feature or error indicator derived from this solution. The objective function takes into account properties of the numerical solution as well as geometric properties of the grid – for example, smoothness and orthogonality of grid lines – which may affect accuracy. The locations of the grid points constitute the unknowns in the objective function. For instance, a composite objective function which incorporates a measure of the solution error E, grid smoothness S and orthogonality O would be of the form

$$F(\boldsymbol{v}) = F(E(\boldsymbol{v}), S(\boldsymbol{v}), O(\boldsymbol{v})) \tag{1}$$

where $\boldsymbol{v}$ is a vector representing the grid point coordinates. Then the optimization problem can be stated as

$$\begin{aligned} &\text{minimize} \quad F(\boldsymbol{v}) \\ &\boldsymbol{v} \in C \end{aligned} \tag{2}$$

Here C is the constraint set, which represents the space of all feasible grids with a given surface and boundary configuration. This constraint is needed to ensure, for example, that the redistributed nodes remain within the device domain, and to preserve the boundary or interface geometry.

The objective function may be generalized to include additional properties, and there are a variety of approaches for constructing its individual components. For example, $S(v)$ and $O(v)$ in (1) may be defined using a geometric approach, which can be illustrated in 2D with the help of Figure 1. The figure shows a representative patch of four cells surrounding any interior grid point (i,j). Local functions for the smoothness and orthogonality may be defined as:

$$\begin{aligned} S_{ij} &= r_1 \cdot r_1 + r_2 \cdot r_2 + r_3 \cdot r_3 + r_4 \cdot r_4 \\ O_{ij} &= (r_1 \cdot r_2)^2 + (r_2 \cdot r_3)^2 + (r_3 \cdot r_4)^2 + (r_4 \cdot r_1)^2 \end{aligned} \tag{3}$$

where r_i are the vectors shown in Figure 1. Similarly, a local feature/error adaptivity indicator w_{ij} can be defined at (i,j) and scaled by the local Jacobian (which is related to the patch area) to construct the adaptivity measure

$$E_{ij} = w_{ij} J_{ij}^2 \tag{4}$$

The functions in (3) and (4) can be accumulated over the entire grid to obtain global measures, and the objective function may be defined as a linear combination:

$$F(v) = \alpha E(v) + \beta S(v) + \gamma O(v) \tag{5}$$

with α, β and γ being positive constants that may be chosen to enforce adaptivity, smoothness and orthogonality to varying degrees.

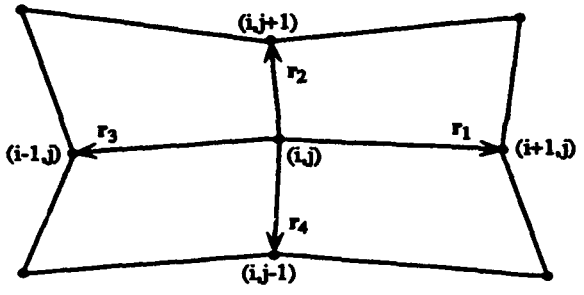


Figure 1: Representative interior patch of 4 cells around grid-point (i,j).

This formulation permits adapting the grid to the error or any desired feature of the problem. For example, in the semiconductor device problem a natural choice is to use the doping profile to adapt the initial grid as indicated in Figure 2.

Hydrodynamic Model and Discretization

The semiconductor device model used in our numerical experiments consists of the non-parabolic hydrodynamic transport system developed by Bordelon et al. [1]. The resulting coupled Poisson and transport equations that relate the electrostatic potential (ϕ), carrier density (n), velocity (v) and energy (w) have the following form:

$$\begin{aligned} \nabla \cdot (\epsilon \nabla \phi) &= q(n - N_D) \\ \frac{\partial n}{\partial t} + \nabla \cdot (nv) &= 0 \\ \frac{\partial (A(w) nv)}{\partial t} + \frac{2}{3m^*} \nabla (B(w) nw) - \frac{q}{m^*} n \nabla \phi &= -\frac{nv}{\tau_p} \\ \frac{\partial (nw)}{\partial t} + \nabla \cdot (\Omega nwv + Q) - qnv \cdot \nabla \phi &= -n \frac{(w - w_0)}{\tau_w} \end{aligned} \tag{6}$$

with $A(w) = 1 + 2\Omega\alpha w$, $B(w) = (1 + \alpha w)/(1 + 2\alpha w)$, Q = "heat flux" $= -nD_w\nabla w$, and with empirical constants $\alpha = 0.5eV^{-1}$, $\Omega = 1.3$ and $D_w \sim 10$ cm^2/sec.

In the present work, the equation system is first analytically mapped to a reference domain to facilitate discretization on general non-uniform grids. The spatial derivatives and metrics in the transformed equations are then discretized using second-order finite-differencing. We use an extension of the Scharfetter-Gummel treatment for the current density and energy flux. For the hydrodynamic system (6), following the usual procedure, we assume locally constant current density and electric field, and linear variation in energy. This yields an approximation of the form

$$\begin{aligned} nv &= \left[\frac{n_{i+1}}{w_{i+1}}B(-\Delta_J) - \frac{n_i}{w_i}B(\Delta_J)\right]\frac{c_1\frac{dw}{d\xi}}{\ln(w_{i+1}/w_i)} \\ \Delta_J &= \left(2 + (c_2\frac{d\phi}{d\xi})/(c_1\frac{dw}{d\xi})\right)\ln(w_{i+1}/w_i) \end{aligned} \tag{7}$$

where subscripts i and $i+1$ denote nodal values in the reference coordinate direction ξ, c_1 and c_2 are coefficients in the scaled momentum equation, and $B(x) = x/(e^x - 1)$ denotes the Bernoulli function. Similarly, assuming locally constant energy flux (S) and exponential n one can derive the approximation

$$\begin{aligned} S &= [w_{i+1}B(-\Delta_S) - w_iB(\Delta_S)]\frac{d_2\nu}{\delta\xi} \\ \Delta_S &= (d_1\delta\xi)/(d_2\nu) \\ \nu &= \ln(n_{i+1}/n_i)[n_{i+1}\ n_i/\ (n_{i+1} - n_i)] \end{aligned} \tag{8}$$

where d_1 and d_2 are coefficients in the scaled energy equation. The resulting semi-discrete equation system is integrated to steady state in a fully-coupled form using backward Euler or semi-implicit Runge-Kutta schemes.

Iterative solution

Solution techniques of special interest in the present work include multigrid-type schemes as well as generalized gradient iterative solvers. Multigrid techniques are based on the use of a sequence of nested grids over the given domain, and solving the problem using a cyclic iterative process which exploits each grid's preferential convergence behavior [4]. The resulting algorithm is usually far more efficient than traditional single-grid iterative algorithms. In this work we focus on nested grid schemes, wherein the solution process begins most economically at the coarsest level and proceeds to finer levels using projection to generate good starting iterates. Figure 3 shows sample results obtained using a two-grid scheme to solve the non-parabolic hydrodynamic model for a 0.08 micron (channel-length) $n^+ - n - n^+$ diode structure. In this example we first computed a drift-diffusion solution on the coarse grid for use as initial approximation to the coarse grid hydrodynamic solution. This was interpolated to the fine grid using an approach consistent with the Scharfetter-Gummel assumptions.

Generalized gradient iterative methods such as bi-conjugate gradients (BCG) and conjugate gradient squared (CGS) are also under investigation. Implicit integration of the fully-coupled hydrodynamic system yields large block-structured algebraic systems at each integration step, which must be solved efficiently for enhancing the efficiency of the overall simulation process. For a simple MOSFET test case such as the one shown in Fig. 2, preliminary calculations have shown an order of magnitude improvement on a 129×33 non-uniform grid when a CGS scheme is used instead of a band solver.

Acknowledgements

This research is being supported by the Semiconductor Research Corporation and the Texas Advanced Technology Program.

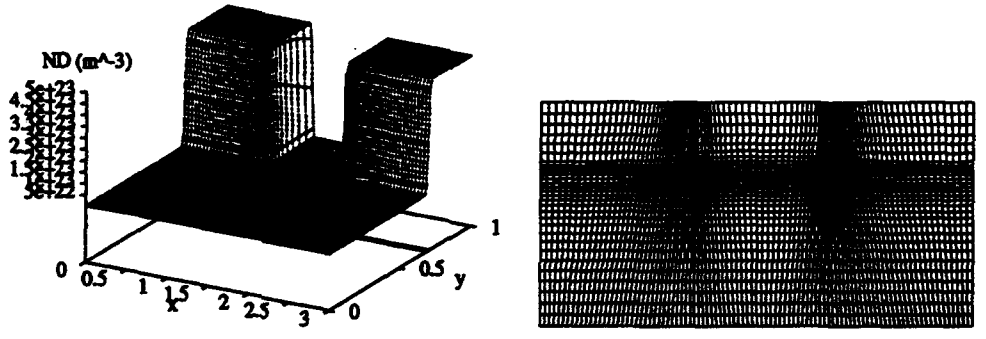
References

[1] T. J. Bordelon, X.-L. Wang, C. M. Maziar and A. F. Tasch, "Accounting for bandstructure effects in the hydrodynamic model: a first-order approach for silicon device simulation," *Solid-State Electronics*, vol. 35, no. 2, pp. 131-139, 1992.

[2] A. Pardhanani and G. F. Carey, "Optimization of computational grids," *Numerical Methods for Partial Differential Equations*, vol. 4, no. 2, pp. 95-117, 1988.

[3] A. Pardhanani and G. F. Carey, "Adaptive redistribution and multilevel techniques for time-dependent and steady problems," in *Multigrid Methods: Special Topics and Applications II*, GMD-Studien Nr. 189, GMD, Sankt Augustin, 1991.

[4] P. Wesseling, *An Introduction to Multigrid Methods*, John Wiley, Chichester (England), 1992.



(a) (b)

Figure 2: (a) Representative MOSFET doping profile; and (b) grid adapted to doping.

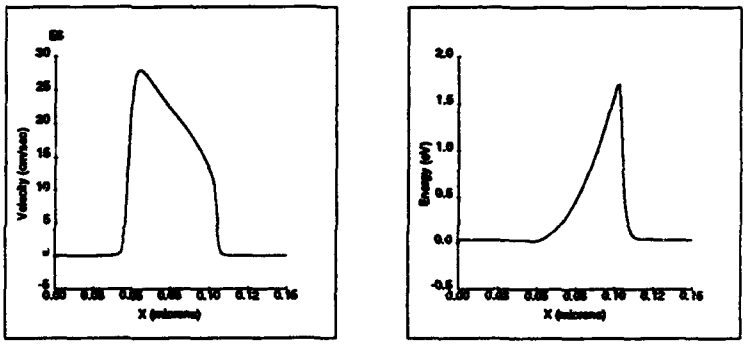


Figure 3: Hydrodynamic simulation of carrier velocity and energy at 3 volts bias for 0.08μm $n^+ - n - n^+$ diode with 129 grid points; doping concentration varies 5 orders of magnitude at junctions.

TRANSPORT EFFECT, HYPERBOLICITY, AND SHOCK CAPTURING ALGORITHMS FOR DEVICE SIMULATIONS

Joseph W. Jerome [1]
Department of Mathematics, Northwestern University, Evanston, IL 60208
and
Chi-Wang Shu [2]
Division of Applied Mathematics, Brown University, Providence, RI 02912

Abstract

We study the effect of the common practice of neglecting the convective terms (inertial approximation) in the hydrodynamic model in the simulation of n^+-n-n^+ diodes and two dimensional MESFET devices. We find that the inertial approximation is invalid near the diode junctions, and near the contact regions of the MESFET device. We also test the hyperbolicity of the first derivative part in the hydrodynamic model, and in related energy transport models. We find that the first derivative part of the system is hyperbolic, for the hydrodynamic model, the modified hydrodynamic model, and the energy transport model. This suggests and validates the use of shock capturing algorithms for the simulation.

I. TRANSPORT EFFECT.

In earlier work (see [5]), we have advocated using modern nonlinear hyperbolic based shock capturing algorithms (e.g., the ENO algorithm in [11]) in device simulations with hydrodynamic (HD) and energy transport (ET) models. Introductions to these models may be found in [10] and [7], respectively. The first use of such methods in device simulation was [2].

A common practice in the interpretation of the hydrodynamic model is to employ the inertial approximation, which in our terminology characterizes the transport effect as small if

$$\tau_p \sqrt{u_x^2 + u_y^2 + v_x^2 + v_y^2} << 1, \tag{1}$$

where τ_p is the momentum relaxation coefficient and (u, v) is the velocity vector. The reader can find this approximation employed in many reduced hydrodynamic models (cf. [9]). As discussed in [10], it allows the extension of the Scharfetter-Gummel method to a hydrodynamic model setting. This hypothesis is well known in fluid mechanics, where the resulting flow is termed a Stokes' flow. In the electrical engineering community, one speaks of neglecting the convective terms.

In this work, we simulate the standard one dimensional n^+-n-n^+ channel and a two dimensional MESFET, using the complete HD model, as introduced in [10], with Baccarani-Wordeman relaxation expressions (see [1]), and then check the validity of (1). From a physical point of view, we wish to check whether the transport effect is uniformly small. If it is, hyperbolic based algorithms probably need not be used, and the inertial approximation would appear justified. The numerical scheme we use is the ENO (Essentially Non-Oscillatory) scheme [11], adapted to device simulations in [2] and [5]. It has the advantage of both high order accuracy and monotone sharp gradient transitions.

The one dimensional n^+-n-n^+ channel is a standard silicon diode with a length of $0.6\mu m$, with a doping defined by $n_d = 5 \times 10^5 \mu m^{-3}$ in $[0, 0.1]$ and in $[0.5, 0.6]$, and $n_d = 2 \times 10^3 \mu m^{-3}$ in $[0.15, 0.45]$,

[1] Research supported by the National Science Foundation under grant DMS-9123208.
[2] Research supported by the National Science Foundation under grant ECS-9214488 and the Army Research Office under grant DAAL03-91-G-0123. Computation supported by the Pittsburgh Supercomputer Center.

joined by smooth junctions. The lattice temperature is taken as $T_0 = 300$ K. We apply a voltage bias of *vbias* = 0.5V, 1.0V and 1.5V, respectively. Other relevant parameters can be found in [5]. We use a high order ENO scheme (third order) and a very refined grid (200 points), in order to ensure that the physical model is fully resolved by our numerical result.

Fig. 1 (left) clearly shows that the transport effect (the quantity in Eqn. (1)) is *not* small near the junctions. In order to verify that this is not an artifact of the spurious velocity overshoot at the right junction, we also simulate with a reduced heat conduction coefficient $\kappa_0 = 0.5$ to reduce this spurious overshoot (see [4]). The result, Fig. 1 (right), shows significant transport effect at the junctions, especially at the left junction.

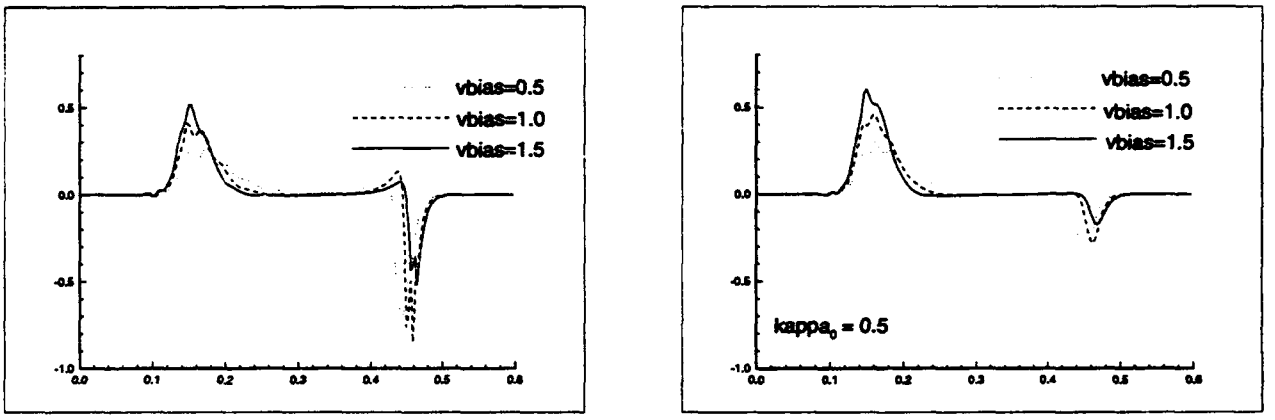


Fig. 1: The transport effect $\tau_p u_x$ for the one dimensional n^+-n-n^+ channel. Left: the HD model; right: the HD model with a reduced heat conduction coefficient $\kappa_0 = 0.5$.

To see the effect of ignoring this transport effect and using a reduced hydrodynamic model, we also make the simulation of the same diode using the reduced HD model with the inertial assumption. This reduced HD model does not have a momentum equation, is a fully parabolic system, hence is much easier to solve numerically. The velocity is a derived quantity from the concentration and energy. We can see from Fig. 2 that the reduced HD model underestimates the velocity.

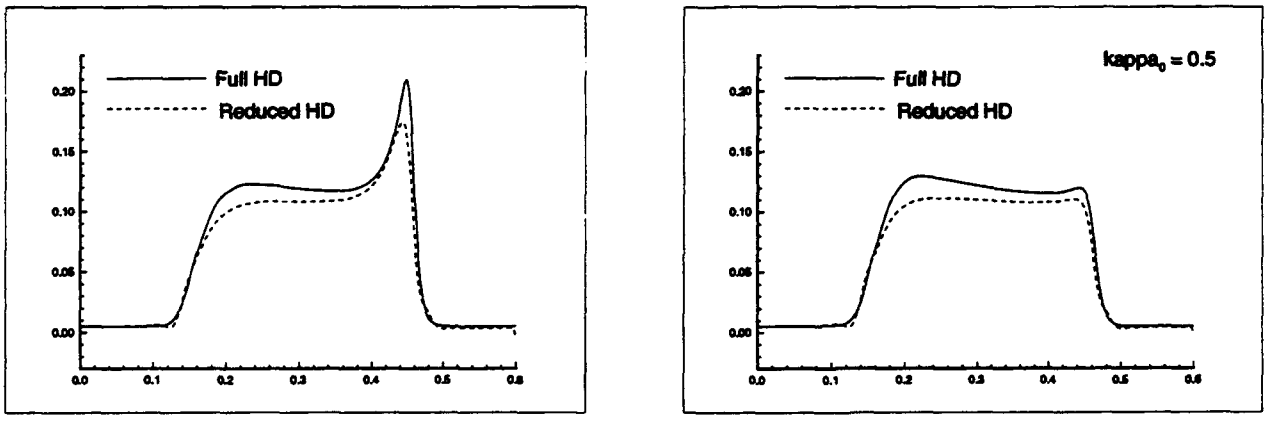


Fig. 2: Velocity at *vbias* = 1.5. Left: the HD model and the reduced HD model; right: the same with a reduced heat conduction coefficient $\kappa_0 = 0.5$.

Next we simulate a two dimensional MESFET of the size $0.6 \times 0.2\mu m^2$. The geometry as well as the doping (in μm^{-3}) is shown in Fig. 3, left. We apply, at the drain, a voltage bias

vbias = 2V. The gate is a Schottky contact, with a negative voltage bias *vgate* = −0.8V and a very low concentration value $n = 3.8503 \times 10^{-8}\mu m^{-3}$. The lattice temperature is again taken as $T_0 = 300$ K. A high order (third order) ENO scheme with a very refined grid of 192×64 points is used. Again, this is to ensure that the physical model is fully resolved by the numerical scheme. Boundary conditions and other parameters can be found in [5].

Fig. 3, right, shows that the transport effect is *not* small near the contacts. For easy presentation, we have listed an integer at every other grid point in the MESFET in Fig. 3, right. This integer is ten times the transport effect formula in Eqn. (1), capped from above by nine.

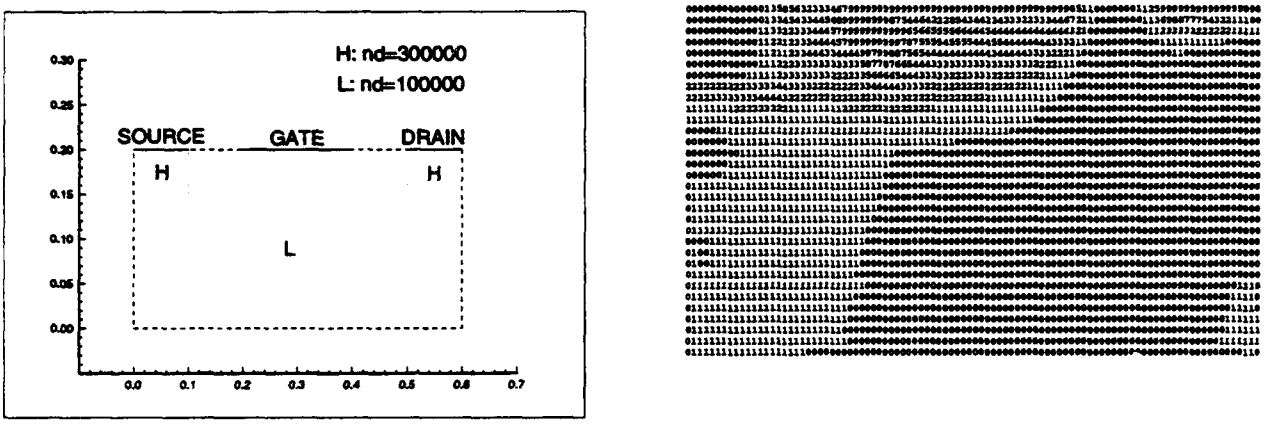


Fig. 3: Two dimensional MESFET. Left: the geometry and the doping n_d; Right: the transport effect. The integers denote the integer part of ten times the transport effect: $\left[10\tau_p\sqrt{u_x^2 + u_y^2 + v_x^2 + v_y^2} \right]$. If it is larger than 9, then 9 is shown.

From these two examples we can conclude that, in device simulations, the transport effect is not uniformly small. This justifies the usage of hyperbolic based shock capturing schemes (e.g., [5]) for device simulation. It also justifies the usage of the full HD model.

II. TEST FOR HYPERBOLICITY.

In this section, we would like to discuss the hyperbolicity check of the first derivative part in the following Eqn. (2), for the HD model, the modified HD model in [13], and for the ET model in [7]. Both HD and ET models can be expressed in the following form:

$$w_t + f_1(w)_x + f_2(w)_y = r(w), \tag{2}$$

where the right-hand-side $r(w)$ contains both the forcing terms due to the relaxation, which are nonlinear functions of w, and the second derivative terms due to the heat conduction. The analysis of the first derivative component of the HD and ET models should not be confused with the mathematical classification of the complete systems (2). Exclusive of the Poisson equation, these are classified as parabolic/hyperbolic and parabolic, respectively. Such classification can be found in, e.g., [3]. Since we are interested in the situation that higher derivatives in the system have relatively small coefficients, and first derivative terms are either dominant or at least are significant, we will study only the first derivative part.

The definition of the first derivative part $f_1(w)_x + f_2(w)_y$ as hyperbolic is: $\xi_1 f_1'(w) + \xi_2 f_2'(w)$ with real ξ_1 and ξ_2, has only real eigenvalues and a complete set of eigenvectors. If the first derivative part is hyperbolic, and if the first derivative part dominates the system or is at least significant, then hyperbolic based algorithms (like ENO) can be very effective. On the other hand,

if the first derivative part is not hyperbolic, the system is of mixed hyperbolic-elliptic type, and the mathematical theory about the solution to (2), when the right-hand-side tends to zero, is very complicated and in many cases still unsolved. Likewise, numerical methods for such mixed type systems are also complicated and under developed (see e.g., [12]). We would thus desire to avoid the appearance of mixed type first derivative part when modifying the models. Notice also that in many modifications to the hydrodynamic models (e.g., [13],[8]), the right-hand-side of Eqn. (2) is changed to contain some first derivative terms also. Although in practical computations these terms are treated as small perturbations and approximated separately, the justification that these terms are indeed "small" can only come from moving these terms to the left-hand-side, absorbing them into $f_1(w)$ and $f_2(w)$, and then checking hyperbolicity. We have performed such a hyperbolicity check for the standard HD model, the modified HD model [13], and the ET model [7]. It is found that all three cases have hyperbolic first derivative parts. The details can be found in [6]. We have to resort to numerical techniques to check the hyperbolicity of the first derivative part for the modified HD model in [13].

IV. REFERENCES.

[1] G. Baccarani and M.R. Wordeman, *An investigation of steady-state velocity overshoot effects in Si and GaAs devices*, Solid State Electr., v28, pp.407–416, 1985.

[2] E. Fatemi, J.W. Jerome, and S. Osher, *Solution of the hydrodynamic device model using high-order nonoscillatory shock capturing algorithms*, IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, CAD-10, pp.232–244, 1991.

[3] C. L. Gardner, *The quantum hydrodynamic model for semiconductor devices*, SIAM J. Appl. Math., v54, pp.409–427, 1994.

[4] A. Gnudi, F. Odeh and M. Rudan, *Investigation of non-local transport phenomena in small semiconductor devices*, European Trans. on Telecommunications and Related Technologies, v1(3), pp.307–312 (77–82), 1990.

[5] J.W. Jerome and C.-W. Shu, *Energy models for one-carrier transport in semiconductor devices*, in IMA Volumes in Mathematics and Its Applications, v59, W. Coughran, J. Cole, P. Lloyd and J. White, eds., Springer-Verlag, to appear.

[6] J.W. Jerome and C.-W. Shu, *Transport effects and characteristic modes in the modeling and simulation of submicron devices*, preprint, 1994.

[7] E.C. Kan, D. Chen, U. Ravaioli and R.W. Dutton, *Formulation of macroscopic transport models for numerical simulation of semiconductor devices*, preprint.

[8] S.-C. Lee and T.-W. Tang, *Transport coefficients for a silicon hydrodynamic model extracted from inhomogeneous Monte-Carlo calculations*, Solid State Electr., v35, pp.561–569, 1992.

[9] B. Meinerzhagen, R. Thoma, H.J. Peifer, and W.L. Engel, *On the consistency of the hydrodynamic and the Monte Carlo models*, Proceedings of the Second International Workshop on Semiconductors, pp. 7–12, Beckman Institute, University of Illinois, 1992.

[10] M. Rudan and F. Odeh, *Multi-dimensional discretization scheme for the hydrodynamic model of semiconductor devices*, COMPEL, v5, pp.149–183, 1986.

[11] C.-W. Shu and S.J. Osher, *Efficient implementation of essentially non-oscillatory shock capturing schemes, II*, J. Comput. Physics, v83, pp.32–78, 1989.

[12] C.-W. Shu, *A numerical method for systems of conservation laws of mixed type admitting hyperbolic flux splitting*, J. Comput. Physics, v100, pp.424–429, 1992.

[13] M.A. Stettler, M.A. Alam, and M.S. Lundstrom, *A critical examination of the assumptions underlying macroscopic transport equations for silicon devices*, IEEE Trans. Electron Devices, v40, pp.733–740, 1993.

NUMERICAL SIMULATION OF HIGH-FIELD TRANSPORT USING A FLUX-CORRECTED TRANSPORT ALGORITHM

M.G. Ancona and C.R. DeVore
Naval Research Laboratory
Washington, DC 20375
ancona@estdsun4.nrl.navy.mil

Abstract

Hydrodynamic simulations of high-field transport are performed using a flux-corrected transport algorithm. This efficient algorithm uses careful control of numerical diffusion to achieve high accuracy in simulating flow phenomena in the presence of steep gradients as can occur in small devices where overshoot phenomena are significant. We apply the flux-corrected transport scheme to a preliminary evaluation of various hydrodynamic descriptions of high-field transport.

I. INTRODUCTION

Continuum or hydrodynamic descriptions of electron transport have long been applied in the analysis and design of semiconductor devices because they provide a useful compromise between computational simplicity and physical fidelity. As devices continue to scale deep into the sub-micron regime such descriptions will continue to be of value although ultimately they must break down. The transport in deep submicron regimes is often characterized by high electric fields, rapidly varying densities and history-dependent phenomena (including inertia) which make the governing equations more hyperbolic in character. From a computational standpoint these factors and particularly the need to obtain accurate results in the vicinity of steep gradients represent significant challenges. One numerical approach capable of handling these difficulties which has been widely used for fluid simulation in other fields is flux-corrected transport (FCT) [1]. In essence, FCT is an explicit, spatially high-order finite-difference scheme in which a conservative "flux-limiting" procedure is used to prevent the otherwise inevitable unphysical ripples which would appear in the numerical solution near steep gradients. In this work, we apply an FCT algorithm to solving hydrodynamic equations describing electron transport in small-geometry n^+-n-n^+ silicon diodes.

In addition to studying the numerical issues, a primary purpose of our effort is to examine and give a preliminary evaluation of various hydrodynamic descriptions of high-field electron transport. Such descriptions are founded on a continuum approximation, i.e., that meaningful density variables can be defined, and in electron transport work are typically derived by taking velocity moments of the Boltzmann equation. Alternatively, hydrodynamic descriptions may be developed using classical field theory [2]. The former approach emphasizes the connections to the underlying microscopic physics whereas the latter, which takes the density variables as *primitives*, focuses on the consequences of general principles of balance, invariance and symmetry, i.e., on what is physically possible given a certain set of primitive densities. Obviously, the larger this set the more physics can be described at the expense of utility. In Sec. II we outline the equations which stem from standard choices for the density variables and then discuss numerical methods and solutions in Secs. III and IV.

II. HYDRODYNAMIC MODELS OF ELECTRON TRANSPORT

We consider describing the flow of the population of conduction band electrons through a semiconductor as the flow of a single fluid through a solid. The primitives of the theory are therefore the quantities which define this electron fluid and its interaction with itself, with the lattice and with the electrostatic field at every point. As a first case, we assume that the fluid is describable by the primitives of mass/charge density, momentum density and energy density. The laws of mass,

momentum and energy balance and of electrostatics then lead to a set of equations constraining these densities as follows [2]:

$$\text{(1a)}\quad n_{,t} + \nabla\cdot(n\mathbf{u}) = 0\,, \qquad mn\frac{d\mathbf{u}}{dt} = -\nabla p^n + \nabla\cdot\tau^n - qn(\mathbf{E}+\mathbf{E}^n)\,, \qquad \nabla\cdot\mathbf{D} = q(N_D - n)\,,$$

$$\text{(1b)}\quad mn\frac{d\varepsilon^n}{dt} = -\nabla\cdot\mathbf{q}^n + \tau^n : \mathbf{d} + \frac{p^n}{n}\frac{dn}{dt} + qn\mathbf{E}_d^n\cdot\mathbf{u} + mns^n\,, \qquad \rho\frac{\partial\varepsilon^l}{\partial t} = -\nabla\cdot\mathbf{q}^l + \mathbf{E}\cdot\frac{\partial\mathbf{P}}{\partial t} + qn\mathbf{E}_r^n\cdot\mathbf{u} + \rho s^l\,,$$

where n, $\mathbf{u}$, $mn\varepsilon^n$, $\mathbf{q}^n$, mns^n are the number density, velocity, energy density, heat flux and energy source density of the electron gas, ρ, N_D, $\rho\varepsilon^l$, $\mathbf{q}^l$ and ρs^l are the mass density, ionized impurity density, energy density, heat flux and energy source density of the lattice, $\mathbf{E}_r^n$ and $\mathbf{E}_d^n$ are the recoverable and dissipative parts of the force (per charge) exerted by the lattice on the gas, p^n and τ^n are the electron gas pressure and viscous stress, $\mathbf{E}$ and $\mathbf{D}$ ($=\mathbf{P}+4\pi\mathbf{E}$) are the electric field and electric displacement, $\mathbf{d}$ is the rate-of-deformation tensor and d/dt is the material derivative.

The differential equations (1) represent a set of physical constraints on the density variables; they are not sufficient however to determine these densities. To make the system determinate constitutive equations specifying the material response must also be supplied. For example, the usual energy transport (ET) model [3] results if we select the following constitutive equations

$$\text{(2)}\quad p^n = kT^n n\,, \qquad \tau^n = 0\,, \qquad \mathbf{E}_r^n = 0\,, \qquad \mathbf{E}_d^n = \frac{\mathbf{u}}{\mu_{LF}}\frac{T^n}{T^l}\,,$$

$$m\varepsilon^n = \frac{3}{2}kT^n\,, \qquad \mathbf{q}^n = -\frac{3}{2}D_{LF}nk\nabla T^n\,, \qquad mns^n = -\rho s^l = -\frac{n}{\tau_w}\left[\frac{m^*}{2}\mathbf{u}\cdot\mathbf{u} + \frac{3k}{2}(T^n - T^l)\right],$$

where m^* is the electron effective mass, T^n and T^l are electron and lattice temperatures, μ_{LF} and D_{LF} are the low-field mobility and diffusivity and τ_w is an energy relaxation time. Now, as discussed in Ref. 2, *when heat conduction is small* the density variables of mass and momentum become adequate to describe the system. In this case, the energy balance equations (1b) need not be solved and the governing equations become (1a) plus constitutive equations which in Ref. 2 were selected as

$$\text{(3a)}\quad p^n = kT^l n\,, \qquad \tau^n = \lambda_v \mathbf{I}\nabla\cdot\mathbf{u} + 2\mu_v\mathbf{d}\,, \qquad \mathbf{E}_r^n = -\chi\frac{d\mathbf{u}}{dt}\,, \qquad \mathbf{E}_d^n = \frac{\mathbf{u}}{\mu}\,,$$

where

$$\text{(3b)}\quad \mu = \frac{2\mu_{LF}}{1 + \sqrt{1 + \dfrac{\mu_{LF}\hat{E}}{\sqrt{2}u_{sat}} + 4\left(\dfrac{\mu_{LF}\hat{E}}{u_{sat}}\right)^2}}\,, \qquad \hat{E} \equiv \left|\mathbf{E} - \tau_r\frac{d\mathbf{E}}{dt}\right|\,,$$

λ_v and μ_v are viscosity coefficients, χ [$\equiv(m^*-m)/q$] is a drag rate coefficient arising from Bragg reflection, u_{sat} is the saturation velocity and τ_r is a "scattering equilibration time". We note the important inclusion in these equations of i) viscous effects and ii) memory or rate effects in the scattering (including as the origin of effective mass). The mobility model in (3b) is that of Ref. 5 with a rate term introduced to represent the delay associated with scattering. This reduced set of *electrohydrodynamic* (EHD) equations, which may be regarded as a physically well-founded version of Thornber's augmented diffusion-drift description [4], has obvious computational advantages and will be explored in our simulations below.

III. FLUX-CORRECTED TRANSPORT

Flux-corrected transport (FCT) is a powerful numerical method for integrating generalized continuity equations [1] which has been widely used for fluid simulation in other fields but has not been applied heretofore to semiconductor transport problems. FCT is an explicit, spatially high-order finite-difference scheme which is especially effective at providing high-accuracy solutions in the vicinity of steep gradients without exhibiting the unphysical ripples often seen in conventional schemes as a result of numerical dispersion. It accomplishes this by carefully controlling the amount of numerical diffusion in the scheme using a conservative "flux-correction" procedure which preserves monotonicity with maximal accuracy. Explicitly, FCT first computes provisional values for

the density (of mass, momentum or energy) $\hat{\rho}_i^k$ at the next time step k at each mesh point i using a low-order, strongly-diffused scheme. It then improves the accuracy of these values by removing as much of the numerical diffusion as possible without generating new or accentuating existing extrema. This is done in a conservative manner via

$$\rho_i^k = \hat{\rho}_i^k - f_{i+1/2}^k + f_{i-1/2}^k , \tag{4a}$$

where

$$f_{i+1/2}^k = S \cdot \max\left\{0, \min\left[S \cdot \left(\hat{\rho}_{i+2}^k - \hat{\rho}_{i+1}^k\right), \mu_{i+1/2}\left(\hat{\rho}_{i+1}^k - \hat{\rho}_i^k\right), S \cdot \left(\hat{\rho}_i^k - \hat{\rho}_{i-1}^k\right)\right]\right\} \tag{4b}$$

is the corrected flux, $S \equiv \mathrm{sgn}\left(\hat{\rho}_{i+1}^k - \hat{\rho}_i^k\right)$ and $\mu_{i+1/2}$ is an antidiffusion coefficient chosen to minimize the residual numerical error [6]. That the corrections are functions of the solution means that the scheme is nonlinear. We note that the idea behind FCT has been incorporated in a number of other numerical schemes known collectively as nonlinear monotone methods. Among these are the essentially non-oscillatory (ENO) schemes [7] which have been applied to semiconductor transport problems [8]. All of these methods have similar advantages; we believe FCT to be preferred only because of its conceptual simplicity which enables, for example, straightforward generalization to more than one dimension [9]. In the calculations of this paper, FCT is used to solve the hydrodynamic equations and the electrostatics is solved conventionally in a Gummel iteration.

IV. SIMULATION RESULTS

As a test problem we model the standard n^+-n-n^+ diode with the geometry and doping levels chosen to match those of Refs. 10 and 11. A one-dimensional boundary value problem modeling this structure is readily formulated in the ET [(1) with (2)] and EHD [(1a) with (3)] descriptions. We first solve this problem using FCT in the familiar ET case. The calculated steady-state velocity profile for a 0.4μm diode biased at 1.5V is shown in Fig. 1 along with the electron temperature profile. These results are essentially the same as those obtained in Ref. 11 using an implicit scheme. The ET description exhibits velocity overshoot which is qualitatively reasonable apart from the well-known spurious peak seen near the anode. The origin of the latter has been widely discussed and is not of interest here. However, in one additional run (Fig. 1) we included viscosity [using τ^n of (3a)] in the ET simulation and found that the spurious peak largely disappears indicating that viscosity needs to be considered if the ET model is to be fully understood. In any event, our main point is that FCT provides an efficient scheme for performing conventional energy transport simulations.

We next apply FCT to solving the EHD equations. Considering the same problem as in Fig. 1, the qualitatively reasonable result shown in Fig. 2 is obtained. We note that the EHD simulation shows no evidence of the spurious peak seen in the ET simulation. In Fig. 2 we also give an

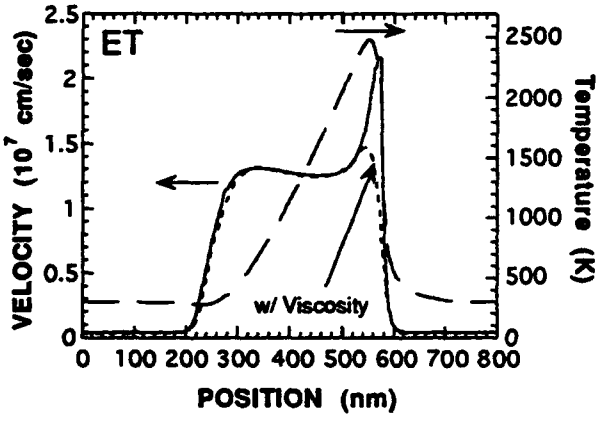


Fig. 1. ET velocity (with and without viscosity) and temperature profiles.

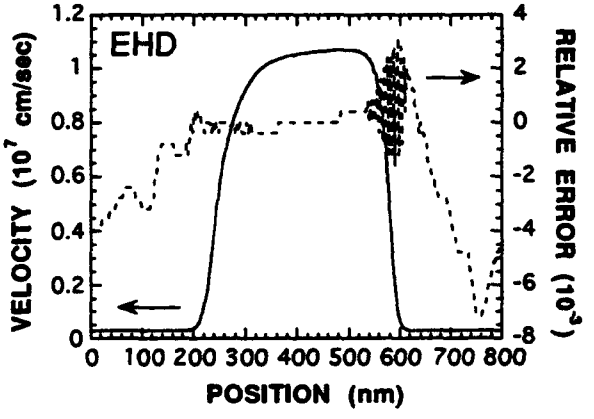


Fig. 2. Profiles of the EHD velocity and the relative error in the steady-state current.

indication of the numerical error introduced by the FCT scheme. We plot the deviations from uniformity in the steady state current as a relative error; the high accuracy possible with an FCT algorithm (here in single precision) is evident. Next, in Fig. 3 we compare the ET and EHD descriptions with profiles computed using diffusion-drift theory (DD) and by Monte Carlo solution of the Boltzmann equation (MC) [10]. This calculation is for a 0.1μm diode biased at 1V and the prediction of DD theory shows that the diode is operating in a strong velocity overshoot regime. In comparison with the "exact" MC solution, the EHD description is seen to do quite well both qualitatively and quantitatively. In this calculation, the one fitting parameter is the choice of τ_r to be 0.13psec. The ET description does significantly less well both in shape and magnitude, however, it should be said that no effort to adjust parameters such as the thermal conductivity has been made in the simulation. Finally, in Fig. 4 we exhibit the important roles of viscous and rate effects in the EHD description. The viscous effects smooth velocity gradients via dissipation, an effect which is partially offset by the rate effects which steepen the solution by delaying the onset of velocity saturation.

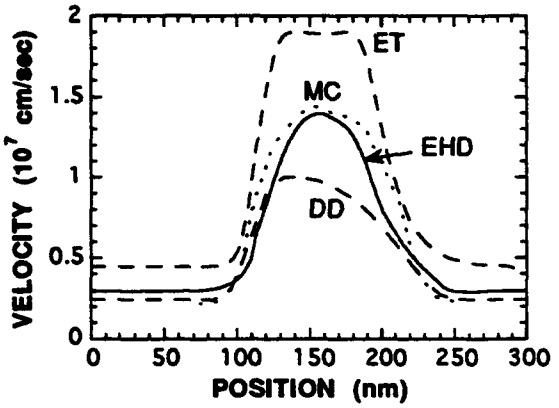


Fig. 3. Comparison of ET and EHD velocity profiles with diffusion-drift (DD) and Monte Carlo (MC).

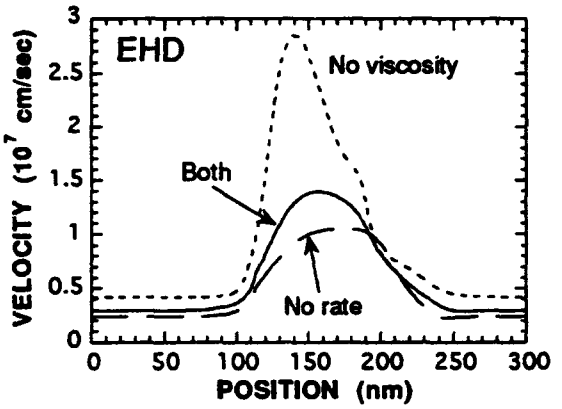


Fig. 4. EHD velocity profiles computed with and without mobility rate and viscous terms.

V. CONCLUSIONS

The main conclusion from this work is that the FCT algorithm provides a robust numerical approach to solving hydrodynamic equations descriptive of high field transport in semiconductors. It is a conceptually simple approach for which there exists a wide body of experience and software. It provides efficient, high accuracy solutions in the presence of steep gradients. and it is readily extendable to more than one space dimension. In this work, we applied this algorithm to the study of two high-energy transport theories. In the context of modeling overshoot phenomena, we find that an electrohydrodynamic description in which an energy balance equation is not solved gives accurate solutions with significant computational savings. Although these results are promising it is clear that more work is needed to fully validate this description and to determine its precise limitations.

Acknowledgements: The authors thank the Office of Naval Research for funding support.

1. J.P. Boris and D.L. Book, *J. Comp. Phys.* **11**, 38 (1973).
2. M.G. Ancona, to be published, *VLSI Design* (1994).
3. K. Blotekjaer, *IEEE Trans. Elect. Dev.* **ED-17**, 38 (1970).
4. K.K. Thornber, *IEEE Elect. Dev. Lett.* **EDL-3**, 69 (1982).
5. D. Roychoudhury and P.K. Basu, *Solid-St. Electron.* **19**, 656 (1976).
6. J.P. Boris and D.L. Book, in *Methods in Comp. Phys.* **16**, Academic, New York, p. 85 (1976).
7. A. Harten and S. Osher, *SIAM J. Numer. Anal.* **24**, 279 (1987).
8. E. Fatemi, J. Jerome and S. Osher, *IEEE Trans. Comp.-Aided Design* **CAD-10**, 232 (1991).
9. S.T. Zalesak, *J. Comp. Phys.* **31**, 335 (1979).
10. E.C. Kan, U. Ravaioli and T. Kerkhoven, *Solid-St. Electron* **34**, 995 (1991).
11. C.L. Gardner, J. Jerome and D. Rose, *IEEE Trans. Comp.-Aided Design* **CAD-8**, 501 (1989).

PARALLEL SOLUTION OF ELLIPTIC BOUNDARY VALUE PROBLEMS

Babak Bagheri, Andrew Ilin, L. Ridgway Scott, and Dexuan Xie
The Texas Center for Advanced Molecular Computation
and the Department of Mathematics, University of Houston

Abstract

We describe the development of some parallel iterative techniques for solving boundary value problems for elliptic partial differential equations. Using domain decomposition techniques, we modify standard sequential iterative techniques to obtain effective parallel methods. We contrast implementations on distributed-memory and shared-memory scalable parallel processors. We describe the use of two different programming paradigms, one involving explicit parallelism in a distributed-memory model and the other utilizing simple loop decompositions in a shared-memory model. Our primary conclusion is that parallel computing on existing commercial parallel supercomputers makes it routine to do three-dimensional modeling of semiconductor devices using drift-diffusion models. The implications this has for the use of more realistic models of submicron devices using Boltzmann-type equations will be mentioned.

I. Introduction

We discuss several techniques for solving elliptic boundary value problems via iterative methods which have a high degree of parallelism. These techniques are being developed to solve as broad a class of problems as possible, but our primary motivation has come from computing the electrostatic potential around molecules of biological significance [8]. Moreover, implementation of the methods has been done as part of an existing code UHBD [5]. This makes the code development more complex but also provides an assessment more realistic than would be available by looking only at computational kernels. In addition, we have applied some of the computational techniques to solve prototypical problems related to semiconductor device simulation [3].

We have studied several variants of standard iterative methods which we have designed to have good parallelism. These include variants of the well known ICCG and SOR iterative methods. In addition, we have proposed new types of iterations especially suitable for parallel computation [12]. We anticipate that all of these methods will be useful as coarse grid solvers for parallel multigrid methods [9].

In addition to studying different parallel iterative methods, we have used different parallel programming paradigms. Two of these are (1) IPfortran [1] and (2) shared memory constructs supported by Kendall Square's KSR-1 Fortran [10]. Both approaches have proved adequate for implementing the parallel algorithms presented here, due to the high degree of regularity of the loops involved. Less regular loops in UHBD, related to its Brownian dynamics phase, have been easier to parallelize using shared-memory constructs [4].

II. PSOR

The Jacobi method for approximating the solution of a linear system is naturally parallel, but the typically more efficient Gauss-Seidel method is essentially sequential. In the Jacobi method, each component X_i of the approximate solution vector $X = (X_1, \ldots, X_N)$ can be computed separately of all others, which we can write schematically as

$$X_i^{k+1} = F_i(X_1^k, \ldots, X_N^k), \quad \text{for} \quad i = 1, 2, \ldots, N, \tag{2.1}$$

where the F_i are functions of N variables. For example, $F = (F_1, \ldots, F_N)$ is an affine function in the case of solving a linear system. Typically F is sparse, depending only on entries near the diagonal, which we indicate by $F_i(\ldots, X_{i-1}, X_i, X_{i+1}, \ldots)$. With Gauss-Seidel, it is frequently the case that X_{i+1}^k depends on X_i^k: schematically it is

$$X_i^{k+1} = F_i(\ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \text{for} \quad i = 1, 2, \ldots, N. \tag{2.2}$$

The same applies for the SOR method, which is just a relaxed (or accelerated) version of Gauss-Seidel.

One approach taken to deal with the sequential nature of SOR is to reorder the unknowns so that one group of components X_i can be computed independently of others. This is often referred to as a coloring of the index set. The most well known case is that of two colors, usually called "red-black" ordering since it is similar to a chess board in simple cases. While this can be quite effective, it requires communication to be done for each color as opposed to just once for each iteration, as is the case for the Conjugate Gradient (CG) method. The number of colors required depends on the extent of the sparcity of F.

A simple technique used in practice is to decompose the index domain (the set of indices i) in a way to minimize the communication (either the number of messages required, or the size) among neighboring domains. Gauss-Seidel (or SOR) is used within each domain, without updating using the appropriate neighboring values. In the two-processor case, it takes the form

$$\begin{aligned} X_i^{k+1} &= F_i(\ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \forall\, i,\ 1 \le i \le N/2, \\ X_i^{k+1} &= F_i(\ldots, X_{N/2}^k, X_{N/2+1}^{k+1}, \ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \forall\, i,\ \frac{N}{2} + 1 \le i \le N. \end{aligned} \tag{2.3}$$

Once the local Gauss-Seidel (or SOR) sweep is done, neighboring values are exchanged, similarly to what would be done in the Jacobi iteration. For this reason, we refer to this method as the Jacobi-Gauss-Seidel (JGS) algorithm (or JSOR for its accelerated or relaxed variant). While appealing for its simplicity, this algorithm frequently requires a much larger number of iterations than the sequential case.

Remarkably, a simple alternative [14] to JGS and JSOR has convergence properties similar to the sequential case, but with communication features similar to JGS/JSOR. We will not attempt a complete description of the most general case, but will simply describe an example and present

numerical results. Consider the following algorithm:

$$\begin{aligned} X_i^{k+1} &= F_i(\ldots, X_{i-1}^{k+1}, X_i^k, \ldots, X_{N/2}^k, X_{N/2+1}^{k+1}, \ldots) \quad \forall\, i,\ 1 \le i \le N/2, \\ X_i^{k+1} &= F_i(\ldots, X_{N/2}^k, X_{N/2+1}^{k+1}, \ldots, X_{i-1}^{k+1}, X_i^k, X_{i+1}^k, \ldots) \quad \forall\, i,\ \frac{N}{2}+1 \le i \le N. \end{aligned} \tag{2.4}$$

This algorithm, which we call PGS (and PSOR for its accelerated or relaxed variant) is parallel for sparse F to the extent that the values $X_{N/2}^{k+1}, \ldots, X_N^{k+1}$ which are produced by the processor computing the second line can be computed and made available to the processor computing the first line before they are needed. In the case that the functions F_i are suitably sparse, this constraint poses no practical limitation to parallelism.

Figure 2.1 shows performance analysis for calculations done with the 5-point discretization of Laplace's equation using a strip decomposition (algorithm (2.4) in the case of two processors). We use this type of performance analysis graph to isolate different parts of a code. The computation time decreases even superlinearly [4] whereas the communication time (due to the use of a strip decomposition) remains nearly constant. The category "other time" simply reflects the part of the total time that cannot be accounted for in either of these categories; in this case it is quite small (being less that a second for two and four processors).

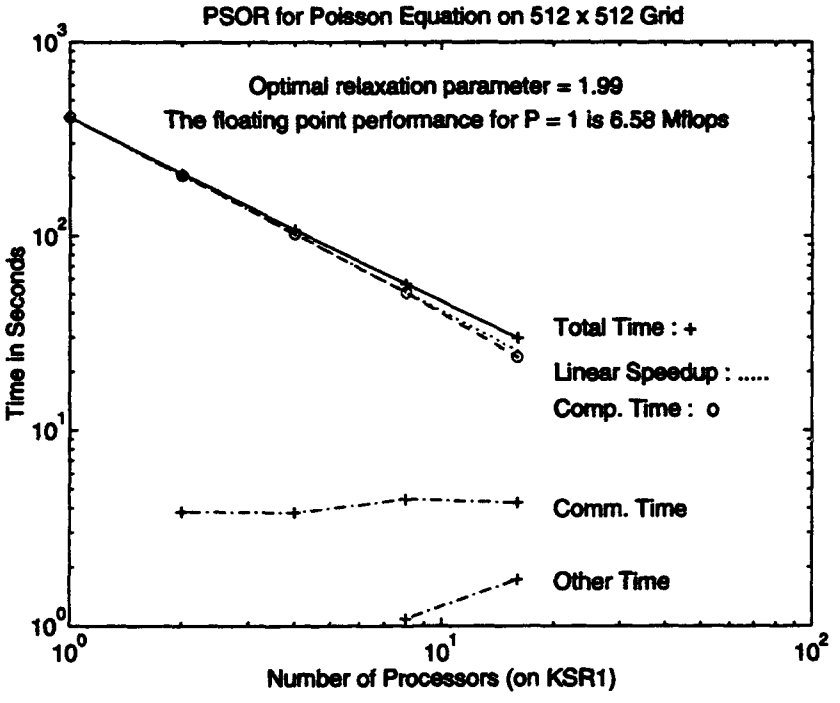


Figure 2.1. Performance analysis for PSOR for the 5-point discretization of Laplace's equation using a strip decomposition on the KSR-1.

We note that the code for this test was implemented in IPfortran and compiled separately for the Delta and KSR-1, without change of source code. The resulting speedup is almost identical for

both systems. In fact, the computation and communications times are largely the same for both systems. Although it is certainly possible to optimize performance for these distinct architectures, this shows that a single programming paradigm can provide efficient execution across a variety of different parallel architectures.

III. PICCG

Our parallel variants of ICCG have been implemented as part of the code UHBD [5,6] which was developed to study the interaction of two molecules of biological significance. One phase involves computing the electrostatic potential around the dominant molecule, and the second phase simulates Brownian motion of the second molecule in this electrostatic force field. The first phase solves the nonlinear Poisson-Boltzmann (NLPB) equation for the electrostatic potential.

We have modified the electrostatic solver to be able to model semiconductor devices [3]. This has provided a stronger test both of the linear and nonlinear parts of the solver, but the principal conclusion is that semiconductor devices can be modeled quite effectively on massively parallel computers. For example, the following table shows that the solver is scalable in the sense that larger problems can be solved without increasing the execution time, by increasing the number of processors used.

Total CPU time in seconds for a MOSFET simulation
on the Intel Delta for P nodes and mesh of size N^3

N^3	$P=1$	2	4	8	16	32	64	128	256
30^3	*22*	14	9	6	4	4	4	5	8
60^3	**192**	99	52	35	*20*	15	12	13	16
90^3			**214**	94	62	36	28	*26*	28
140^3					**184**	127	90	91	76
200^3							**252**	228	197
260^3									**440**

One particular case of interest is the so-called *memory constrained* scaling, the times for which are indicated in bold face. This is the case using the smallest number of processors which can run the problem, i.e., can fit the problem in local memory. We note nearly constant run times for this case. The slanted numbers indicate a different scaling which corresponds to a number of processors yielding an execution time that is an order of magnitude smaller. In this case, local memory is not utilized fully.

Most importantly, this table indicates that very large problems can be solved in just a few minutes (or just a few seconds, depending on resources available), allowing repeated designs to be tested or even optimized. We note also that the best decomposition has not been used for the case of large P and moderate N. If a block decomposition were used in this case, even better performance would be realized for the times away from the diagonal in the table.

One striking conclusion of our work so far [2, 3, 4] is that the total execution time for the elliptic solver portion of UHBD is essentially the same for quite disparate computer architectures and programming paradigms, as shown in Figure 3.1. The computations on each machine have quite distinct internal characteristics. For example, each calculation in done in each machine's single precision, which is 8-bytes on the KSR-1 and 4-bytes on the Delta. Due to the shorter word length, more iterations actually are done to reach the prescribed tolerance (the same for both machines).

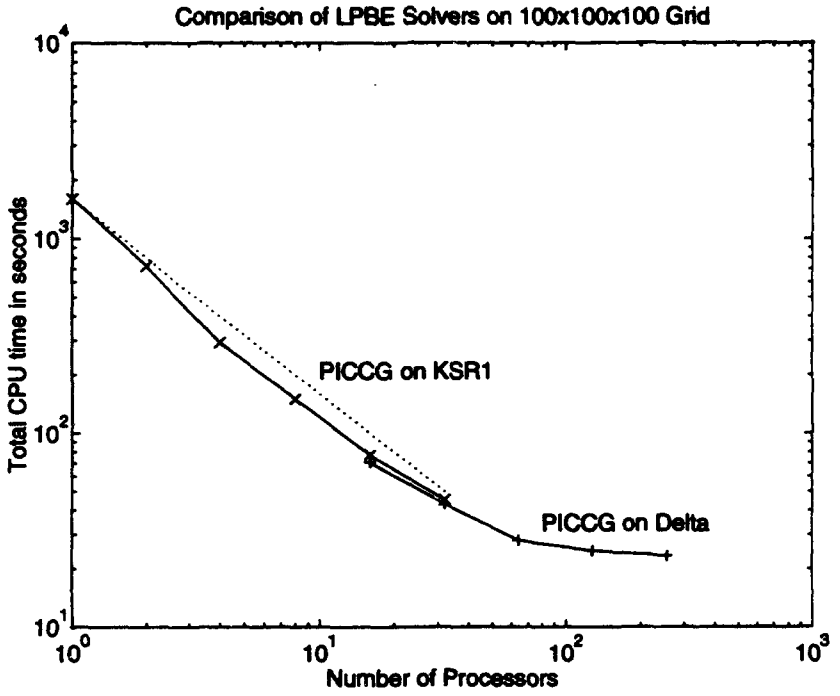


Figure 3.1. Timing for the linear and nonlinear solvers in UHBD on a test problem with a single atom.

In addition, quite different programming paradigms are being used in each case. For the Delta computations, we used IPfortran [1], an explicitly parallel language. For the KSR computations, we used the KSR "tiling" directives [10]. However, the total time is almost identical for 16 and 32 processors for a uniform mesh of size 100^3.

IV. Conclusions and future work

We view the current state of affairs in our work as incomplete. We have identified a number of promising parallel iterative methods, but we have not yet begun to quantify their domains of applicability (and superiority). Moreover, we anticipate these will ultimately find their best application as coarse grid solvers in a parallel multigrid technique.

On the other hand, just using these parallel variants of standard iterative methods, we are able to solve two and three dimensional problems of substantial industrial interest remarkably quickly. For example, the simulation shown in Figure 3.1 solves a three dimensional problem with a million

unknowns in less than a minute using 32 processors. For this reason, it seems appropriate to consider more accurate models of semiconductors, e.g., the Boltzmann equation [7], together with methods for accelerating such calculations using a diffusion approximation [13]. The understanding of such methods in the context of neutral particles (photons, neutrons, etc.) has advanced dramatically recently [11]. However, application of these ideas to electron transport is still in a formative stage. We hope to address this at a later date.

References

[1] Bagheri, B., Clark, T. W., and Scott, L. R. IPfortran: a parallel dialect of fortran. *Fortran Forum 11* (Sept. 1992), 20–31.

[2] Bagheri, B., Ilin, A., and Scott, L. R. Parallelizing UHBD. Research Report UH/MD 167, Dept. Math., Univ. Houston, 1993.

[3] Bagheri, B., Ilin, A., and Scott, L. R. Parallel 3-D MOSFET simulation. In *Proceedings of the 27th Annual Hawaii International Conference on System Sciences* vol. 1, T.N.Mudge and B.D. Shriver, ed's, IEEE Computer Soc. Press, 1994, pp. 46–54.

[4] Bagheri, B., Ilin, A., and Scott, L. R. A Comparison of Distributed and Shared Memory Scalable Architectures. 1. KSR Shared Memory, accepted for the Scalable High Performance Computing Conference, May, 1994.

[5] Davis, M. E., Luty, J. D., Allision, B. A., and McCammon, J. A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program. *Computer Physics Communications 62* (1990), 187–197.

[6] Davis, M. E., and McCammon, J. A. Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradient methods. *Journal of Computational Chemistry 10* (1989), 386–391.

[7] P. Degond and F.J. Mustieles, A deterministic partilce method for the kinetic model of semiconductors: the homogeneous field model, *Solid-State Electronics* **34** (1991), 1334-1345.

[8] Gilson, M. K., Straatsma, T. P., and McCammon, J. A. Open 'back door' in a molecular dynamics simulation of acetylcholinesterase. *Science March 4* (1994), 386–391.

[9] Holst, M., and Saied, F. Multigrid solution of the Poisson-Boltzmann equation. *Journal of Computational Chemistry 14* (1993), 105–113.

[10] Kendall Square Research Corporation. *KSR Fortran Programming.* Kendall Square Research, Waltham, MA, 1992.

[11] K.M. Khattab and E.W. Larsen, Synthetic acceleration methods for linear transport problems with highly anisotropic scattering, *Nuclear Sci. & Eng.* **107** (1991), 217–227.

[12] Scott, L. R. Elliptic preconditioners using fast summation techniques, Proc. Domain Decomposition 7, Penn State, October, 1993. To appear in Contemporary Mathematics, Americam Math. Soc., Providence.

[13] Scott, L. R. Computer design: a new grand challenge, In *Proceedings of the 27th Annual Hawaii International Conference on System Sciences* vol. 1, T.N.Mudge and B.D. Shriver, ed's, IEEE Computer Soc. Press, 1994, pp. 3–6.

[14] Xie, D. Research Report UH/MD, Dept. Math., Univ. Houston, to appear.

PARALLEL IMPLEMENTATION OF A GAAS MESFET ELECTRO-THERMAL SIMULATION ON A TRANSPUTER-BASED SYSTEM

C.S. Tsang-Ping, D.M. Barry, C.M. Snowden
Department of Electronic and Electrical Engineering, Leeds University, Leeds, LS2 9JT, UK

ABSTRACT

The numerical simulation of a GaAs MESFET device on a transputer-based parallel system is presented. The physical modelling consists of a comprehensive two-dimensional energy transport model taking into account thermal heating effects within the device lattice. The semiconductor equations were solved by an SOR point iterative method using a finite difference discretisation scheme. Algorithms targeted at message passing Multiple-Instructions Multiple-Data (MIMD) distributed memory architectures are described. The efficiency and stability of the parallel algorithms are briefly discussed. A parallel speed-up of 14.3 was obtained on an array of 16 transputers.

1. INTRODUCTION

Semiconductor simulation is a very important tool in the design and understanding of new semiconductor devices. However, the use of accurate physical models usually requires expensive high performance computing resources. The recent advances in parallel processing technology offer a cheap and scalable alternative computing solution. However, parallel algorithms are still needed to complement these systems. This paper presents the parallel numerical simulation of the characterisation of an n-channel MESFET semiconductor device using a time-dependent SOR iterative method. Parallel algorithms for this type of iterative solver were designed specifically for a message-passing MIMD distributed memory transputer architecture. The simulation used a comprehensive electro-thermal MESFET model and the numerical solution was achieved with the finite difference discretisation scheme.

2. THE GAAS MESFET ELECTRO-THERMAL MODEL

The electro-thermal model used in this work is based on an *Energy Transport* model [1] coupled with a *Thermal* model [2]. This comprehensive model accounts for hot carriers effects and the influence of lattice heating on the electron flow not included in the standard *Drift-Diffusion Transport* model. A *Scharfetter-Gummel* formulation for current density and energy flux was used. The governing equations for the MESFET device are as follows:

Poisson
$$\nabla^2\psi = -\frac{q}{\varepsilon_o\varepsilon_r}(N_D - n) \tag{1}$$

Current Continuity
$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla\cdot J_n + G \tag{2}$$

Current Density
$$J_n = q\mu_n(w_n)nE + qD_n(w_n)\nabla n \tag{3}$$

Energy Density Conservation
$$\frac{\partial W_n}{\partial n} = J_n\cdot E - \nabla\cdot S_n - \frac{W_n - W_{no}}{\tau_{w_n}(w_n)} \tag{4}$$

Energy Flux
$$S_n = -\mu_n(w_n)nE + qD_n(w_n)\nabla W_n \tag{5}$$

Thermal
$$\nabla\cdot(k_L\nabla T_L) + H_S = 0 \tag{6}$$

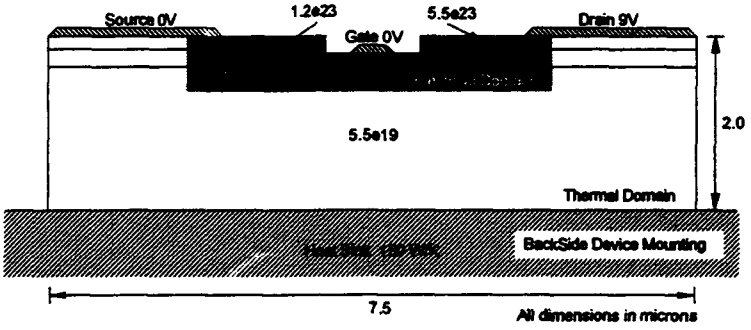
Heat Generation
$$H_s = q\frac{W_n - W_{no}}{\tau_{w_n}(w_n)} \tag{7}$$

where $\psi, N_D, n, t, J_n, G, \mu_n, E, D_n$ are, respectively, the potential, donor density, electron concentration, time, current density, recombination, mobility, electric field, diffusion coefficient, $W_n, S_n, W_{no}, \tau_{w_n}$ are, electron energy flux, energy density, equilibrium energy flux, energy relaxation time and k_L, T_L, H_s are lattice thermal conductivity, lattice temperature, heat generation respectively.

3. TIME-DEPENDENT NUMERICAL ALGORITHM

The numerical solution is achieved using a finite difference discretisation scheme. The semiconductor equations consisting of the closely coupled Poisson, continuity and energy equations are solved by a Gauss-Seidel point iteration method with successive relaxation for a time-dependent solution [3].

The thermal model couples with the carrier transport model to form a coupled electro-thermal model. The solution of this model is obtained by sequentially solving the Poisson (1), current continuity (2) and energy density conservation (4) partial differential equations once per time-step. This time-stepping process is repeated until a steady state solution is reached. The elliptic thermal equation (6) is also solved with the SOR technique but for steady state conditions as the lattice temperature usually remains constant at the DC bias condition [2]. The lattice temperature is therefore solved at regular intervals in the time stepping process, i.e. for a total simulation time of 2.5ps, for time-steps of 5fs, the thermal equation is solved every 0.25ps and the new lattice temperature obtained is subsequently used.



Low Field Mobility = 0.8 m^2Vs^{-1} Gate Width = 400e-6 m

Figure 1. Generic Power GaAs MESFET Device

The thermal boundary has been restricted to the same domain as the carrier transport equations, therefore equivalent third-order boundary conditions are required. To obtain accurate results, the domain for analysis has to be extended horizontally for 2 to 3 times the source-drain spacing and to a depth of 5 to 10 times the active region of the device [2]. Instead of solving the equation over the electrical domain (i.e. 3.5 by 0.5 microns - greyed) a much wider thermal domain (i.e. 7.5 by 2.0 microns - white) has to be solved as depicted in Figure 1. As a result, more computational resources both in memory and processor time are needed to solve the device problem This is naturally suited to a scalable parallel system.

4. PARALLEL ALGORITHMS & IMPLEMENTATION

The parallel system used consists of an array of TRAnsputer Modules (TRAMs) connected in a ring network topology. Each TRAM has a transputer and some local memory. The device simulation problem has been parallelised using a one dimensional geometric domain decomposition [4] as shown in Figure 2. A sub-domain of the whole MESFET is locally stored on each TRAM's memory and each transputer computes the solution of its sub-domain concurrently.

The finite difference discretisation of the semiconductor equations results in a five-point discretisation scheme. This requires some communication of boundary data between neighbouring processes. The efficient design of communication protocols is very important in a parallel environment. The aim is to minimise communication overheads.

The two main overheads are attributed to the *control* and *data exchange* protocol. The *control* protocol is based on a driver-slave principle. The driver sends messages via the control channel to the slaves to perform specific tasks. The iterative nature of the numerical algorithms requires messages from all the slaves to the driver so that the latter can test for the convergence of solution across the whole domain. The *data exchange* communication protocol ensures data consistency across boundaries between neighbouring processes. The algorithm communicates data via the exchange channels (refer to Figure 2).

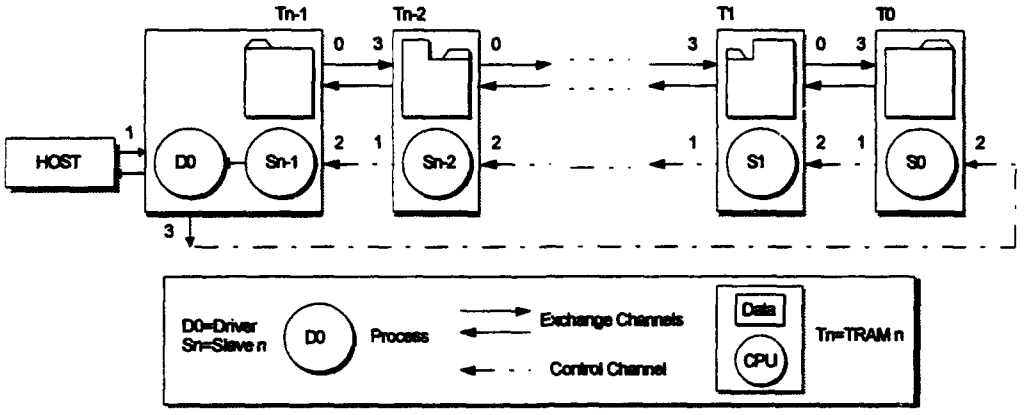


Figure 2. Configuration and Mapping of Simulation on an Array of TRAMs (n=16)

The convergence of the iterative solution requires a modified red/black checker-board one-dimensional partitioning method named as the *RB 1D SOR* [5]. This parallel ordering method is essential for the optimum convergence of the point iterative scheme.

5. RESULTS AND PERFORMANCE

Simulation and system performance results were obtained from the parallel system by simulating the recessed-gate GaAs MESFET depicted in Figure 1 using a uniform grid for the bias condition shown. Figure 3 shows the electron and lattice temperature profiles of the GaAs MESFET device.

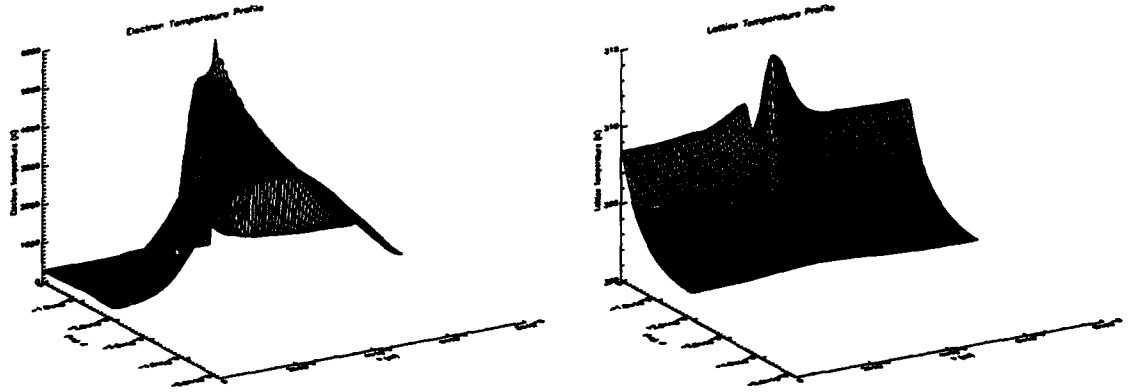


Figure 3. Electron and Lattice Temperature of GaAs MESFET device

Although the parallel simulator was shown to performed adaptive meshing efficiently[5] for an electrical model, the electro-thermal simulation has been restricted to a uniform grid of 96x60 owing to memory limitations on the TRAMs. Table 1 shows the performance of the parallel system, solving the full electro-thermal equations iteratively using the *RB 1D SOR* partitioning method, with the solution computed over a time of 2.5ps (500 time steps of 5fs) for steady state solution.

The numerical solution for this particular problem converges in roughly the same number of iterations when simulated on the parallel system. Using a single transputer, it took about 6.5 hours to compute that solution but only about 0.5 hour with 16 transputers.

Nodes	Iteration Number				Time
	Poisson	Current	Energy	Thermal	(hr)
1	21380	2718	2061	491	6h36
4	21351	2718	2062	492	1h43
8	21324	2659	2042	492	0h53
12	21295	2716	2058	492	0h36
16	21256	2709	2060	491	0h28

Table 1. Performance of the Parallel Simulation

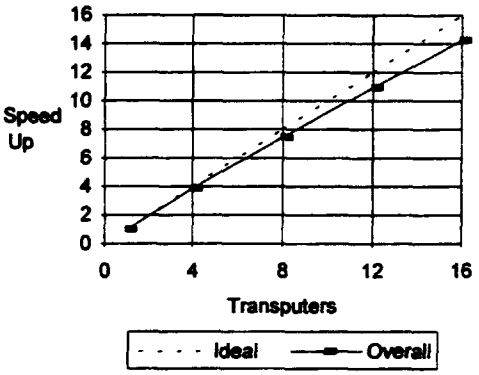


Figure 4. Overall Parallel Performance

As shown in Figure 4 a speed increase of up to a factor of 14.3 on 16 transputers is possible with the current algorithm. The drop in performance for large number of transputers is attributed to the increase in the communication to computation ratio due to a fix domain size. Increasing the size of the problem as the number of processors is usually recommended for efficient use of parallelism. Otherwise most time is spent on communication of data rather than useful computation as is the case for very small domains.

6. CONCLUSION

A complete parallel numerical simulator for the characterisation of a GaAs MESFET device has been presented. Geometric domain decomposition was the natural way of parallelising the problem enabling a logical map on a distributed memory system. In addition to the numerical algorithms, efficient parallel algorithms are required to manage the communications protocols in this distributed-data environment. We found that careful implementation of the communication protocols is important to achieve high parallel efficiencies. A modified red/black partitioning updating method is also required to provide optimum convergence. The Gauss-Seidel point iterative method with successive relaxation was very suitable for parallel implementation. An advantage of this method is that the inclusion of additional equations such as the energy density conservation and thermal equations follow the same parallel principle that is used for Poisson and current continuity equations. The parallel methods developed and implemented in this work proved that parallel processing is a feasible computing alternative that can be used to provide fast characterisation of semiconductor devices.

REFERENCES

[1] C.M. Snowden, D. Loret, "Two-Dimensional Hot-Electron Models for Short-Gate-Length GaAs MESFETs," *IEEE Transactions on Electron Devices*, Vol. ED-34, No.2, February 1987, pp.212-223.

[2] G. Ghione, P. Golzio, C. Naldi, "Self-Consistent Thermal Modelling of GaAs MESFET's: A Comparative Analysis of Power Device Mountings," *Alta Frequenza*, Vol. LVII, No. 7, 1988, pp.311-319

[3] T.M. Barton, *Characterisation of the physical behaviour of GaAs MESFETs*, PhD Thesis, Department of Electrical and Electronic Engineering, Leeds University, UK, 1988.

[4] C.G. Fox, M. Johnson, G. Lyzenga, S. Otto, J. Salmon, D.Walker, *Solving problems on concurrent processors: General techniques and regular problems*, Prentice-Hall, 1988.

[5] C.S. Tsang-Ping, D.M. Barry, C.M. Snowden, "Parallel Simulation of a GaAs MESFET Semiconductor Device with Adaptive Meshing on a Transputer Network," *Proceedings of the 2nd Euromicro Workshop on Parallel and Distributed Processing*, Malaga, Spain, IEEE Computer Society Press, January 1994

A PARALLEL MULTIGRID SOLVER FOR SEMICONDUCTOR DEVICE EQUATIONS

X.Han, D.M.Barry, and M.J.Howes
Department of Electronic and Electrical Engineering, University of Leeds, LS2 9JT, UK
Email: een5xh@leeds.ac.uk, dmb and mjh@elec-eng.leeds.ac.uk

Abstract

This paper presents a parallel multigrid algorithm for solving semiconductor device equations in two-dimensions and its implementation on a MIMD parallel machine with distributed memory. The numerical experiments of a GaAs MESFET device simulation demonstrate the combined high efficiency of the domain decomposition method and the multigrid method. The parallel multigrid method using 16 processors is up to 60 times faster than a single-grid iterative solver using a processor.

I. INTRODUCTION

Semiconductor device simulations are an important tool for both physicists and device design engineers to analyse physical phenomena inside semiconductor devices and to predict the performance of new devices prior to fabrication. These physical models require the solutions of the Poisson equation for electrical potential and the current continuity equations for electron and hole concentrations. More complex models will also solve for the energy and temperature distributions. Numerical techniques must be used to solve these equations which require extensive computing power. The increasing complexity of both devices and physical models have challenged the available computing resources.

Many attempts have been made to design fast simulations. Multigrid (MG) methods offer a fast and robust iterative method for solving PDEs and have found applications in semiconductor device simulation [1-3]. The simulation of semiconductor devices possesses an inherent parallelism which requires many repeated operations on different grid points. It is clear that by exploiting the parallelism of the numerical algorithms leads to good speed ups of the simulation. The direct solution method has been used on MIMD parallel computers for device simulation [4], however this does not show good efficiency in the case of normal grid sizes. The Jocobi-SOR, Frankel iterative methods[5] and conjugate gradient iterative methods[6] have been implemented on SIMD Connection Machines. This approach is limited when modelling the irregular structures of modern devices. Transputer networks have also been used in semiconductor device simulation using a Monte Carlo method in [7] and finite difference methods in [8].

In this paper we present for the first time a parallel implementation of the multigrid iterative method for the solution of two-dimensional device equations on a medium-grain MIMD parallel machine with distributed memory and demonstrate its parallel efficiency for device simulation.

II. SEMICONDUCTOR DEVICE EQUATIONS

The semiconductor model used is based on the drift-diffusion approximation. It includes the Poisson equation, the current continuity equations for electrons and holes. The coupled system consists of an elliptic differential equation and two parabolic partial differential equations with dependent variables potential ψ, electron concentration n and hole concentration p. After discretisation using a finite difference scheme on a rectangular grid with N grid points, together with boundary conditions, we obtain a set of

$3\times N$ equations with $3\times N$ variables ψ, n, and p. We can write them in symbolic form as

$$F(\boldsymbol{u}) = f \qquad (1)$$

where F denotes the nonlinear difference operators, f a constant and $\boldsymbol{u}$ is a matrix including ψ, n, and p. The Gauss-Seidel (GS) technique coupled with successive over-relaxation (SOR), GS-SOR, is used for the Poisson equation along with successive under-relaxation (SUR), GS-SUR, for the continuity equations. The coupled Poisson and the continuity equations are solved using Gummel's approach[9].

III. Multigrid Method

It is obvious that PDEs need to be solved on a fine-enough grid to obtain accurate solutions. In semiconductor simulation a large truncation error may cause convergence problems. In some circumstances the larger the grid spaces the smaller the time step is required to maintain solution stability [10]. Classical iterative methods slow down with increasing grid point number. Multigrid iterative methods are highly efficient solvers for PDEs, in which the combination of fine grid relaxation and coarse grid correction produces a fast convergence rate. The multigrid full approximation scheme (MG-FAS) [11] is used in this work. To solve the system $F(\mathbf{u})=\mathbf{f}$ the MG method uses a sequence of grids, $G_k(1\le k\le K)$, where G_1 is the finest grid and G_K is the coarsest grid. There exists a system $F_k(\mathbf{u}_k)=\mathbf{f}_k$ on each grid G_k. P^k_{k+1} is a prolongation operator from a coarse grid G_{k+1} to a fine grid G_k and R^k_{k-1} is a restriction operator from a fine grid G_{k-1} to a coarse grid G_k. MG-FAS cycles may be defined as follows.

(a) *Interpolate u_1 and f_1 to each of the grids and solve $F_k(u_k)=F_k(R^k_{k-1}u_{k-1}) + R^k_{k-1}(f_{k-1} - F_{k-1}u_{k-1})$*

(b) *IF G_k is the coarsest grid ($k=K$)*

solve $F_K(U_K)=F_K$ exactly

Prolongation correction $V_K = | u_K^{new} - u_K^{old} |$ to the next fine grid G_{K-1}

ELSE

Solve $F_k(u_k)=f_k+P^{k+1}{}_kV_{k+1}$

Prolongation correction $V_k = | u_k^{new} - u_k^{old} |$ to the next fine grid G_{k-1}

(c) *Do (b) until to the finest grid G_1*

The Gauss-Seidel method is used as smoothing operator on all grids except on the coarsest grid where the GS-SOR/SUR method is employed to speed-up the solution. The interpolation operators are bi-linear prolongation and half-weighting restriction.

IV. Parallelisation and Implementation

There are two methods to exploit parallelism in multigrid methods [12]. One is straightforward and based on the domain decomposition technique. In this method the multigrid is divided into several smaller sub-grids, where each sub-grid includes all levels from the finest to the coarsest. These sub-grids are then distributed onto several processors and all processors do the same multigrid operations on different sub-grids in parallel. This method is referred to as data-parallel multigrid (DPMG). The other method is to carry out the multigrid operations concurrently by many processors on different grid levels, in which the multigrid is not partitioned but each level of the multigrid is assigned to a processor and several MG operations on different grid levels are done at the same time. The second method may be called operation-parallel multigrid (OPMG). The DPMG method is used in this work.

Fig.1 shows a 3 level multigrid decomposition and sub-domain mapping onto 3 processors. All operations mentioned above, solving, smoothing, computing errors, and interpolations between different level, are local to each processor. At inter-sub-domain boundaries relaxations and interpolations in a sub-domain will

need the boundary data of the adjoining sub-domains. Column dummy points at each side of sub-domains along the boundaries are allocated and communications are required at all grid levels to update the values of the dummy points.

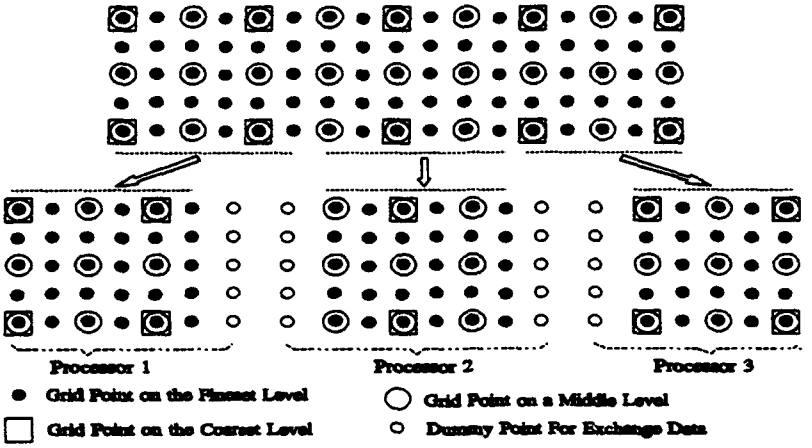


Fig.1 Partitioning and Distributing the Multigrid onto 3 Processors

Thus all levels of the multigrid are partitioned and sub-grids are mapped onto the processors in a way that the locality of the interprocessor communication profits from the locality of the discretisation grid to reduce the communication overheads. A transputer-based machine hosted by a workstation is used in this implementation. The machine includes 16 processors with 2 Mbytes of memory each. These are connected as a ring. The parallel version of multigrid operations includes the following steps,

(a) *Solving equations (or smoothing errors) on a grid level and updating boundary points of the grid level after every iteration.*
(b) *Calculating the errors on a grid level.*
(c) *Prolongation from a coarse grid level to the next fine grid level.*
(d) *Updating boundary points on the fine grid level.*
(e) *Restriction from a fine grid level to the next coarse grid level.*
(f) *Updating boundary points on the coarse grid level.*

It is obvious that the amount of communication is proportional to the number of grid points on the side boundaries of sub-grids for a ring of processors. The amount of the computation, including all multigrid operations, is proportional to the number of grid points in the sub-domains. The parallel overhead is the ratio of the amount of the communication and the amount of the computation. So the overhead becomes larger as the grid becomes coarser.

V. Experimental Results

A GaAs MESFET device with 1 micron channel length has been used in the current work. For a GaAs MESFET device, which is entirely unipolar, only the Poisson equation and the continuity equation for the majority carrier of electrons need to be solved. The bias voltages applied to the contact source and the contact gate are 0V with a built-in voltage of - 0.8V on the gate. The external drain voltage is increased linearly from 0V to 5V over 1 ps and then fixed at 5V. Comparisons are made between the GS-SOR/SUR methods and the MG-FAS method both on a single processor and 16 processors. The simulation time required using the two methods are listed in Table 1. The speedup factors due to the multigrid method and due to the data-parallel algorithm are calculated from the execution time.

It can be seen from the table that the multigrid method is implemented on the transputer-based system without significantly impairing the efficiency of the multigrid. The parallel multigrid still produces a typical parallel speedup of domain decomposition techniques. Since only a few grid points are left on a very coarse grid, communication overhead becomes more significant; the speedup of the parallel multigrid method drops from 7.10 to 4.62. The speedup due only to the data-parallel method also decreases from 13.1 to 8.51 for the same reason. The speedup of both sequential and parallel multigrid is affected by the fact that the solution of a time-step can be started from the solution of the previous time-step as a very good initial guess and so the initial error is normally smaller compared with general initial problem. Even so the overall speedup from the combination of the multigrid method and the domain decomposition method is still over 60.

VI. CONCLUSIONS

Table 1 Execution Time(in seconds) and Speedup(Sp)

Processor No.	GS-SOR/SUR	MG-FAS	MG-Sp
1	4087	575.6	7.10
16	312.1	67.60	4.62
Data-Para. Sp	13.10	8.51	60.46†

† Overall Speedup

It has been demonstrated that it is possible to obtain fast solutions for the coupled semiconductor equations using the multigrid method on a medium-grain parallel machine with distributed memory. Both the parallel efficiency of the underlying data-parallel algorithm and the convergence rate of multigrid method are maintained. By using the parallel multigrid method on a 16 processor machine the simulation time is reduced by more than 60 times compared with the single grid sequential solver with the Gauss-Seidel method coupled with successive over/under-relaxation scheme. The parallel efficiency is expected to improve with an increase in the problem size.

REFERENCES

[1] A S Shieh, On the Solution of Coupled System of PDE by a Multigrid Method, *IEEE Trans. on Electron Devices*, Vol.ED-32, No.10, pp.2083-2086, 1985

[2] P W Hemker, A Nonlinear Multigrid Methods for One-Dimensional Semiconductor-Device Simulation - results for the Diode, *J. of Comput. and Appl. Math.*, Vol.30, No.1, pp.117-126, 1990

[3] P M de Zeeuw, Nonlinear Multigrid Applied to a One-Dimensional Stationary Semiconductor Model, *SIAM J. on Sci. and Stat. Comput.*, Vol.13, No.2, pp.512-530, 1992

[4] K Motegi and S Watanabe, A Parallel Algorithm for Solving Two Dimensional Device Simulation by Direct Solution Method and Its Evaluation on the AP-1000, *IEICE Trans. Fundamentals*, Vol.E75-A, No.7, pp.920-922, 1992

[5] J P Darling and I D Mayergoyz, Data Parallel Algorithms for the Numerical Modeling of Semiconductor Devices, *Proc. of the 5th SIAM Conf. on Parallel Processing for Sci. Comput.*, Houston, TX D910325-27, CH.4, pp.394-400, 1992

[6] D M Webber, E Tomacruz, R Guerrieri, T Toyabe, and A Sangiovanni-Vincentelli, A Massively Parallel Algorithm for Three-Dimensional Device Simulation, *IEEE Trans. on CAD*, vol.10, No.9, pp.1201-1209, 1991

[7] B Boittiaux, G Goncalves, and M P Haye, A Transputer Network for a Device Simulator, *Microprocessing and Microprogramming*, Vol.35, pp.127-132, 1992

[8] C S Tsang-Ping, D M Barry, and C M Snowden, Parallel Implementation of A GaAs MESFET Electron-Thermal Simulation on a Transputer-Based System, *In this proceedings.*

[9] H K Gummel, A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations, *IEEE Trans. on Electron Devices*, Vol.ED-11, pp.455-465, 1969

[10] M Reiser, Large-Scale Numerical Simulation in Semiconductor Device Modeling, *Comp. Math. Appl. Mech. Eng.*, Vol.1, pp.17-38, 1972

[11] A Brandt and N Dinar, Multigrid Solution to Elliptic Flow Problems, *Numerical Methods in Partial Differential Equations*, Ed. S Parter, pp.53-147, 1977

[12] R S Tuminaro, Multigrid Algorithms on Parallel Processing Systems, *PhD Thesis, Stanford University*, 1990

AN EFFICIENT FULL-ZONE K·P METHOD FOR 'ON THE FLY' CALCULATION OF VALENCE BAND ENERGIES IN HOLE TRANSPORT STUDIES

John P. Stanley and Neil Goldsman
Department of Electrical Engineering
University of Maryland, College Park, MD, 20742

Abstract

A new computationally efficient full-zone k·p method for use in valence band transport and optical studies of Si and Ge has been developed. This method generalizes the traditional k·p method by avoiding, in part, the use of perturbation theory. New band parameters have also been computed.

I. INTRODUCTION

Full-zone carrier transport investigations are becoming increasingly popular. Band-structures used in these simulations are often calculated with the pseudopotential method, which uses plane waves as basis functions. However, due to the rapidly varying angular behavior exhibited by the true eigenfunctions, a large number of plane waves are needed (typically 50-100). Thus, band structure calculations usually require considerable CPU time, and full-zone transport calculations often resort to interpolation and to look-up tables, where the results of these pseudopotential calculations are stored. Under certain conditions, however, it may be preferable to calculate the band structure values while the transport simulation is taking place ('on the fly'). To allow for these circumstances, we have developed a full-zone k·p method, which very quickly calculates an eigenvalue anywhere in the zone for the relevant valence bands. Determining an eigenvalue requires approximately 5 milliseconds on a 486PC-33, and values agree well with results from pseudopotential calculations. The procedure is also invertible to give $\mathbf{k}(\varepsilon)$, making this method attarctive for Monte Carlo calculations. The calculations have also lead to the development of new band parameters.

II. METHOD OF SOLUTION

The k·p method is based on second-order degenerate (Rayleigh-Schrödinger) perturbation theory applied to a one-electron crystal Schrödinger wave equation [1,2]; given a knowledge of a relatively small number of band energies at a particular point $\mathbf{k}_0$ in k-space the method allows the band structure to be calculated very accurately for small deviations $\delta\mathbf{k}$ about $\mathbf{k}_0$. For sufficiently small $\delta\mathbf{k}$ analytic solutions are possible [1], while for somewhat larger $\delta\mathbf{k}$, a numerical solution is required [2,3]. The k·p method has also been applied extensively to superlattice structures [4,5].

With our k·p calculations, we do not resort to perturbation theory, but solve exactly an 18 band k·p secular determinant with spin-orbit coupling included. Specifically, we have generalized the well-known 3×3 k·p valence band Hamiltonian of Dresselhaus *et al.* (DKK) [1] to include all momentum coupling arising from an underlying 18-band k·p Hamiltonian. This allows direct use of the Kane cubic [2], but now with energy-dependent coefficients. We have computed direct (momentum) coupling matrix elements using the five Luttinger parameters [6,7], and have computed new indirect coupling matrix elements using a fitting procedure.

Once the secular equation is established, we find the eigenvalues using a determinant decomposition method, which to our knowledge has never been applied to band calculations, and is

closely related to "Schur complement domain decomposition" [8]. With this method, the matrix is block-partitioned and then reduced to lower (block)-triangular form. This allows the associated determinant to be factored into a product of lower-order determinants. We then only have to solve the low-order determinant associated with the valence bands of interest. This 3rd order secular determinant is quickly solved with Newton's method to obtain the eigenvalues. The mathematical basis of our method is as follows.

We wish to solve the eigenvalue problem

$$(H-\varepsilon)\,|\phi\rangle = (H_0 + H' - \varepsilon_n)\,|\phi\rangle = 0 \tag{1}$$

where the Hamiltonian has been split into two parts, one of which representing an independently soluble problem

$$(H_0 - \varepsilon_n)\,|\psi_n\rangle = 0;\;\; H_0 = -\nabla^2 + V;\;\; \{\psi_n\}_n \text{ complete on } \mathcal{H} \tag{2}$$

In the present context, H' represents the k·p terms, as well as (formally) spin-orbit coupling. Now, we partition $\mathcal{H}$ as

$$\mathcal{H} = \mathcal{H}_A \oplus \mathcal{H}_B;\;\; \mathcal{H}_A = \text{ span}\{\psi_n\}_{n\in A};\;\; \mathcal{H}_B = \text{ span}\{\psi_n\}_{n\in B} \tag{3}$$

and $|\phi\rangle$, accordingly, as

$$|\phi\rangle = P_A\,|\phi\rangle + (1-P_A)\,|\phi\rangle\,;\;\; P_A = \sum_{n\in A} |\psi_n\rangle\,\langle\psi_n| \tag{4}$$

Substitution of (4) into (1) then leads to the following eigenvalue problem, restricted to $\mathcal{H}_A$

$$\sum_{j\in A}\{\langle\psi_i|P_AH'P_A|\psi_j\rangle + \delta_{ij}\varepsilon_i$$
$$- \langle\psi_i|P_AH'(1-P_A)[(1-P_A)(H-\varepsilon)(1-P_A)]^{-1}(1-P_A)H'P_A|\psi_j\rangle - \delta_{ij}\varepsilon\}\langle\psi_j|\phi\rangle = 0 \tag{5}$$

This result follows in a more straightforward manner directly from (1), if we partition the matrix H in accordance with (3) and employ a determinant reduction formula involving the Schur complement

$$(H-\varepsilon)|\phi\rangle = 0 \rightarrow |H-\varepsilon| = \begin{vmatrix} A-\varepsilon & S \\ S^\dagger & B-\varepsilon \end{vmatrix} = |A - S(B-\varepsilon)^{-1}S^\dagger - \varepsilon| = 0, \tag{6}$$

the final reduced-order determinant being equivalent to (5) with the following identifications

$$\begin{aligned} (A)_{ij} &= \langle\psi_i|P_AH'P_A|\psi_j\rangle; \\ (S)_{ij} &= \langle\psi_i|P_AH'(1-P_A)|\psi_j\rangle; \\ (B-\varepsilon)^{-1}_{ij} &= \langle\psi_i|[(1-P_A)(H-\varepsilon)(1-P_A)]^{-1}|\psi_j\rangle \end{aligned} \tag{7}$$

Standard k·p perturbation theory follows from (6) by treating $(B-\varepsilon)^{-1}$ as a perturbation. We do not do this, but instead analytically reduce (6) to an expression involving matrices of $O(A)$ only. We employ 18 (cubic harmonic) basis functions, partitioning this set in accordance with standard k·p perturbation theory applied to diamond-type semiconductors ($A_{(3\times3)} \sim \Gamma_{25}$ valence bands, $B_{(15\times15)} \sim$ all other bands). Our new reduced-order 3×3 k·p Hamiltonian replaces the DKK valence band Hamiltonian [1], and we then make use of the Kane solution [2], otherwise unmodified.

We have computed the required momentum matrix elements based on the measured Luttinger parameters of Hensel and Suzuki [6] for Ge and Lawaetz [7] for Si, and have generated new matrix elements associated with intra-conduction band coupling.

To compute valence band energies, we solve the reduced secular determinant using Newton-Raphson iteration, obtaining the three (heavy, light, and splitoff band) energies individually. Typically, fewer than 5 iterations are required to compute each energy to better than 6 decimal places for arbitrary (reduced) **k**. Running on a 486PC-33 computer this corresponds to about 20–25 milliseconds per energy determination for random **k** throughout the Brillouin zone, and to about 2–4 milliseconds per energy for closely spaced wavevectors.

III. COMPUTED VALENCE BANDS IN SI AND GE

Dispersion relations for the valence bands in Si computed using our extended k·p method are shown in Fig. 1, along with selected values given by the pseudopotential calculation of Chelikowsky and Cohen [9]. Agreement with the pseudopotential result is quite good throughout the Brillouin zone. It is also evident that for large **k** our bands tend to fit the local pseudopotential result somewhat better that the nonlocal result. The corresponding result for Ge is provided in Fig. 2. Agreement with the pseudopotential result is also good, though somewhat less dramatic than in Si. It is worth noting that each of these plots requires just over one second to compute on a 486PC-33 computer. Each dispersion curve consists of 100 energy evaluations (hence 300 per plot). Energy contours for the heavy and light hole bands in Si are illustrated in Figs. 3 and 4; these are surface projections in the ⟨100⟩ plane.

IV. SUMMARY

We have developed a computationally efficient method for computing valence band energies throughout the Brillouin zone in bulk diamond-type semiconductors. We have applied this method to Si and Ge, obtaining very good agreement with pseudopotential calculations. We have also computed new k·p valence band parameters. Our approach is to solve the underlying secular equation on a lower dimensional subspace exactly, rather than perturbatively as in traditional k·p perturbation theory. Our method is quite general, and in particular, application to both valence and conduction bands of zincblende materials is possible.

Acknowledgements

This work has been completed under support of a National Needs Graduate Fellowship, and partial support from the Semiconductor Research Corporation.

[1] G. Dresselhaus, A. F. Kip, and C. Kittel, Phys. Rev. **98**, 368 (1955)
[2] E. O. Kane, J. Phys. Chem. Solids **1**, 82 (1956)
[3] J. M. Hinckley and J. Singh, Phys. Rev. B **41**, 2912 (1990)
[4] G. Bastard, in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1991), Vol. 44, p. 229
[5] D. L. Smith and C. Mailhiot, Rev. Mod. Phys. **62**, 173 (1990)
[6] J. C. Hensel and K. Suzuki, Phys. Rev. B **9**, 4219 (1974)
[7] P. Lawaetz, Phys. Rev. B **4**, 3460 (1971)
[8] T. F. Chan and D. Goovaerts, SIAM J. Matrix Anal. Appl. **13**, 663 (1992)
[9] J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B **14**, 556 (1976)

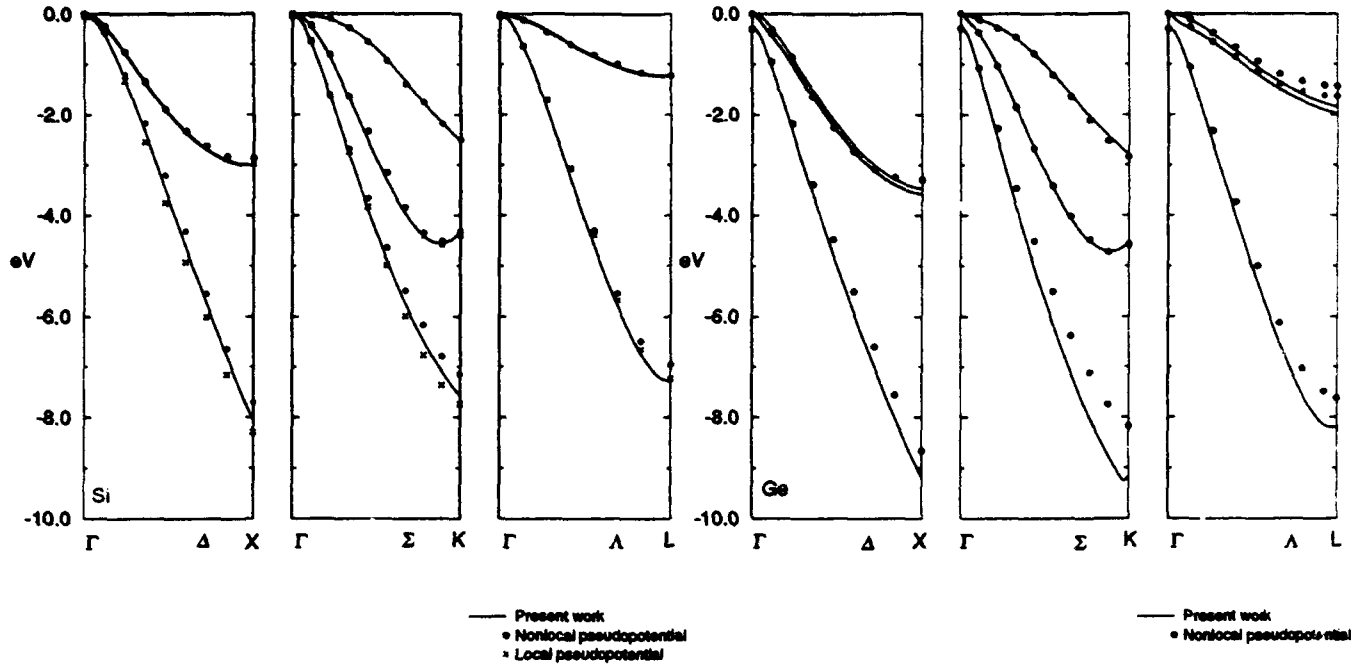


Figure 1: Comparison of present full-zone k·p valence band dispersion relations in Si with the pseudopotential result of Chelikowsky and Cohen [9].

Figure 2: Comparison of present full-zone k·p valence band dispersion relations in Ge with the pseudopotential result of Chelikowsky and Cohen [9].

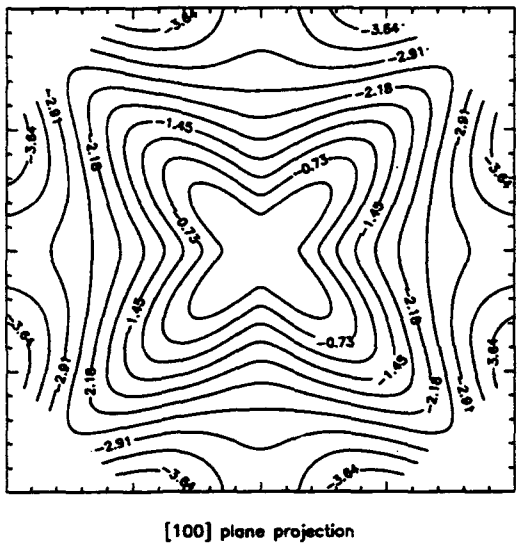


Figure 3: Constant energy surfaces of the heavy hole band in Si.

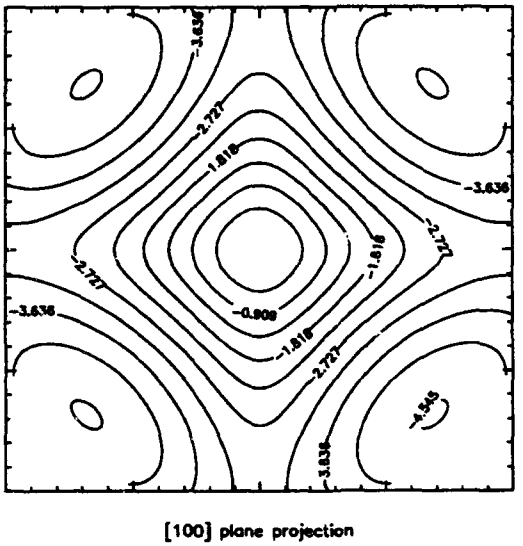


Figure 4: Constant energy surfaces of the light hole band in Si.

A GENERALIZED MONTE CARLO APPROACH FOR THE SIMULATION OF THE COHERENT ULTRAFAST DYNAMICS IN PHOTOEXCITED SEMICONDUCTORS

Fausto Rossi and Stefan Haas
Philipps-Universität Marburg
Fachbereich Physik und Zentrum für Materialwissenschaften
Renthof 5, D-35032 Marburg, Germany
Tilmann Kuhn
Inst. für Theoretische Physik der Universität Stuttgart
D-70550 Stuttgart 80, Germany

Abstract

In this paper a generalized Monte Carlo method recently developed by the authors for the solution of the coupled set of quantum kinetic equations for the distribution functions and the interband polarization is presented. The aim of this method is to combine the advantages of the description within a fully quantum mechanical picture with the power of the Monte Carlo technique for the treatment of stochastic processes. It is based on a decomposition of the kinetic equations in a coherent and an incoherent part. The former is integrated directly while the latter is sampled by means of a Monte Carlo simulation. This allows us to treat on the same kinetic level carrier thermalization and relaxation as well as dephasing processes.

I. INTRODUCTION

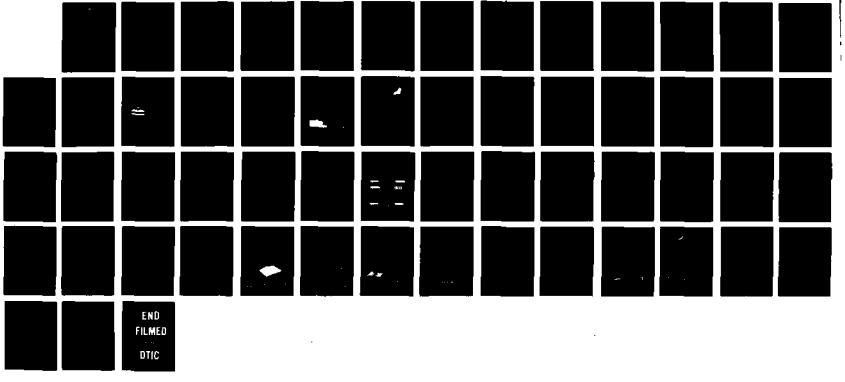
The Monte Carlo (MC) method, which has been applied for more than 25 years to the analysis of semiclassical charge transport in semiconductors, is the most powerful numerical tool for microelectronic device simulation [1]. On the other hand, the present-day technology allows the investigation of relaxation and dephasing phenomena in semiconductors with a time resolution which has now reached a few femtoseconds [2]. On such a time-scale, coherent aspects play an important role and the carrier dynamics cannot be treated in terms of the traditional semiclassical transport theory. Therefore, in order to study this partially coherent dynamics, a generalization of the conventional MC method is required.

The aim of the present invited paper is to review a method recently proposed by the authors [3] and to discuss its application to the analysis of ultrafast carrier dynamics in photoexcited semiconductors [4, 5]. The main peculiarity of the method is to retain the big advantages of the MC method in treating scattering processes and, at the same time, to take into account on the same kinetic level also coherent phenomena. Compared to the conventional MC technique, which simply provides a solution of the semiclassical Boltzmann Transport Equation (BTE), this generalized MC approach provides a solution of the Semiconductor Bloch Equations (SBE). In addition to a simulation of the various distribution functions, this will result in a simulation of the interband polarization induced by the coherent light field.

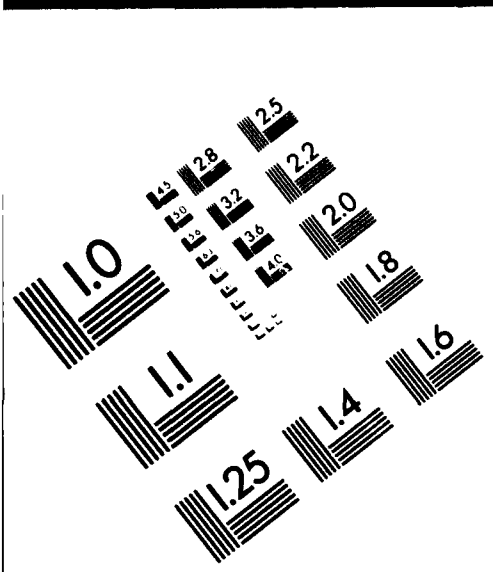
Such an approach allows a selfconsistent description of the carrier photogeneration process [5]. The energy broadening due to the finite pulse duration and due to the decay of the interband polarization has not to be introduced as a phenomenological parameter as in any conventional MC simulation [6] but it comes out selfconsistently with its full time dependence.

AD-A284 214 PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON COMPUTATIONAL ELECTRONICS (3RD) HELD IN PORTLAND OREGON ON MAY 18-20 1994(U) NATIONAL SCIENCE FOUNDATION WASHINGTON DC 20 MAY 94 XJ-XD

UNCLASSIFIED NL



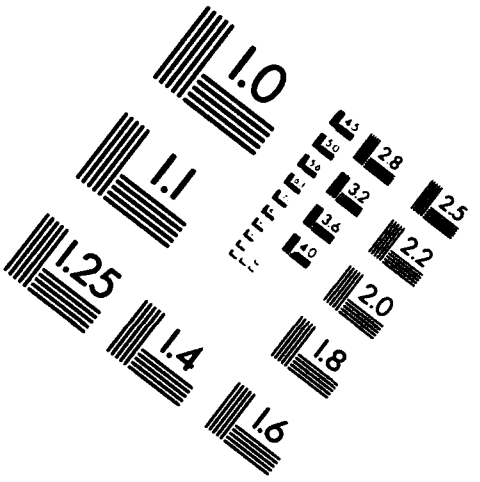
END
FILMED
DTIC

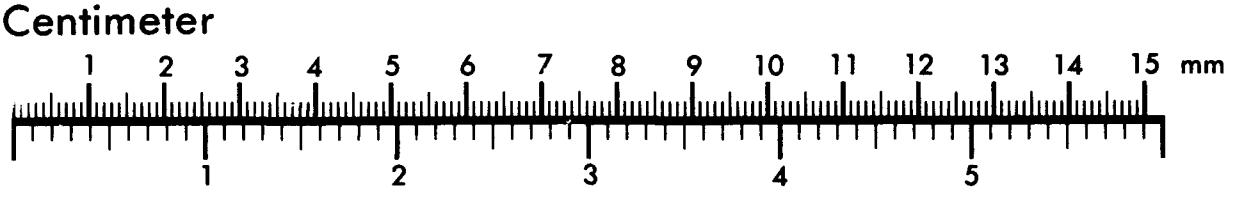


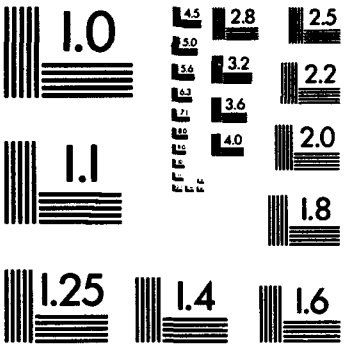
Association for Information and Image Management

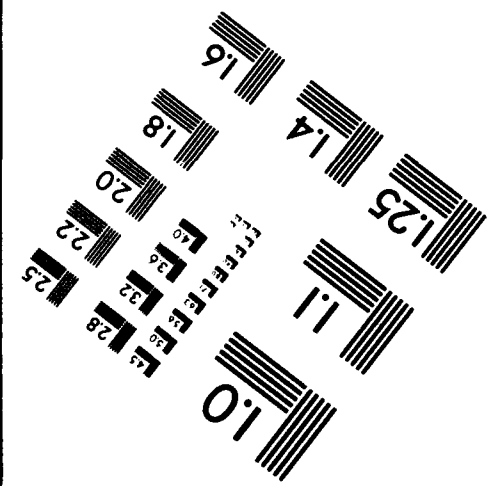
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910

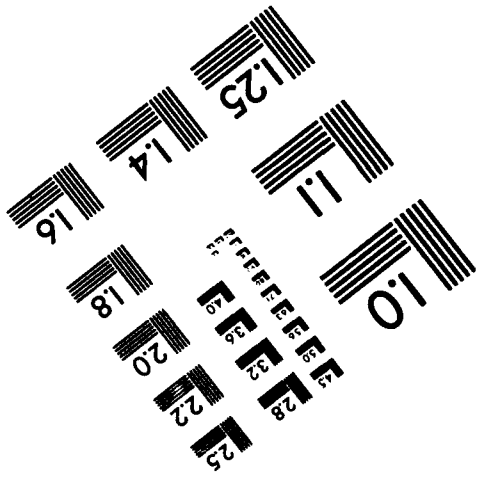
301/587-8202











MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

II. PHYSICAL SYSTEM AND THEORETICAL APPROACH

Let us consider a bulk-semiconductor model characterized by two spherical and parabolic bands. In a semiclassical picture, only the carrier distribution functions over single-particle states are considered as kinetic variables. All interactions between carriers and other types of quasiparticles, and in particular also the interaction with the external light field, are treated as perturbations. They are usually described in terms of scattering processes within Fermi's Golden Rule. Such approximations lead to the BTE for the distribution functions of electrons ($f^e_{\mathbf{k}}$) and holes ($f^h_{\mathbf{k}}$):

$$\frac{d}{dt} f^{e,h}_{\mathbf{k}} = \sum_{\mathbf{k}'} [s^{e,h}_{\mathbf{k},\mathbf{k}'} f^{e,h}_{\mathbf{k}'} - s^{e,h}_{\mathbf{k}',\mathbf{k}} f^{e,h}_{\mathbf{k}}] \tag{1}$$

with scattering rates $s^{e,h}_{\mathbf{k}',\mathbf{k}}$. On this level all coherence or correlation effects are neglected.

For the analysis of coherent phenomena, the phase relation between different types of carriers, induced by the light-matter interaction, has to be treated explicitly [3, 7, 8]. Therefore, the kinetics cannot be simply described in terms of distribution functions (intraband density matrices) but we have to include as kinetic variables also the interband polarization (interband density matrix) $p_{\mathbf{k}}$. For the unperturbed dynamics a closed set of equations can be easily obtained [3]. However, this fully coherent dynamics is modified by the presence of the various interaction mechanisms. They give rise to an infinite hierarchy of equations of motion which has to be truncated at some level.

Here, we will limit ourselves to contributions up to second order in the interaction matrix elements. The second-order terms are treated within the usual Markov approximation [3] and all second-order contributions which involve second or higher powers of the polarization are neglected. Within such an approximation scheme, the resulting system of SBE takes the general form [3]:

$$\frac{d}{dt} f^e_{\mathbf{k}} = g_{\mathbf{k}}(t) + \sum_j \frac{d}{dt} f^e_{\mathbf{k}}\Big|^j_{inco}, \qquad \frac{d}{dt} f^h_{\mathbf{k}} = g_{-\mathbf{k}}(t) + \sum_j \frac{d}{dt} f^h_{\mathbf{k}}\Big|^j_{inco}, \tag{2}$$

$$\frac{d}{dt} p_{\mathbf{k}} = \frac{1}{i\hbar}\Big\{(\mathcal{E}^e_{\mathbf{k}} + \mathcal{E}^h_{-\mathbf{k}} + \Omega_{\mathbf{k}}) p_{\mathbf{k}} + (M_{\mathbf{k}} A_0(t) e^{-i\omega_L t} + \Delta_{\mathbf{k}})(1 - f^e_{\mathbf{k}} - f^h_{-\mathbf{k}})\Big\} + \sum_j \frac{d}{dt} p_{\mathbf{k}}\Big|^j_{inco}, \tag{3}$$

with a generation rate

$$g_{\mathbf{k}} = \frac{1}{i\hbar}[(M_{\mathbf{k}} A_0(t) e^{-i\omega_L t} + \Delta_{\mathbf{k}}) p^*_{\mathbf{k}} - (M^*_{\mathbf{k}} A^*_0(t) e^{i\omega_L t} + \Delta^*_{\mathbf{k}}) p_{\mathbf{k}}], \tag{4}$$

where $A_0(t)$ is the envelope of the vector potential of the external light field with angular frequency ω_L, $M_{\mathbf{k}}$ is the dipole matrix element, and the index j refers to the various interaction mechanisms. They result to modify the system dynamics with two different contributions: (i) Coherent terms which lead to a renormalization of the free-carrier energies $\mathcal{E}^{e,h}_{\mathbf{k}}$ by a self-energy $\Omega_{\mathbf{k}}$ and of the external light field by an internal field $\Delta_{\mathbf{k}}$, and (ii) incoherent terms which lead to relaxation and dephasing processes.

Denoting by $\mathcal{F}_{\mathbf{k}}$ the generic kinetic variable (distribution functions or polarization), the SBE (2,3) can be schematically rewritten as

$$\frac{d}{dt}\mathcal{F}_{\mathbf{k}} = \frac{d}{dt}\mathcal{F}_{\mathbf{k}}\Big|_{co} + \frac{d}{dt}\mathcal{F}_{\mathbf{k}}\Big|_{inco}, \tag{5}$$

with a coherent part

$$\frac{d}{dt}\mathcal{F}_{\mathbf{k}}\Big|_{co} = C^0_{\mathbf{k}}(\{\mathcal{F}\}) + \sum_j C^j_{\mathbf{k}}(\{\mathcal{F}\}), \tag{6}$$

where $C^j_{\mathbf{k}}$ is some functional of the kinetic variables. Within the approximations discussed it is easy to realize a strong formal similarity between the various kinetic equations. In particular, the incoherent contributions

have exactly the same formal structure of the "Boltzmann collision term" in Eq. (1) also for the case of the interband polarization:

$$\frac{d}{dt}\mathcal{F}_{\mathbf{k}}\Big|_{inco} = \sum_{j}\sum_{\mathbf{k}'}[\mathcal{S}^{j}_{\mathbf{k},\mathbf{k}'}\mathcal{F}_{\mathbf{k}'} - \mathcal{S}^{j}_{\mathbf{k}',\mathbf{k}}\mathcal{F}_{\mathbf{k}}] \,, \tag{7}$$

where $\mathcal{S}^{j}_{\mathbf{k},\mathbf{k}'}$ denotes the scattering rate for a transition $\mathbf{k}' \rightarrow \mathbf{k}$ induced by the j-th interaction mechanism. This strong similarity constitutes the starting point of our generalized MC approach.

III. GENERALIZED MONTE CARLO PROCEDURE

As a starting point, let us briefly recall the basic ideas of the conventional MC simulation. As discussed above, the semiclassical transport theory is based on the BTE (1). This is, in general, a non-linear equation which is usually transformed into a locally linear one by means of a time-step solution. Due to this local linearity, the distribution function at any time t within the time-step can be written as

$$f^{e,h}_{\mathbf{k}}(t) = \sum_{\mathbf{k}'} G^{e,h}_{\mathbf{k},\mathbf{k}'}(t,t_o) f^{e,h}_{\mathbf{k}'}(t_o) \,, \tag{8}$$

where G, called Boltzmann propagator, has a direct physical interpretation: it describes the probability that a particle in state $\mathbf{k}'$ at time t_o will be found in state $\mathbf{k}$ at time t. Equation (1) and (8) can be regarded as the starting point of the traditional Ensemble Monte Carlo (EMC) technique [1, 11] which simply provides a MC sampling of the sum in Eq. (8). Such sampling is performed through a stochastic simulation of a suitable ensemble of carriers. These "simulative carriers" are, in general, not real physical particles; such an ensemble of particles is only representative for the real carrier system. For each simulative carrier, a sequence of random "free flights", interrupted by random "scattering events", is generated. It can be shown, that such a "random walk" in k-space is just a MC sampling of the Boltzmann propagator $G_{\mathbf{k}_f,\mathbf{k}_i}(t,t_o)$, where $\mathbf{k}_i$ and $\mathbf{k}_f$ denote, respectively, the initial and the final state of the generic random walk [11].

Let us now come back to the system of quantum kinetic equations (5). Since this is again a system of nonlinear equations, as in the semiclassical case, we introduce a time discretization. In the proposed numerical procedure, for each time step Δt, the coherent contributions (6) are evaluated by means of a direct numerical integration while the incoherent contributions (7) are "sampled" by means of a generalized MC simulation. Let us now focus our attention on the explicit form of the incoherent contributions (7): For all the kinetic variables (including the polarization field) the various $\mathcal{S}_{\mathbf{k},\mathbf{k}'}$ are within our approximations positive-definite quantities, i.e. they can be regarded as "true" scattering probabilities from state $\mathbf{k}'$ to state $\mathbf{k}$. However, the function $\mathcal{F}$ is now a complex function. Due to the local linearity of our transport equation over the time-step, the kinetic variable at time t can be written as

$$\mathcal{F}_{\mathbf{k}}(t) = \sum_{\mathbf{k}'} \mathcal{G}_{\mathbf{k},\mathbf{k}'}(t,t_o)\mathcal{F}_{\mathbf{k}'}(t_o) \,, \tag{9}$$

where $\mathcal{G}$ is now a generalized Boltzmann propagator corresponding to the kinetic variable. As for the semiclassical case, the propagator $\mathcal{G}$ results to be a positive-definite solution of the generalized Boltzmann equation (7). Therefore, it can be again sampled by means of a conventional EMC simulation.

Equation (9) constitutes the starting point of our generalized MC approach. As for the semiclassical case, such sampling is again performed through a stochastic simulation of a suitable ensemble of carriers which, in general, have nothing to do with real physical particles. The structure of the proposed MC procedure can then be summarized as follows: The total time is divided into time-steps. The simulation starts before the laser has been switched on. The system is chosen to be in its fundamental state, i.e. the vacuum of electron-hole pairs. The simulation then results in a loop over the various time steps. For each time step:

(i) we introduce an ensemble of $N_{\mathbf{k}}(t_o)$ "simulative particles" with $N_{\mathbf{k}}(t_o) \propto |\mathcal{F}_{\mathbf{k}}(t_o)|$; (ii) we attach to each "particle" i a phase-factor w_i according to the phase of $\mathcal{F}_{\mathbf{k}}(t_o)$; (iii) for each "particle" we sample its propagator $\mathcal{G}$ by means of a conventional EMC simulation [1] i.e. a random sequence of free flights and scattering events; (iv) at the end of the time-step the new value of $\mathcal{F}$ is evaluated: $\mathcal{F}_{\mathbf{k}}(t_o + \Delta t) = \sum_i w_i$. The usual "counting" of the particles in k is then replaced by a sum of these phase-factors w_i which reflects the complex nature of the kinetic variable $\mathcal{F}$. A similar approach has been recently used by the authors for a MC simulation of four-wave mixing experiments [4]. In this case, a MC simulation of the various, in general complex, Fourier components of the distribution functions is required.

IV. APPLICATIONS

We will now present some numerical results concerning simulations characterized by a laser energy far from the band gap (excess energy $\mathcal{E}_{ex} = 0.18$ eV, pulse duration $\tau_L = 50$ fs). This is the typical situation for energy-relaxation experiments [2].

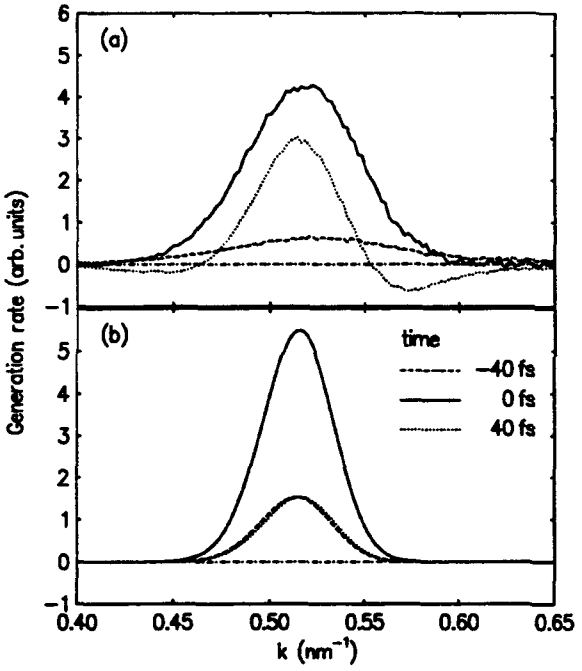


FIG. 1. Generation rates for a final density $n = 10^{16}$ cm^{-3}, (a) obtained from the SBE, and (b) in the semiclassical limit.

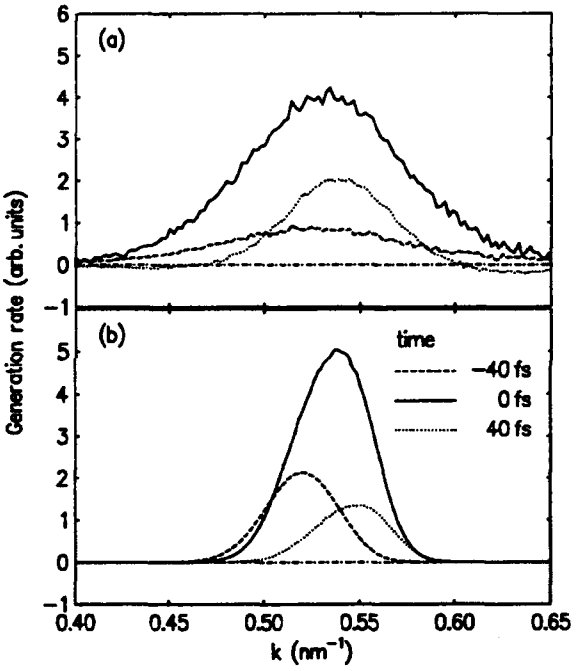


FIG. 2. Same as FIG. 1, but for a final density $n = 10^{18}$ cm^{-3}.

For a better understanding, let us first consider a "simulated experiment" characterized by a final carrier density $n = 10^{16}$ cm^{-3}. In Fig. 1 the self-consistent generation rates obtained from this MC simulation are shown as a function of the wave-vector k for different times during the laser pulse. Figure 1 (a) shows the generation rates for the full generation model while in (b) the corresponding rates for the semiclassical case are plotted. Due to the Markovian limit, the latter ones do not contain regions with negative values. On the contrary, the rates in (a) exhibit a strong time-dependence also in the shape. In particular, at short times, the shape of the generation rate is found to be much broader than estimated from the uncertainty principle using the pulse width as uncertainty of time. The reason is that the "observation time" has to be used for a correct

estimation of the line width. For lc imes we note a narrowing of the generation rate, this narrowing, however, is accompanied by the b of negative regions off-resonance which can be interpreted as a stimulated recombination process. Thus, the distribution of the generated carriers does not only become narrower with increasing time due to a generation mainly in resonance but also due to a recombination of those carriers which have been generated performing "energy non-conserving transitions" at short times.

The above result shows that a self-consistent treatment of the generation process can be important if either the evolution is analyzed already during the pulse or if some scattering mechanism is so strong that it can remove those carriers generated with the "wrong" energy before they can recombine. The latter one is exactly the situation that we obtain by repeating the above "simulated experiment" for the case of a final density $n = 10^{18}\,\mathrm{cm}^{-3}$. The self-consistent generation rates obtained in this case are shown in Fig. 2. Due to the strong efficiency of carrier-carrier scattering, already during the laser pulse carriers are removed from their initial distribution and, therefore, the stimulated recombination of Fig. 1(a) is strongly reduced. As a consequence, in this case the carrier distribution after the of the laser pulse results to be significantly broader compared to the corresponding semiclassical ca rgy shift of the generation rate especially in the high density case is due to band gap renormalization. This effect is present in the semiclassical model as well as in the full generation model. Phase space filling effe lead to the asymmetries in the generation rate in the high density case at 0 fs and especially at 40 fs.

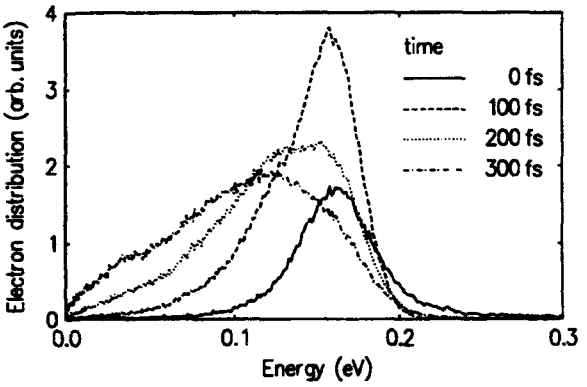


FIG. 3. Electron energy distribution for different times.

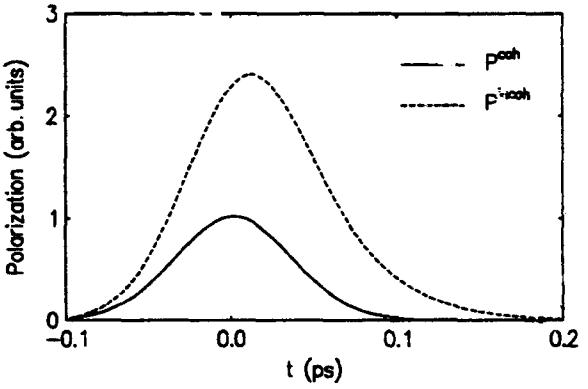


FIG. 4. Polarization as function of time.

In Fig. 3 the electron distribution as a function of energy at different times is also shown. It corresponds to the first simulated experiment ($n = 10^{16}\,\mathrm{cm}^{-3}$) with the full generation model (see Fig. 1(a)). The figure shows the typical scenario of carrier thermalization due to carrier-carrier scattering and energy relaxation due to carrier-phonon scattering. In the electron energy distribution we notice still some structure related to the discrete emission of optical phonons, which becomes more pronounced in the case of lower densities.

In Fig. 4 the polarization of the carrier system is shown as a function of time: The solid curve refers to the absolute value of the total polarization $P^{coh} = |\sum_{\mathbf{k}} p_{\mathbf{k}}|$. It decays due to the inhomogeneous broadening in k-space since each contribution $p_{\mathbf{k}}$ in the sum rotates with a different frequency. Its decay strongly depends on the width of the carrier distribution and, therefore, on the properties of the laser pulse, and its time scale is typically much shorter than that related to incoherent phenomena. The dashed curve refers to the incoherently summed polarization $P^{incoh} = \sum_{\mathbf{k}} |p_{\mathbf{k}}|$. It is a measure of the degree of coherence still present in the system and after the pulse it decays due to incoherent scattering processes. The dephasing time is of the order of 100 fs which in this case is mainly due to carrier-carrier interaction.

V. CONCLUSIONS

We have presented a numerical method which enabled us to include coherent phenomena in a MC simulation. The theory is based on the SBE for the distribution functions of electrons and holes, as well as for the interband polarization. Within this method the generation process is treated in a self-consistent way with its full time-dependence. We have obtained the time-dependence of the total polarization as well as of the incoherently summed polarization, which describes the degree of coherence still present in the system. Thus we can analyze the various times relevant for the dephasing process.

REFERENCES

[1] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer, Wien, 1989).

[2] J. Shah, Solid State Electron. **32**, 1051 (1989) and H. Kurz, Semicond. Sci. Technol. **7**, B124 (1991).

[3] T. Kuhn and F. Rossi, Phys. Rev. Lett. **69**, 977 (1992) and Phys. Rev. B **46**, 7496 (1992).

[4] A. Lohner, K. Rick, P. Leisching, A. Leitenstorfer, T. Elsaesser, T. Kuhn, F. Rossi, and W. Stolz, Phys. Rev. Lett. **71**, 77 (1993).

[5] F. Rossi, S. Haas, and T. Kuhn, Phys. Rev. Lett. **72**, 152 (1994).

[6] T. Elsaesser, J. Shah, L. Rota, and P. Lugli, Phys. Rev. Lett. **66**, 1757 (1991) and L. Rota, P. Lugli, T. Elsaesser, and J. Shah, Phys. Rev. B **47**, 4226 (1993).

[7] S. Schmitt-Rink, D. S. Chemla, and H. Haug, Phys. Rev. B **37**, 941 (1988).

[8] M. Lindberg and S. W. Koch, Phys. Rev. B **38**, 3342 (1988)

[9] R. Zimmermann and M. Hartmann, phys. stat. sol. (b) **150**, 365 (1988).

[10] R. Binder, D. Scott, A. E. Paul, M. Lindberg, K. Henneberger, and S. W. Koch, Phys. Rev. B **45**, 1107 (1992).

[11] F. Rossi, P. Poli, and C. Jacoboni, Semicond. Sci. Technol. **7**, 1017 (1992).

MONTE CARLO MODELING OF THE DYNAMIC SCREENING EFFECTS ON ULTRAFAST RELAXATION OF PHOTO-EXCITED CARRIERS IN GaAs

N. Nintunze and M.A. Osman
Department of Electrical Engineering and Computer Science
Washington State University
Pullman, WA99164-2752

Abstract

The relaxation of photo-excited carriers in GaAs is investigated at 300K using an Ensemble Monte Carlo approach. The screening of the carrier-carrier(c-c) interaction is treated dynamically using a momentum and frequency dependent dielectric function in the Random Phase Approximation and in the Plasmon Pole Approximation. Calculated effective carrier temperatures agree with experimental data for time delays longer than 200fs.

I. INTRODUCTION

Femtosecond lasers allow the study of thermalization and initial relaxation of photo-excited carriers in bulk semiconductors [1, 2] as well as in quantum well structures[3]. Experimental investigations are paralleled by theoretical calculations, mostly Monte Carlo(MC) simulations[1, 2, 4]. Recently polarization related phenomena have also been introduced in an MC model[5], and such model is expected to give more information about the early moments following semiconductor laser excitation. MC simulations with the screening of the c-c interaction treated statically deviate from experimental data during the first 100-500fs after laser excitation, because static screening underestimates c-c scattering rates[1]. On the other hand, Molecular Dynamics(MD) provides a better agreement with the experiment[2] because MD realizes dynamic screening. However, MD requires an extensive CPU time. In this paper, we report on an alternative model for treating dynamic screening of the c-c interaction in MC simulations. The model takes into account the wave-vector and frequency dependence of the dielectric function. Simulations are performed for two different implementations. The first uses the Random Phase Approximation(RPA) to the dielectric function as in the analytical calculations of Meyer and Bartoli[6]. The second approach implements the Plasmon Pole Approximation(PPA) to the dielectric function, recently used by Collet[7] in analytical calculations. The model is applied to carrier relaxation in p-type and i-type GaAs. A laser of 2eV photon energy and 50fs duration is assumed. Two excitation densities of $2 \times 10^{18} cm^{-3}$ and $2.5 \times 10^{16} cm^{-3}$ are simulated, so that c-c and carrier-phonon processes could be observed. Results show a good agreement between the proposed dynamic approach and the experiment for time delays longer than 200fs.

II. COMPUTATIONAL MODEL

The Monte Carlo model for the electrons includes non-parabolic Γ, L and X valleys. The MC program for electrons considers elastic acoustic phonon scattering, the intervalley deformation potential, the screened polar-optical phonon, dynamically screened electron-electron(e-e) and electron-hole(e-h) scattering. The MC program for holes is based on a 3-band model where the heavy and light hole bands are warped, while the split-off band is spherical parabolic. Scattering mechanisms included are elastic acoustic phonon, optical phonon scattering, self-consistent screened polar optical phonon and screened hole-hole scattering.

An alternative formulation of the c-c interaction is implemented, where use is made of a wavevector and frequency dependent dielectric function in the interaction potential. Following the approach of Meyer and Bartoli[6], the interaction potential can be written as:

$$U(\mathbf{q}) = \frac{e^2}{q^2 \epsilon_o \epsilon(q, \omega = \mathbf{q}.\mathbf{v_{cm}})} \tag{1}$$

where $\epsilon(q,\omega)$ is the dielectric function, q is the relative wavevector, and $\mathbf{v_{cm}}$ is the velocity of the center of mass. The general expression of the RPA is used in the determination of the free-carrier contributions to the dielectric function. Making use of the above interaction potential, the e-h scattering rates are given by:

$$\Gamma_{e-h}(\mathbf{k_h}) = \frac{2p\mu e^4}{2\pi\hbar^3 \epsilon_o{}^2 g} \sum_{k_h} f_{\mathbf{k_h}} \int_o^g \frac{dq}{q^3 |\epsilon(q,\omega)|^2} \tag{2}$$

where e is the electron charge, p is the hole density, μ is the reduced mass of the interacting particles, and g is a relative wavevector defined as:

$$\mathbf{g} = 2\mu(\frac{\mathbf{k_e}}{m_e} - \frac{\mathbf{k_h}}{m_h}) \tag{3}$$

where $\mathbf{k_e}(\mathbf{k_h})$ is the electron(hole) wavevector, and $m_e(m_h)$ is the electron(hole) mass. In the e-e interaction, p is replaced by n, and electron wave vector and mass rather than hole wavevector and mass are used in g.

The scattering rates are calculated in the Monte Carlo program as follow. First the components of the center of mass velocity of the interacting particles and the components of the relative wavevector are found. In order to determine q, the knowledge of both the initial and final carrier states are required, which is a problem as scattering has not taken place yet. An approximate value can be found by using virtual scattering processes: for a given electron, we choose at random an ensemble of target particles to scatter with, determine the resulting virtual next state and change in wavevectors for each collision. However, the states of the scattering charge carriers are not updated. Thereafter, it becomes possible to determine the components of q and to calculate the frequency. The next step is the determination of the total dielectric function, where the free carrier contributions are obtained by numerically integrating the corresponding expressions. After repeating this process with an ensemble of target carriers, an average of the c-c scattering rate for the given electron is found. The average c-c scattering rates obtained in the current calculation were about 3 to 7 times higher than the values obtained in the self-consistent static approach of reference [4]. Ideally, the c-c scattering rates should be calculated every time a scattering process takes place. However, the computation time can be reduced by using a three-dimensional scattering table in k-space, where a given scattering rate is stored in a corresponding k-cell to be accessed whenever necessary. For an iteration step of 2.5fs with the scattering table updated every iteration for the first 200fs, one picosecond simulation of 5000 particles took about 11 hours on the IBM3090.

Recently Collet[7] used the PPA dielectric function. This is a simpler analytical approximation, which results in saving of CPU time, as the dielectric function is free of integrations. The scattering rates are still determined using eq.2, and following the procedure presented above. This implementation requires about 3 times less CPU time when compared to the RPA approach.

III. RESULTS

The cooling of photo-excited carriers in p-GaAs was investigated at 300K. The assumed 2.0eV laser photon energy excites carriers from the heavy, light and split-off bands in the ratio 0.46:0.32:0.22, in accordance with recently published data[1]. Figure 1 shows the time evolution of the electron

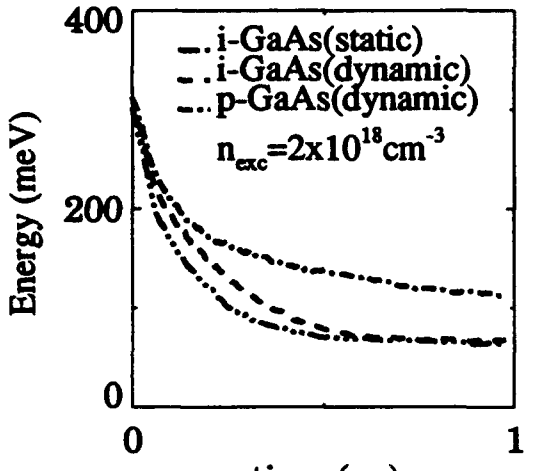


Fig.1. Electron energy

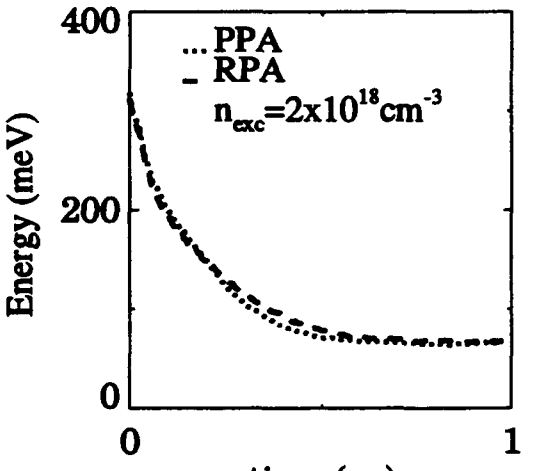


Fig.2. RPA and PPA results

energy, where the origin of the time axis corresponds to the peak of the exciting laser. The figure shows that dynamic screening of the c-c interaction results in faster cooling of the electrons when compared to the time-dependent static model of reference [4]. The faster cooling during the first 500 femtoseconds is mostly due to electron energy loss to holes. Accordingly, holes heat up, but the gained energy is rapidly dissipated through non-polar and polar phonon emission. The faster cooling in p-GaAs demonstrates that added hole concentration due to doping results in additional electron energy loss to holes. In the dynamic case, electrons redistribute energy and momentum among themselves and lose energy to the holes at a faster rate when compared to the static approach. Figure 2 compares the results of the RPA and PPA implementations of the dynamic screening of the c-c interaction. It can be seen that the PPA approximates well the RPA results. The two cases are within 10% of each other.

Figure 3 shows the theoretical effective carrier temperatures extracted from the slope of the near exponential tail of the luminescence intensity for the RPA, the PPA, the static screening model, and the experimental data[1] in i-GaAs, for $n_{exc} = 2 \times 10^{18} cm^{-3}$. The calculated electron temperature is within 10% of the experimental values for time delays longer than 300fs in i-GaAs, and within 10% at 200fs in p-GaAs. It can be seen that the static screening model always predicts electron temperatures higher than the experiment and the present model for time delays shorter than 500fs. However, the dynamic screening models still predict electron temperatures that are higher than experimental data at time delays shorter than 200fs. This is partly due to the approximations used in the calculation of the dielectric functions. Additionally, this model does not include the full details of the energy band structure, ignores higher order quantum corrections in the scattering rate calculations, and polarization and quantum coherence which become important at short time scales.

The simulation was repeated at the lower excitation and doping density of $2.5 \times 10^{16} cm^{-3}$ to ex-

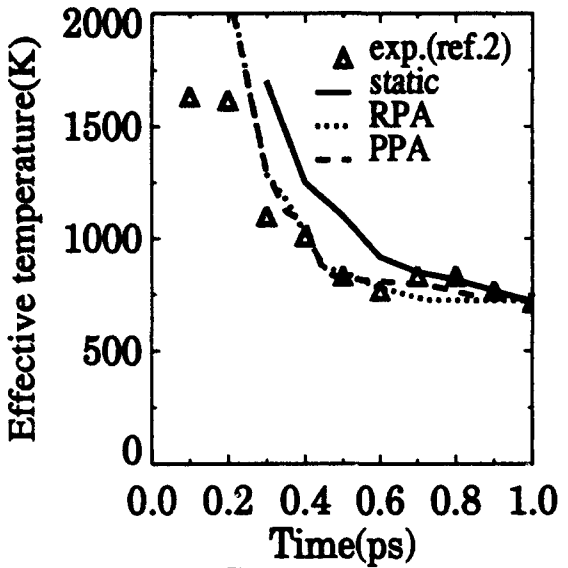


Fig.3. Carrier temperature

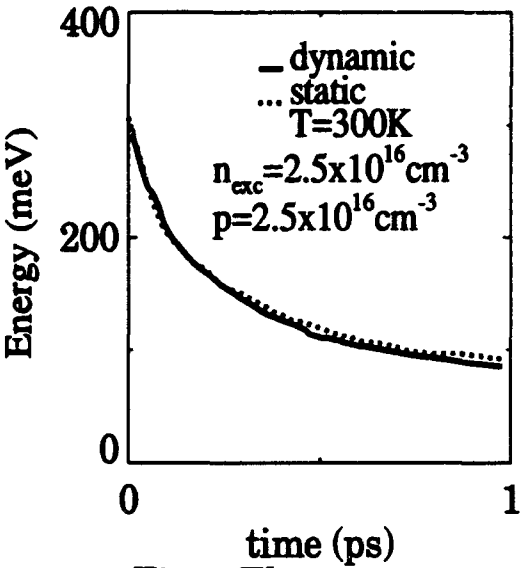


Fig.4. Electron energy

amine the role of c-c and c-phonon processes for both dynamic and static screening. It can be seen in Fig.4 that the average energy is not strongly affected by the screening model used in calculating c-c scattering rates, because, at low carrier densities, the e-phonon interactions are the main energy loss channels.

IV. CONCLUSION

An expression for dynamically screened c-c scattering that takes into account the wavevector and frequency dependence of the dielectric function in the RPA and in the PPA was developed and implemented in an MC program. The relaxation of photo-excited carriers in GaAs was examined using these two screening models, in addition to static screening. The obtained carrier effective temperatures are in good agreement with the experiment for time delays longer than 200fs in the dynamic cases, compared to 500fs in the static case.

References

[1] U. Hohenester, et.al, Phys.Rev. B47(20), 13233(1993)

[2] L. Rota, et.al, Phys.Rev. B47(8),4226(1993)

[3] S.M. Goodnick, P.Lugli, W.H. Knox, and D.S. Chemla, Solid-State Electronics 32, 1737(1989)

[4] M. A. Osman and D.K. Ferry, Phys.Rev. B 36, No.11, 6018(1987)

[5] T. Kuhn and F. Rossi, Phys.Rev. B46, 7496(1992)

[6] J.R. Meyer and F.J. Bartoli, Phys. Rev. B28, 915(1983)

[7] J.H. Collet, Phys.Rev.B 47, 10279(1993)

FULLBAND ENSEMBLE MONTE CARLO MODELING OF HIGH-FIELD TRANSPORT IN THE ZnS PHOSPHOR OF AC THIN FILM ELECTROLUMINESCENT DEVICES

Shankar S. Pennathur, Keya Bhattacharrya, John F. Wager, and Stephen M. Goodnick
Department of Electrical and Computer Engineering
Oregon State University, Corvallis, OR 97331-3211

Abstract

A fullband ensemble Monte Carlo investigation of high-field transport in the ZnS phosphor of ac thin-film electroluminescent (ACTFEL) devices is presented. A full band dispersion computed using empirical pseudopotentials is used to model the first two conduction bands in ZnS. Computed electron energy distributions at high fields reveal a reasonable fraction of electrons energetic enough to impact-excite luminescent centers in the phosphor layer.

I. INTRODUCTION

Alternating current thin-film electroluminescent devices are used in the production of high-resolution, flat-screen displays and are being increasingly researched [1-5]. An ACTFEL device essentially consists of a wide bandgap semiconductor such as ZnS (referred to as the phosphor layer) sandwiched between two insulating layers. Carriers that are sourced into the semiconductor by surface states in the semiconductor-insulator interface, are accelerated under the influence of very high electric fields. The energetic electrons traversing the phosphor layer are then responsible for impact exciting the (intentionally introduced) luminescent centers. Luminescence is obtained as the excited electron states in these centers radiatively relax to their ground states. An understanding of the high-field carrier transport in the phosphor layer and the physics of the different threshold processes such as band-to-band impact ionization and impact excitation of luminescent impurities is essential for device design, especially when newer phosphors are being continually developed in the quest for a full-color EL display.

In this paper, we present the results obtained using a full band model for ZnS for fields in the range 1-2 MV/cm. By including band-to-band impact ionization as well as impact excitation of Mn^{2+} centers, a unified picture of the physical processes crucial to electroluminescence is achieved.

II. MONTE CARLO MODEL

We use a full band dispersion for ZnS, computed using empirical (local) pseudopotentials[6]. The first two conduction bands included in the simulation, span in energy to values sufficiently higher than the most energetic electrons encountered for the electric fields considered. The low-energy scattering rates are computed using a non-parabolic dispersion for the different valleys in the first conduction band. The scattering rates however are corrected at higher energies using the full band density of states thereby forcing the scattering rates to behave as the density of states. Figure 1 shows the total scattering rates computed in the Γ valley of ZnS at 300K. While polar-optic phonon scattering is the dominant scattering mechanism at low energies, intervalley scattering mechanism becomes important beyond 1.5 eV (which is roughly the energy separation between the Γ valley and the X and L valleys). Other scattering mechanisms included in the model are ionized impurity scattering, acoustic phonon scattering, band-to-band impact ionization (at high energies), and impact excitation of Mn^{2+}. Impact ionization is included in the model using a simple Keldysh formulation [7], which specifies the ionization rate as,

$$\Gamma_{ii}(E) = \Gamma_{ph}(E_{th})\, P\, (\frac{E - E_{th}}{E_{th}})^2 \qquad (1)$$

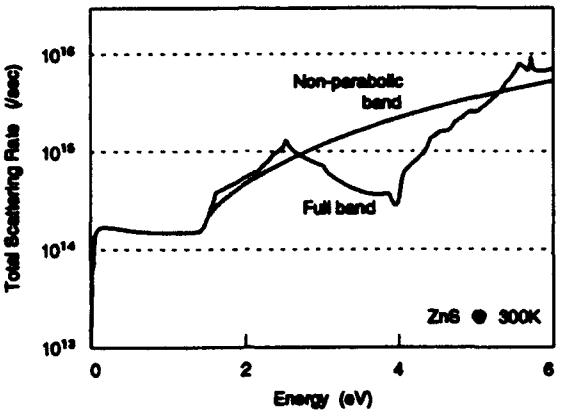


Figure 1: Total scattering rate in the central valley of ZnS at 300 K. Rates for non-parabolic band and adjusted rates for fullband are shown.

where Γ_{ii}, the energy dependent impact ionization rate is proportional to $\Gamma_{ph}(E_{th})$, the total phonon scattering rate at E_{th}, the threshold energy for impact ionization, given as

$$E_{th} = \frac{2m_e + m_h}{m_e + m_h} E_G \tag{2}$$

where m_e, m_h and E_G are the electron and hole band edge effective masses and the energy gap respectively. Our work included a threshold (computed as above) of 4.3 eV and the value of P used was 100. Figure 2 shows the variation of the impact ionization coefficient α_{ii} with the inverse of the electric field, obtained from the Monte Carlo model, fitted to reported experimental values [8] by tuning the deformation potentials for intervalley phonon scattering in the second band, showing a reasonable fit in the electric field values of most interest. Impact excitation is the process in

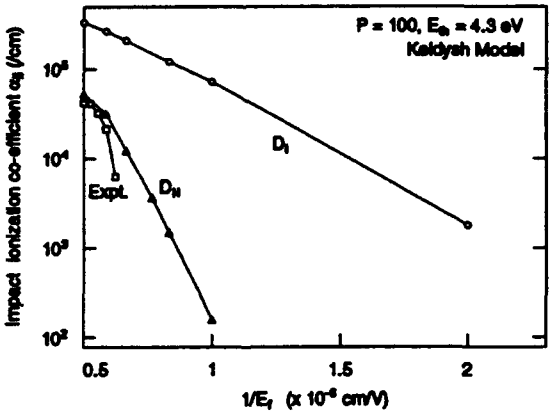


Figure 2: Plot of impact ionization coefficient α_{ii} as a function of inverse electric field. Plots D_I and D_{II} correspond to two different sets of deformation potentials for phonon scattering in the second band.

which a hot electron interacts with a luminescent impurity in the host phosphor, to excite valence electrons in the luminescent center to excited states, losing energy in the process. In this work, we

model the excitation cross-section for a center with a threshold energy E_{ie} as [9]

$$\sigma(E) = \frac{c^2 e^4}{4\pi\epsilon_\infty^2 E^2}\sqrt{\frac{E - E_{ie}}{E}} \tag{3}$$

where e is the electronic charge, and ϵ_∞ is the high frequency dielectric constant. The constant factor c^2 is related to the overlap integral between the hot electron and the interacting impurity's electron wavefunctions, and was fitted to obtain a peak value of $1\times10^{-15}\text{cm}^2$ for the cross-section. The associated scattering rate is then simply given as

$$\Gamma_{ie} = \sigma(E) v_d N_{li} \tag{4}$$

where v_d is the average velocity of the carriers, and N_{li} the density of the centers in the host phosphor (typically about 0.5 atomic %).

III. RESULTS AND DISCUSSION

Figure 3 shows the energy distribution of electrons for three different (typical) phosphor fields, along with the excitation cross-section (in arbitrary units) of Mn^{2+} ions. It is observed that the distribution gets hotter (increasing average energies) with increasing phosphor fields. For Mn^{2+} centers (used for yellow luminescence) with an excitation threshold energy of about 2.1 eV, it is seen that a considerable number of electrons in the ensemble are energetic enough to cause impact excitation. By counting the number of impact excitation events occurring during the simulation

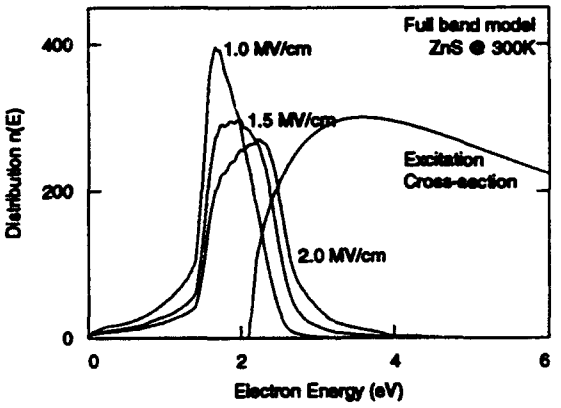


Figure 3: Electron energy distribution at three different phosphor fields plotted along with the excitation cross-section of Mn^{2+} centers.

at steady state over a length of time, an estimate is obtained of the average number of impact excitations effected by an electron as it traverses the entire length of the phosphor layer. This parameter is linked to the maximum observable brightness of the devices. Figure 4 shows a plot of this internal quantum yield parameter as a function of the phosphor field, revealing an almost linear variation, and the existence of a possible cutoff field as the lower limit. Figure 4 also shows the variation of the number of ionization events per transferred electron. There are few ionization events for fields below 1.5 MV/cm, but a significant number of events is observed at higher fields. This variation is consistent with the threshold energies of the impact excitation and impact ionization processes being 2.1 eV and 4.3 eV respectively.

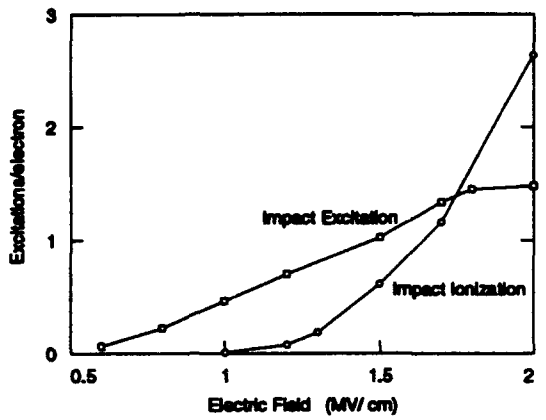


Figure 4: Plot of recorded impact ionization and impact excitations events per transferred electron as a function of the phosphor field.

IV. CONCLUSIONS

A full band Monte Carlo modeling of ZnS phosphor in ACTFEL devices reveals that band-to-band impact ionization plays a crucial role in stabilizing the electron distributions. The steady state electron energy distributions obtained for different phosphor fields reveal a significant fraction of the electrons energetic enough to participate in impact exciting luminescent centers in the host phosphor. Increasing phosphor fields results in hotter energy distributions and relatedly, the estimated internal quantum yield varies linearly with the phosphor field. Impact excitation processes while being the basis of the functionality of these devices do not affect the hot electron energy distributions to any significant degree.

ACKNOWLEDGMENT

The usage of the supercomputing resources of Sandia National Laboratories and NCSA at University of Illinois, in performing the simulations presented in this work, is acknowledged. We thank Dr. Ravaioli at University of Illinois, and Dr. P. Vogl of Walter Schottky Institute, Munich, for their assistance, and Dr. Bringuier for use of his unpublished results. We also acknowledge the support of the U.S. Army Research Office under Contract No. D11L03-91G0242.

REFERENCES

[1] B. K. Ridley, *J. Phys. C: Solid State Phys.*, **16**, 3373 (1983).

[2] E. Bringuier, *J. Appl. Phys.*, **70**, 4505, 1991.

[3] R. Mach and G. O. Müller, *J. Cryst. Growth*, **101**, 967, 1990.

[4] K. Bhattacharyya, S. M. Goodnick, and J. F. Wager, *J. Appl. Phys.*, **73**, 3390, 1993.

[5] K. Brennan, *J. Appl. Phys.*, **64**, 4024, 1988.

[6] M. L. Cohen and T. K. Bergstresser, *Phys. Rev.*, **166**, 789, 1966.

[7] L. V. Keldysh, *Sov. Phys. JETP*, **21**, 1135, 1965.

[8] T. D. Thompson and J. W. Allen, *J. Phys. C: Solid State Phys.*, **20**, L499, 1987.

[9] E. Bringuier and K. Bhattacharyya, Unpublished work.

Advances and Opportunities in the Design and Modeling of Vertical-Cavity Surface-Emitting Lasers

J. W. Scott, B. J. Thibeault, S. W. Corzine and L. A. Coldren
ECE Dept., University of California at Santa Barbara. Santa Barbara, California 93106

Abstract

The numerical models available for optoelectronic devices are quite limited. In the case of vertical cavity surface emitting lasers, almost no models have been developed. For the numerical modeler, this presents the opportunity to develop new insights but makes it difficult to determine which effects are dominant. Our group has maintained a very active program in vertical cavity lasers in which there has been a strong interaction between numerical modeling and experiment. Based on that experience, numerical simulations have been developed which predict the optical, electrical and current to light characteristics for index-guided vertical cavity lasers. This paper discusses the various physical effects we have modeled and points out the research areas which demand more involved calculations.

Vertical cavity lasers represent a relatively new class of semiconductor lasers. The development of epitaxially grown distributed Bragg reflectors with reflectivities in excess of 99% has enabled their realization in recent years. Interest in the lasers was originally based on their low divergence beams and potential for array applications. Initial experimental results showed low output powers and high drive voltages. Improvements have lead to low drive voltages, high differential efficiency, sub-milliamp thresholds and output powers well above 1 mW. More recently, it has been demonstrated that they can be designed to have temperature stabilized operation. The complex nature of the devices and the time and expense of fabrication cycles has driven the development of numerical models to aid the device design. Along with optical models to determine the electromagnetic fields, we have developed an LI simulator which includes thermal effects, carrier diffusion, stimulated emission and spatial hole burning[1].

Due to their small size and the distributed nature of the reflectors, many of the models used for conventional in-plane semiconductor lasers must be modified. The two lasers are contrasted in Fig. 1a. The low optical losses of the vertical cavity require accurate calculations. For example, additional

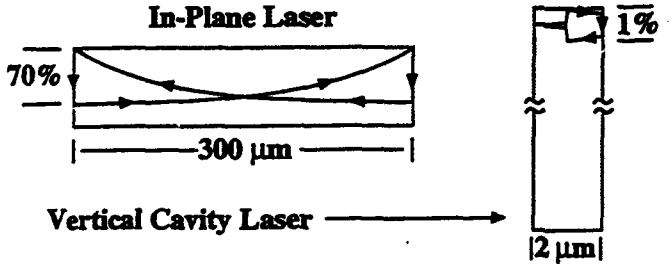


Fig 1a. Schematic comparing the optical power flow in an in-plane laser and a vertical cavity laser. The low losses of the high Q cavity requires accurate calculations

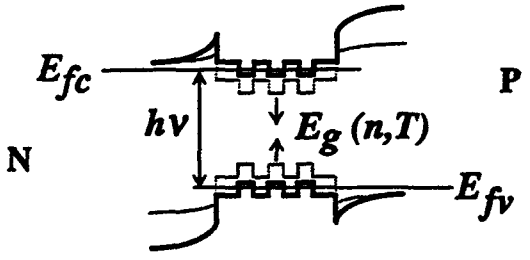


Fig 1b. Relationship of the cavity mode hν and the band gap of the laser active region. The quasi-Fermi level separation is required to be greater than the photon energy.

round trip propagation losses of only 0.5% would reduce the optical efficiency nearly in half, resulting in a large reduction in the slope of the LI curve. The short cavity length results in a wide spacing of the Fabry-Perot modes, and thus only a single longitudinal mode falls within the optical gain spectrum of the quantum wells. As shown schematically in Fig. 1b, the bandgap shrinks toward lower energy (longer wavelengths) due to both ohmic heating and increasing carrier densities while the cavity mode, $h\upsilon$, essentially stays fixed. For lasing to occur, the necessary population inversion

requires the quasi-Fermi level separation to be greater than the photon energy. At the optimum alignment of the cavity mode with the gain peak, the threshold current is at a minimum. Depending on their relative position at room temperature, a gain offset can be used to produce temperature stabilized operation. Once temperature rises have reduced the bandgap below the photon energy of the mode, carrier densities rise rapidly and carrier leakage can then become very important. Confirmation of these effects in three quantum well $In_{0.2}Ga_{0.8}As$ vertical cavity lasers has been reported in reference[2]. An earlier design of the 11μm diameter laser showed a threshold minimum near 35°C and a strong increase in carrier leakage over the $Al_{0.2}Ga_{0.8}As$ cladding layers at higher temperatures. Grown with a longer cavity and higher barrier $Al_{0.5}Ga_{0.5}As$ cladding layers, the newer design exhibited a minimum threshold current of 1.6 mA at 70°C with a variation of less than 0.5 mA over a 80°C range.

The small size of these lasers makes it possible to achieve very low threshold currents. At the same time, the small size makes surface effects very important. Even with the relatively low surface recombination velocities of these InGaAs wells, ~ 2 x 10^5 cm/s, surface recombination accounts for more than half of the threshold current for etched pillar, bottom emission designs. With improvements in the growth and fabrication technologies, new structures are being investigated. A schematic of an intra-cavity contacted laser[3] is shown in Fig. 2a. Both contacts to this top surface emitting laser are made using p and n doped layers within the optical cavity. A current constriction etch above the active layer is used to force the current into the optical mode, removing surface recombination from the region of current injection. The resulting LI characteristics are shown in Fig. 2b. The lasers have sub-milliamp thresholds with output powers well above 1 mW. Spatial hole

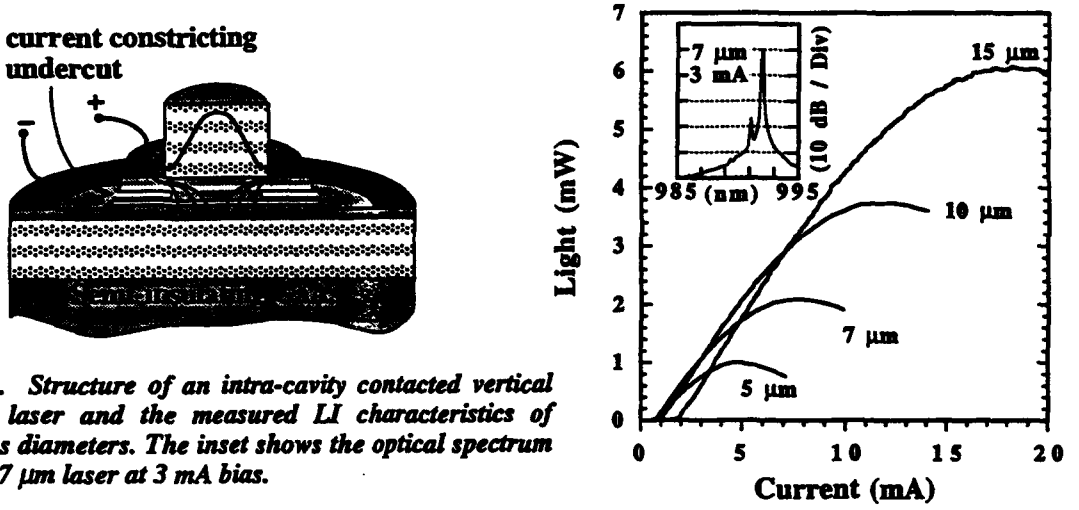


Fig. 2. Structure of an intra-cavity contacted vertical cavity laser and the measured LI characteristics of various diameters. The inset shows the optical spectrum of the 7 μm laser at 3 mA bias.

burning and surface scattering losses are competing effects, resulting in a wide variation in the transverse modal properties of the lasers depending on diameter. The 5 μm device lases in a single fundamental mode (MSR > 30 dB) while the 15 μm laser has four competing modes. The use of intra-cavity contacts allow both mirrors to be undoped, enabling microwave characterization using high speed probes and co-planer waveguides to make the transition from the probes to the lasers. To model these devices the following models have been developed.

I. Optical Model

The problem has been assumed to be separable into axial (growth direction) and transverse mode profiles. The transverse modes are calculated using the standard approach for determining the HE modes of a dielectric waveguide in cylindrical coordinates, using the averaged value for the index in the semiconductor. A transmission matrix approach is used to calculate the resonant cavity wavelength, threshold gain and optical losses for the axial fields. A transmission matrix is calculated for each layer in the cavity, and then they are multiplied together to find the transmission and reflection coefficients for the entire structure. To determine the lasing condition, a search is made in

wavelength and gain to find the poles in reflectivity, so that light is emitted for no incoming field. The resulting gain is the threshold optical gain required for the particular design. Included in the formulation are complex dielectric constants, allowing the addition of optical gain or loss in any layer. It is important to use accurate models for the dispersion of the index for the various materials, in the AlGaAs system the data from Afromowitz[4] is often used. For the optical losses, the dominant losses are free carrier absorption. This plasma effect is modelled phenomenologically by using an absorption coefficient of $11cm^{-1}$ per $10^{18}cm^{-3}$ p-type carriers and $5cm^{-1}$ per $10^{18}cm^{-3}$ n-type carrier for a wavelength of 1 μm in GaAs. Very little data of the accuracy required exists in the literature. Proper calculation of this effect requires complex bandstructure models. It is important because resistance due to lower doping leads to heating which limits output power while higher doping leads to optical losses which reduce the output power. A balanced design requires more detailed knowledge of the tradeoffs.

To determine the optical efficiency, defined as the fraction of photons generated that are emitted out of the cavity, the threshold gain is calculated with and without optical losses. The ratio is the optical efficiency, typically between 50 and 70% for our designs. The transmission coefficient, T_r, can be calculated using the ratio of the field inside and outside the cavity. As shown schematically in Fig. 1, the round trip gain must compensate for the losses of transmission and internal loss. This is expressed as $G = L + T_r$ where the round trip gain G is related to the material gain g by:

$$G = 2gl_{act}\zeta_{enh} \qquad (1)$$

where the two is for two passes (round trip), l_{act} is the total quantum well thickness and ζ_{enh} is the enhancement factor due to the standing wave effect. For our three 80Å quantum well design, ζ_{enh} has a value of 1.83 instead of the ideal 2 for an infinitely thin layer placed at the antinode. A final note is that the inclusion of diffraction losses requires complete 3D solutions, a much more complex problem given the relatively large index discontinuities at each interface of the distributed Bragg reflectors. Furthermore, gain-guided structures pose an even more complex problem. The transverse modes are dominated by the weak index guide generated by the thermal gradients associated with the current flow, and thus the thermal, electrical and optical properties must be solved self consistently in 3D. We restrict ourselves here to strongly index-guided structures. An additional point is that the local temperature may need to be included in the calculation as the bandgaps (and hence indices of refraction) of the various layers shift at different rates relative to the lasing wavelength resulting in a changing transmission coefficient for the mirror. While we have not yet included this effect, others have reported[5] on significant reductions in the transmission coefficient at elevated temperatures.

II. Gain Model

The gain model for vertical cavity lasers must provide the material gain as a function of carrier density, temperature and wavelength. Due to the small size of the lasers, their thermal impedance is high, and typically junction temperature rises as the output "rolls over" are above 100°C. In addition, carrier densities exceed $10^{19}cm^{-3}$ due to spatial hole burning and bandgap shifts from heating. The position of the cavity mode shifts due to index dispersion at a rate of ≈0.8Å/°C while the bandgap shifts at ≈3.4Å/°C. Thus the gain spectrum must be known as well since the gain peak shifts its relative position during laser operation. To determine the gain spectrum, we use a first principle gain model that includes valance band mixing and the effects of strain[6]. Typical output is shown in Fig. 3. It has proved to be critical to include the band shrinkage effect in order to explain the threshold characteristics observed as a function of temperature. This has been included using the phenomenological formula $\Delta E_g = -Cn^{1/3}$ and can be seen as the shift of the band edge towards longer wavelengths with increasing carrier densities in Fig. 3a. Finally, the gain model also provides the spontaneous emission as a function of carrier density. This is calculated using the band structure, the matrix elements, and assuming a virtual photon in each radiation mode. As can be seen in Fig. 3b, it is inappropriate to assume a linear relationship for the peak gain as a function of carrier density. For the following models, either curve fits or lookup tables for the gain data shown in Fig. 3 have been used to speed the calculations.

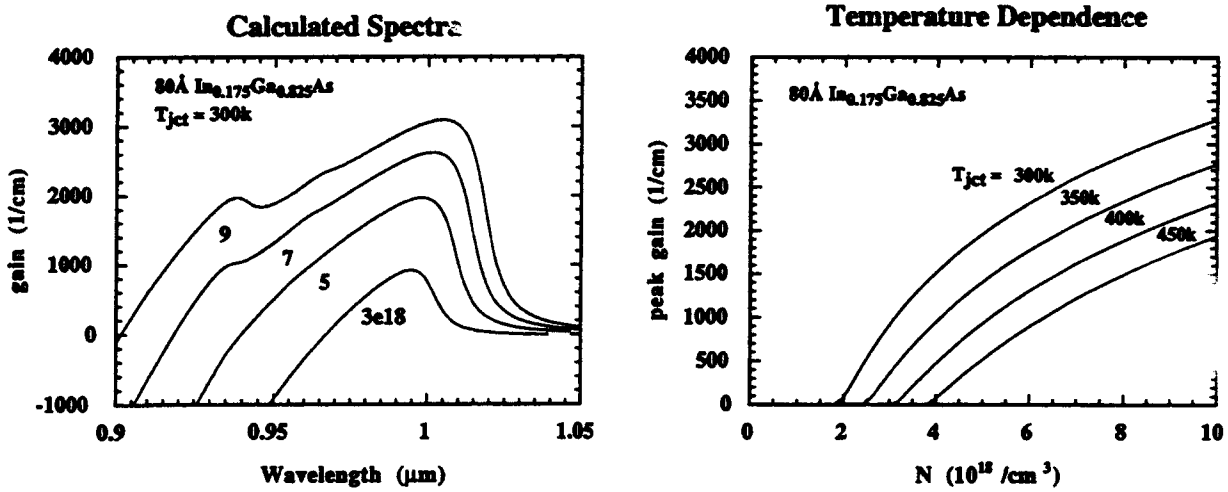


Fig. 3 Calculated gain spectra and peak gain for the strained InGaAs quantum wells

III. LI Model

The current to light (LI) model is shown schematically in Fig. 4a. The carrier density profile is solved self consistently in cylindrical coordinates. The radial ambipolar diffusion currents, stimulated emission, spontaneous emission, Auger recombination, surface recombination and carrier leakage currents are balanced in each cell. Input parameters include the transverse mode profile, the injected current density profile for each voltage, the cavity mode shift with temperature, and an effective thermal conductivity. The temperature rise is assumed constant across the junction and calculated using the analytic formula for a disc on a semi-infinite substrate[7]:

$$\Delta T_{jct} = 1/4 r_{act} k_{sub} \qquad (2)$$

where ΔT_{jct} is the junction temperature rise, r_{act} is the active region radius and k_{sub} is the effective thermal conductivity. Note that the thermal conductivity of the ternary and quaternary materials can be 10-20 times higher than the binaries such as GaAs due to random alloy scattering of phonons. The simple etched pillar structures that we fabricate make the analytic approximation reasonable. Fully buried structures such as the proton implanted gain-guided designs require numerical calculation. Complete continuous wave (CW) LI characteristics are calculated using this approach. Most of our calculations assume a single transverse mode for simplicity, however, the calculations can be run with multiple transverse modes at the expense of slower convergence.

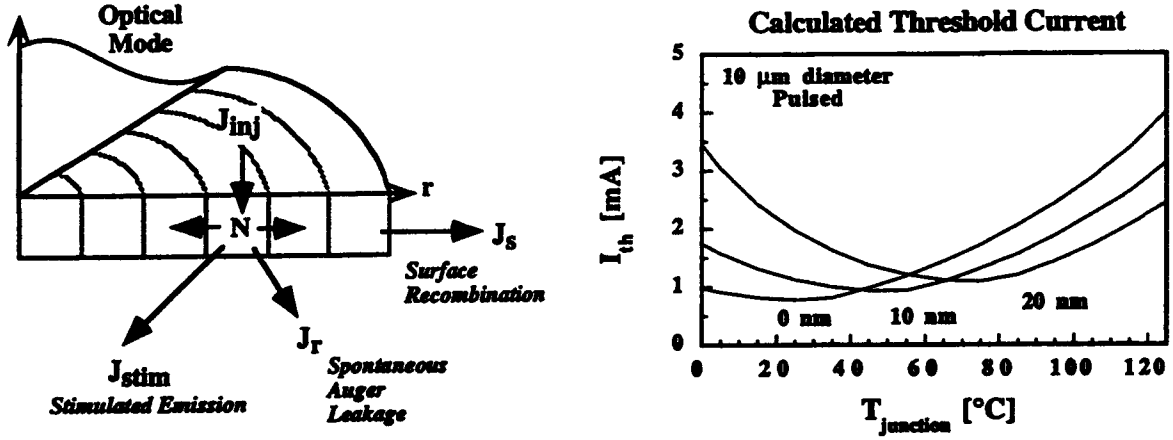


Fig. 4. Schematic of the LI model and a simulation of the threshold current for varying gain offsets

Results of a calculation with the LI model for determining the threshold current under pulsed operation ($\Delta T_{jct} = 0$) are shown in Fig. 4b. Several different curves are shown, corresponding to varying offsets of the cavity mode and the gain peak. The calculation used an ambipolar diffusion constant of 20cm^2/s and a surface recombination velocity of 2x10^5 cm/s. This temperature insensitive operation is radically different from the behavior of in-plane lasers, whose threshold current always increases with increasing temperature. The reduced threshold with increasing temperature is a result of the interplay of the band shrinkage effect and the spectral gain curve while the increase in threshold at elevated temperatures is primarily due to Auger recombination. The coefficient used for our calculations of these InGaAs quantum wells has been fit to the data using the typical $C_A n^3$ dependence with an Auger coefficient C_A =1x10^{-29} cm^6/s, three times higher than bulk GaAs. It is interesting that bandgap renormalization and Auger recombination play a dominant role in these lasers, typically they are second order effects for in-plane lasers in the GaAs system. As researchers attempt to make vertical cavity lasers at other wavelengths, accurate models for these effects will become more important. In particular, Auger recombination at high carrier densities can be reduced by modifying the bandstructure with strain. This will be very important in the telecommunication wavelengths and requires much more complex numerical calculations.

IV. Current Injection

For top emitting laser structures or those using dielectric distributed Bragg reflectors, ring contacts such as those shown in Fig. 1 are inevitable. The concern is that current crowding will occur at the periphery of the laser where the fundamental optical mode is weak. The result will be reduced internal efficiency and the tendency to promote multimode operation by enhancing the gain near the perimeter. We have taken care to model the current injection for our intra-cavity contacted designs. Accurate models for the JV characteristics of the p-i-n and heterojunctions are required as it is the differential resistance which determines the distribution of current once the diodes have been forward biased. Particularly in Be doped AlGaAs, the dopants can diffuse during growth resulting in unknown dopant profiles. To model our diodes we have grown test active regions and measured their JV characteristics under uniform injection conditions. The measured characteristics are used in the simulation. To calculate the injected current density as a function of radius the laser is divided into a mesh in cylindrical coordinates as shown in Fig. 5a where the nonlinear materials are lightly shaded. The voltage and current distribution is found using an Alternating Direction Iteration (ADI) technique where the diodes have been linearized. The diode resistance values are adjusted during the iteration process so that the final solution uses the correct JV characteristics. The results of such a calculation are shown in Fig. 5b. For this particular doping, diodes and geometry, the current flowing through the p-i-n junction shows current crowding at the edge only above 6 mA of drive current. Since the threshold current for this laser is below 1 mA, this is an acceptable design.

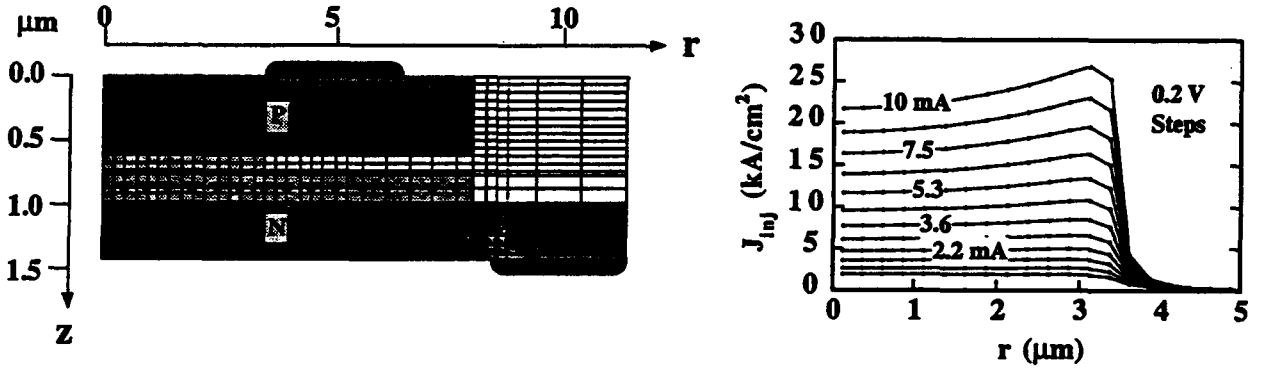


Fig. 5. Grid used for IV simulation and the injected current density profile at the p-i-n junction

V. Intra-Cavity Laser Simulation

Combining the output of the current injection calculations with the LI simulation gives a complete current, voltage, light characteristic simulation of the intra-cavity contacted devices of Fig. 2. The results of such a calculation are shown in Fig. 6a for the 7 μm diameter laser. With all input variable determined by the model for the uniformly injected lasers, the only parameter adjusted for fitting was

the gain offset. Spectral measurements and room temperature photoluminescence of the active material had indicated a gain offset on the order of 10 nm, in good agreement with the chosen gain offset of 5nm. The calculated carrier density profiles are shown in Fig. 6b. Two important points can be determined from the plot. First, that the effects of surface recombination have been greatly

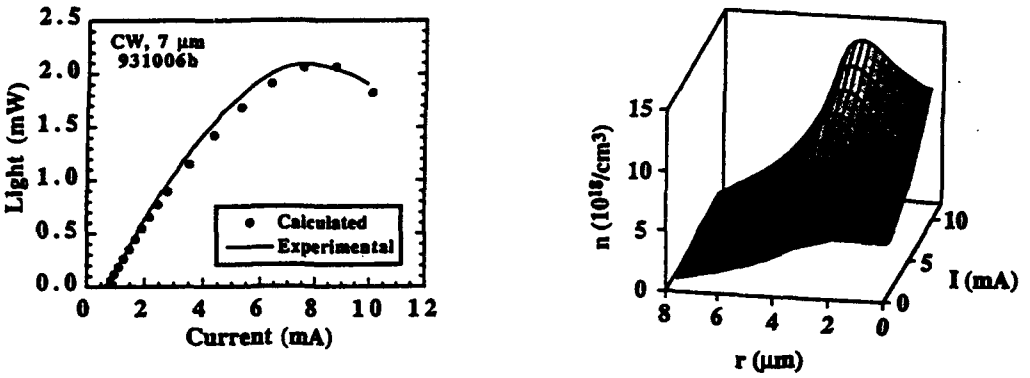


Fig. 6. Comparison of calculated and experimental LI characteristics for the 7μm intra-cavity laser of Fig. 2. The radial carrier density profile for varying bias currents is shown on the right.

reduced by this current constricted design. Second, that there is a large inefficiency resulting from the carriers injected at the edge of the optical waveguide where the optical mode is weak. If, instead, the current constriction could be made to a diameter less than the transverse mode diameter, greater internal efficiency would be observed while the single mode operation would be enhanced. These more difficult "gain apertured" designs are currently under investigation in our lab and others.

VI. Conclusion

Models for the optical,electrical and LI characteristics of vertical cavity lasers have been presented to demonstrate the current state of the laser simulation and to point out the dominant device physics. While agreement with experiment is good, most of the approximations have been made possible by restricting our analysis to strongly index guided structures. As research begins to focus on designs which combine the features of the index guided and gain guided structures, fully self consistent solutions of the current, thermal and optical problems will be required. In addition, many important physical effects have been included using phenomenological models. The high optical gain requirements of the short vertical cavity pose challenges to develop efficient vertical cavity lasers at shorter and longer wavelengths. Comprehensive models of free carrier absorption, Auger recombination and band gap renormalization in strained and unstrained materials may provide insights into better designs for these more challenging material systems. This work was sponsored by ARPA via the Optoelectronics Technology Center.

References

[1] J. W. Scott, R. S. Geels, S. W. Corzine and L. A. Coldren, " Modeling temperature effects and spatial hole burning to optimize vertical cavity surface emitting laser performance" *IEEE J. Quantum Electron.*, **29** (5) (1993)

[2] J. W. Scott, D. B. Young, S. W. Corzine and L. A. Coldren, " Modeling the current to light characteristics in index guided vertical-cavity surface-emitting lasers," *Appl Phys Lett.*, **62** (10) pp1050-1052 (1993)

[3] J. W. Scott, B. J. Thibeault, D. B. Young and L. A. Coldren, "High Efficiency Sub-Milliamp Vertical Cavity Lasers with Intra-Cavity Contacts" *IEEE Photonics Tech. Lett.*, **6** (6), (1994)

[4] M. A. Afromowitz, "Refractive index of $Ga_{1-x}Al_xAs$," *Solid State Comm.*, **15**, (59) (1974)

[5] J. Piprek, H. Wenzel and G Sztefka, "Modeling Thermal Effects on the Light vs Current Characteristic of Gain-Guided Vertical-Cavity Surface-Emitting Lasers," *IEEE Photonics Tech. Lett.*, **6** (2), pp 139-142 (1994)

[6] S. W. Corzine, R. H. Yan and L. A. Coldren, "Theoretical Gain in Strained InGaAs/AlGaAs Quantum Wells, Including Valence Band-Mixing Effects," *Appl. Phys. Lett.*, 57 (27) pp 2835-2837 (1990)

[7] S. S. Kutateladze, V. M. Borishanskii, *A Concise Encyclopedia of Heat Transfer*, Oxford: Permagon,1966, p43

A TRANSITION MATRIX STUDY OF LASER DYNAMICS

Muhammad A. Alam and Mark S. Lundstrom
Purdue University, West Lafayette, IN 47907

ABSTRACT

Semiconductor laser dynamics are simulated by a transition matrix approach. We analyze a set of experiments to illustrate the intrinsic gain dynamics of a laser and to clarify the role of various scattering rates in determining carrier relaxation under lasing conditions. We conclude with a study of the effects of photons and hot phonons on gain compression. We find that while hot phonons effects are important, they are not the rate limiting factor of a semiconductor laser.

I. INTRODUCTION:

Recently, there has been considerable interest in the microscopic dynamics of semiconductor quantum well lasers both because of their technological importance and the interesting device physics issues. Although carrier relaxation in quantum wells (QWs) in the presence of carrier-carrier and carrier-phonon interactions has been investigated by a number of experimental and theoretical groups[1], its implication for lasers is less well understood. To understand QW laser carrier dynamics, one must consider ultrafast stimulated emission, carrier thermalization by the large thermal carrier population at threshold, transport in the separate confinement layers, and the capture of carriers from the barrier region to the quantum well itself.

In this paper, we explore the intrinsic response of a quantum well laser systematically using a new transition matrix approach (TMA)[2]. Specifically, we examine the role of electron-electron and electron-nonequilibrium phonon interactions on laser gain and the distribution function. We also study how carrier dynamics are affected by strong electron-photon interactions.

II. TRANSITION MATRIX APPROACH:

Briefly, the TMA is a Monte Carlo technique to directly solve the time dependent but space-independent Boltzmann equation. A transition matrix relates the distribution of particles (electrons, phonons, or photons) at a given time t to the particle distribution at time $t + \delta t$. For density independent scattering events, such as carrier relaxation in absence of light or Coulomb interaction, this matrix is time invariant, however, when the scattering processes depend on the density, then this matrix will evolve in time. To compute the transition matrix, the input momentum space is first divided into a large number of bins, a large number of particles is injected in each of these bins, and these particles are tracked for a time δt by a 2-D Monte Carlo simulation [2,3]. At this point, positions of the particles in momentum space are noted, and the ratio of the particles for a pair of initial and final bins gives the transition matrix element. If one has a initial distribution of particles in momentum space, by repeatedly multiplying the evolving distribution by the transition matrix, one can track particle evolution as a function of time.

Our model includes electron, phonon, and photon dynamics. Electron transport has been treated semiclassically. The scattering mechanisms included are polar optical phonons, acoustic phonons, and electron-electron scattering [3]. Static screening was assumed for e-e scattering, but dynamic screening can be readily treated by the TMA. Transition matrix elements were first computed assuming that the destination states were always empty and the partner state (for electron-electron scattering) was always full. During simulation, one obtains dynamic estimate of

the distribution function, f, and the matrix elements are modified accordingly. For example, ee scattering matrix element at the n-th time step is given by

$$t_{ij,mn}^{(n)} = t_{ij,mn}^{(0)} f_j^{(n)}(1 - f_m^{(n)})(1 - f_n^{(n)}), \tag{1}$$

where $t_{ijmn}^{(0)}$ is the scattering rate of an electron at state i colliding with a particle at state j, with final destination to states m and n. This rate was computed by assuming that $f_j^{(0)}$ equals unity and $f_m^{(0)}$ and $f_n^{(0)}$ equal zero. At each time step, this rate is then modified by the Fermi factors as shown above to account for the band filling effects.

We considered hot phonon effects by solving a Boltzmann equation for phonons in the relaxation time approximation [4]. The phonons emitted during a time step are sorted according to their wavevectors, q. In the next time step, a fraction of these phonons will be reabsorbed, and some of the phonons will decay by nonelectronic means with a finite lifetime. The excess phonons affect the POP scattering rates in the following way,

$$t_{ij}^{(n)} = t_{ij}^{0} \frac{N_q^{(n)} + \frac{1}{2} \pm \frac{1}{2}}{N_q^{(0)} + \frac{1}{2} \pm \frac{1}{2}}, \tag{2}$$

where t_{ij} is the POP scattering rate from state j to state i, $N_q^{(0)}$ is the equilibrium phonon population and $N_q^{(n)}$ is the phonon population at n-th time step.

Finally, the electron-photon interaction is described through standard multiband effective mass formalism. The hole bandstructure was computed by using a 4x4 k.p Hamiltonian for the quantum well. The optical transition matrix elements were computed using polarization and wave-vector dependent band to band scattering rates assuming strict wavevector and energy conservation. The dynamic, energy dependent broadening of the joint density of states were subsequently accounted for by a energy dependent Lorenzian broadening factor [5].

III. RESULTS:

Using the model described above, we discuss three sets of experiments to clarify gain dynamics issues. First, we analyze a pump-probe experiment in which the quantum well is initially empty, next we discuss pump-probe experiments for a gain-inverted laser diodes operated in the amplifier mode, and finally, we examine the effects of gain compression and hot phonons on laser performance.

In the first type of experiments, the quantum well may either be empty or modulation doped. A light pulse excites carriers from the valence band to the conduction band, and a delayed laser beam then probes the carrier distribution as it relaxes. These experiments study carrier relaxation in quantum well via POP and electron-electron scattering. Since they have been simulated in detail by Goodnick and Lugli [3], they provide a good test for our approach.

In this simulation, we use a rectangular laser pulse to excite carriers at 0.23 eV above the conduction band. Figure 1 shows the relaxation dynamics as a function of time. Initially, carriers relax by emitting polar optical phonons. Therefore, the distribution function develops a set of well defined peaks separated by the phonon energy (35 meV). At higher densities, these peaks are washed away be electron-electron scattering within 800 fs; at lower densities, the peaks persist for a longer time. These conclusions are in reasonable agreement with Goodnick's results, which were obtained by direct Monte Carlo simulation.

The second set of experiments, also a pump-probe variety, are more relevant to laser gain dynamics. In these experiments, facets of a laser diode are coated with an antireflection coating, so that it operates in the amplifier mode. Various bias currents put this diode in various levels of

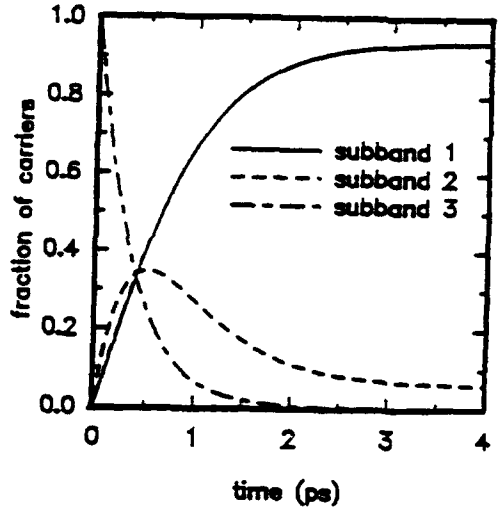


Fig. 1 Relaxation of carriers in an AlGaAs/GaAs/AlGaAs QW. Carriers are injected at t=0 into the third subband and they subsequently relax too subband 1.

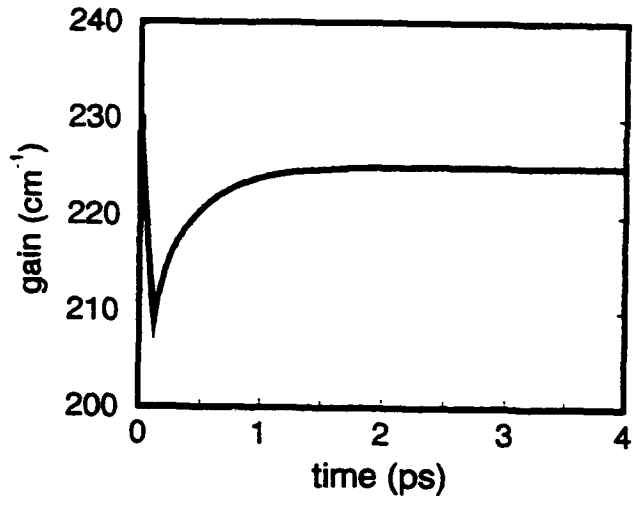


Fig. 2 The gain dynamics of a 150 A QW after a 100 fs of TE laser pulse.

inversion. A pump beam is then launched in the cavity. Depending on the bias level, the pump beam will either stimulate emission (in gain region) or stimulate absorption (in the loss region). A delayed probe beam, as before, monitors the distribution function as the spectral hole is filled in or as the spectral 'heap' diffuses away (in energy). This experiment is relevant to lasers, because lasers are generally biased near threshold, so the nature of screening and the relative importance of various scattering events is expected to be very similar. This is also a significantly more difficult system to describe by simulation, because the change in probe gain may be only a few percent, therefore, noise in the simulation is unacceptable.

Figure 2 shows the dynamics in the gain region. The pump beam is a rectangular pulse of 100 fs duration. The lowest subband population goes down immediately due to stimulated emission, so does the probe gain. Within a few ps, the subband population relaxes by redistributing the carrier population and the relaxation time of $\approx$ 0.72 ps is well within experimental range. As the spectral hole fills up, so does the gain. The gain finally saturates at a lower value, however, because of net decrease of carrier concentration in the quantum well. In the absorption region, on the other hand, electrons are pumped into the conduction band. The increase in number of carriers is reflected in both gain spectrum and subband population. Also, the net increase of carriers is reflected in higher saturated values for gain. In both the gain and absorption region, the average energy goes up - cold carriers are removed in the gain region from the conduction band, hot carriers are added in the loss region.

The role of electron-electron scattering is significant in these experiments and needs some discussion. When a spectral hole is burned in the distribution function by a pump beam, e-e scattering is not very effective in filling up the hole, because electron-electron scattering requires two empty final levels, and the spectral hole provides only one. This explains why it takes picoseconds to fill up a spectral hole, as opposed to femtoseconds as one would expect for a system dominated by electron-electron scattering. Therefore, the time requirement is more consistent with phonon scattering requirements.

Finally, we discuss gain compression and hot phonon issues. Gain compression is a important figure of merit, and it affects the maximum frequency of oscillation and maximum gain in a very significant way. Recently, there has been a lot of discussion on the factors determining gain including band filling effects, spectral hole burning, hot phonon effects. We can now address these issues from

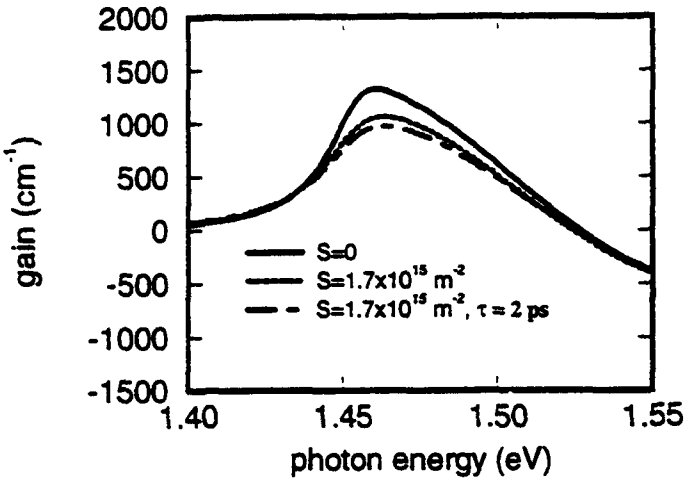


Fig. 3 Gain compression due to stimulated emission and hot phonon effects.

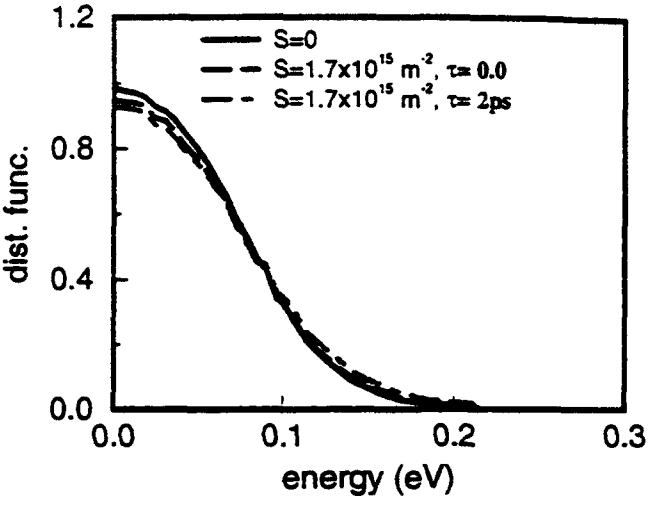


Fig. 4 Electron distribution function at subband 1 in the presence of stimulated emission and hot phonons.

a semiclassical point of view.

Figure 3 shows three gain curves, and Fig. 4 shows the corresponding distribution functions. The first one is the gain curve for a laser biased to $3.0\text{x}10^{16}$ per m^{-2} without any light or hot phonon effects. Broadening due to polarization dephasing has been accounted for by including both the inscattering and outscattering rates of POP and e-e scattering. The second curve is the gain curve in presence of photons ($1.7\text{x}10^{15}\ \text{m}^{-2}$) at 1.46 eV. If we increase phonon lifetime, there is even larger heating and gain compression. In the first case, the gain compression factor is $1.2\text{x}10^{-16}\ \text{m}^{-2}$, in the second case this factor increases to $1.6\text{x}10^{-16}\ \text{m}^{-2}$. If a photon lifetime of 2 ps is assumed, then the maximum frequency of oscillation is 195 GHz without hot phonons and 187 GHz with hot phonons. In contrast to recent phenomenological treatments, we conclude that the net effect of hot phonons is not very significant, and intrinsic dynamics is not the rate limiting factor of diodes now in fabrication.

IV. SUMMARY:

Using a new transition matrix approach, we have analyzed a set of experiments to clarify the gain dynamics of quantum well lasers. Our investigation sheds light on gain dynamics explored by recent pump-probe experiments. Also, we find that gain compression due to hot phonons is not as significant as it was previously thought. We ascribe this to the q dependence of hot phonon effects. To understand the gain dynamics of issues one should also include transport, this is a research issue we shall investigate in the future.

REFERENCES

1. Jagdeep Shah, Ed., *Hot Carriers in Semiconductor Nanostructures: Physics and applications*, Academic Press, (1992).
2. Muhammad A. Alam and Mark Lundstrom, SPIE OE LASE Proc., (1994).
3. S. M. Goodnick and P. Lugli, Phys. Rev. B 37, 2578 (1988).
4. P. Lugli et. el., Phys. Rev. B 39(11), 7852 (1989).
5. M. Asada, *Quantum Well Lasers*, 97, Academic Press (1993).

COUPLING CLASSICAL CARRIER TRANSPORT, CAPTURE, AND SIZE QUANTIZATION IN A QUANTUM WELL LASER SIMULATOR

Matt Grupen and Karl Hess
The Beckman Institute
Department of Electrical and Computer Engineering
University of Illinois
405 N. Mathews Ave., Urbana, IL 61801, 217-333-9734

Abstract

A model for the coupling of the classical and quantum regions of a quantum well laser diode is presented. The calculation of carrier transport in different regions of the laser is first discussed. Then a discretization of the carrier continuity equations in and around the quantum well is presented in detail. The model is finally tested through the simulation of three different $GaAs/AlGaAs$ lasers. The results are discussed, and the importance of carrier capture is demonstrated.

I. INTRODUCTION

There are several different approaches to quantum well laser simulation. One approach is based on the Statz-de Mars rate equations [1] and is widely used because of its simplicity and flexibility. However, it can not be used to calculate the actual transport of carriers, and therefore it can not yield local variations in the carrier densities. An alternative approach involves the self-consistent solution of the semiconductor equations with carrier transport. Several such simulators have been produced for a two-dimensional cross section of the laser [2,3,4]. Some self-consistent simulators use drift-diffusion theory throughout the entire device [5], while others have complemented drift-diffusion theory with thermionic emission theory at abrupt heterojunctions [2,4]. However, the code presented in this paper is the only laser simulator that treats the capture of carriers and size quantization in the quantum well.

This paper will discuss the simulation of carrier transport in the laser simulator called Minilase. In particular, the numerical implementation of capture and quantization will be presented in detail. Modulation responses will then be calculated by Minilase to illustrate the importance of carrier capture.

II. THE LASER MODEL

The device equations for a laser diode consist of those that describe the carrier dynamics and those that describe the photon dynamics. The carrier dynamics can be expressed by Poisson's equation and the continuity equations for electrons and holes. Since the spatial distributions of free carriers in the quantum well are determined by size quantization, Schrödinger's equation for the conduction and valence subbands must also be solved in this region [6].

The photon dynamics can be calculated by including optical gain, spontaneous emission, and cavity losses in a photon rate equation for each longitudinal mode. The mode gain and spontaneous emission are determined by the amount of inversion, the optical matrix element, and the photon

density. The inversion charge is determined by the carrier dynamics, and the optical matrix element can be calculated from Fermi's Golden Rule [7]. The photon populations can be mapped into local photon densities through the solution of Maxwell's equations for the laser cavity [6].

Since the continuity equations include the divergence in carrier flux, carrier transport must be calculated throughout the device. Because the electric field is normally moderate in the bulk regions of the device, the parabolic band approximation and drift-diffusion theory are appropriate in these regions [8]. Abrupt heterojunctions, however, are modeled in Minilase as discontinuities in the band edges. Carrier fluxes at such discontinuities are better calculated with the ballistic transport found in thermionic emission theory [9]. This theory requires a carrier injected from one material into another to immediately scatter into thermal equilibrium with the carriers in the latter material. This is a good approximation for a single heterojunction because most carriers will inelastically scatter within a mean free path after crossing the interface.

Since the quantum well is the region of optical gain, it is very important to accurately calculate the carrier dynamics in and around this region. Drift-diffusion theory can be used in the surrounding bulk regions, and ballistic transport determines the injection of carriers into the well. However, since the size of the well is on the order of a mean free path, thermionic emission must be augmented to account for the probability that a carrier traverses the well without inelastic scattering. This can be done by carefully discretizing the quantum well as shown in figure 1.

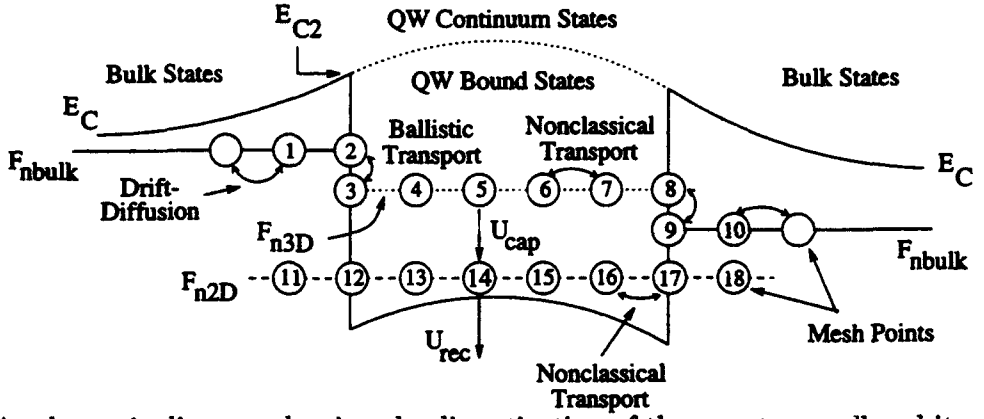


Figure 1: A schematic diagram showing the discretization of the quantum well and its coupling to the classical regions.

The quantum well must be partitioned into the continuum states above the well and the lower energy bound states. Our choice of the boundary between continuum and bound states is neither purely classical nor quantum mechanical. It accounts in part for quantum reflections, and quantum resonances are incorporated through the scattering rate between the two partitions. Associated with each partition is a unique quasi-Fermi level. Carriers injected from the bulk into the quantum well are restricted to the continuum states. This ballistic injection can be easily calculated from the quasi-Fermi levels for the bulk and the continuum using an expression similar to thermionic emission. For example, using the node numbers in the figure,

$$j_{n,2\to3} = A^* T^2 \left[\exp\left(\frac{F_{n\text{bulk},2} - E_{C2}}{kT} \right) - \exp\left(\frac{F_{n3D,3} - E_{C2}}{kT} \right) \right] \tag{1}$$

A carrier in the continuum can transfer back into the classical region via an analogous expression, or it can inelastically scatter and be captured into a bound state. The scattering process is modeled with a net capture term similar to Hall-Shockley-Read recombination [10]. The spatial distributions of quantum well carriers are determined by their wave functions. The continuum wave functions are approximated by plane waves, and those in the bound states are obtained from Schrödinger's

equation. Also, according to size quantization, all the mesh points in the continuum must have the same quasi-Fermi levels, and similarly F_{n2D} must be constant.

To illustrate the discretization of the quantum well, the electron continuity equations for the continuum and the bound states, discretized in one dimension, are shown below. The two-dimensional case is analogous. First, let us pick the continuum node 5 (any continuum node can be chosen). The equation for this node is then

$$\sum_{i=3}^{8} \frac{\partial n_i}{\partial t} l_i = j_{n,2\rightarrow 3} - j_{n,8\rightarrow 9} - \sum_{i=3}^{8} U_{\mathrm{cap},i} l_i \tag{2}$$

where l_i is the length associated with node i according to the box integration method [11]. The equations for continuum points j=3,4,6,7,8 are simply $F_{n3D,j} = F_{n3D,5}$. Similarly, let us pick bound mesh point 14. Its continuity equation is

$$\sum_{i=11}^{18} \frac{\partial n_i}{\partial t} l_i = \sum_{i=11}^{18} (U_{\mathrm{cap},i} - U_{\mathrm{rec},i}) l_i \tag{3}$$

and the equations for bound points j=11$\rightarrow$13,15$\rightarrow$18 are simply $F_{n2D,j} = F_{n2D,14}$.

Although the quantum well mesh points that share the same real space coordinates have different quasi-Fermi levels, the electrostatic potential is single valued and is determined by the total charge at each real space coordinate. To illustrate, let us pick heterojunction node 2 (nodes 3 or 12 could also have been picked) and write the discrete Poisson's equation in one dimension.

$$0 = \sum_{i=1,4} \varepsilon_{2,i} \frac{\psi_2 - \psi_i}{|x_2 - x_i|} - (N^+_{\mathrm{D},2} - N^-_{\mathrm{A},2})(l_2 + l_3) - \sum_{j=2,3,12} (p_j - n_j) l_j \tag{4}$$

Note that node 2 belongs to the bulk and node 3 belongs to the continuum. Therefore, $l_2 = (x_2 - x_1)/2$ and $l_3 = (x_4 - x_3)/2$ according to the box integration method. Since the bound states overlap into the bulk regions, $l_{12} = (x_{13} - x_{11})/2$. Equation (4) determines ψ at node 2, and the potentials at nodes 3 and 12 immediately follow because ψ is continuous (to a very close approximation) at the interface, i.e. $\psi_j = \psi_2$ where j=3,12.

III. Results

The importance of the carrier dynamics in and around the quantum well is made very clear by the modulation response. To demonstrate this, three 100 Å $GaAs$ single quantum well laser diodes were simulated. One diode contained an $Al_{0.3}Ga_{0.7}As$ separate confinement region (SCR) measuring 1450 Å on each side of the well. Another diode contained an SCR that measured 2950 Å on each side. In the third device, the SCR was asymmetric, measuring 2950 Å on the p-side and 950 Å on the n-side. The results are shown in figure 2. The frequencies and heights of the resonant peaks are influenced by photon dynamics. The rates of roll-off near zero frequency, however, are directly related to carrier capture, and they differ for the three devices. Reducing the size of the n-side has caused the asymmetric device to roll off more gradually than the wide device. This is due primarily to the number of holes that were not captured by the well. Figure 3 shows these holes for the two devices. These carriers are not used to modulate the optical output power but, instead, contribute to parasitic diffusion capacitance. The number of wasted holes is clearly smaller for the asymmetric device, and therefore this device has smaller diffusion capacitance and a more gradual roll-off in the modulation response.

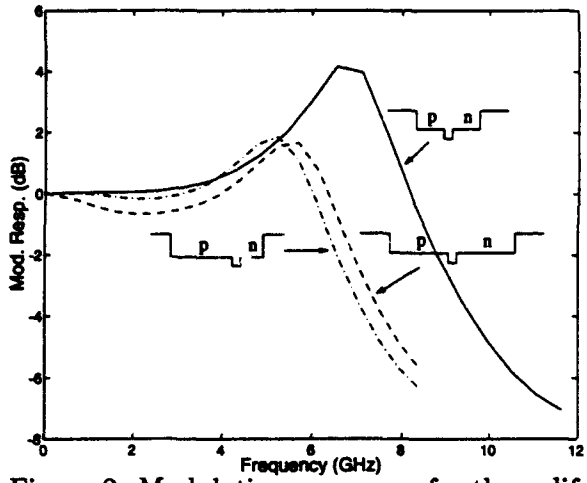


Figure 2: Modulation responses for three different separate confinement regions.

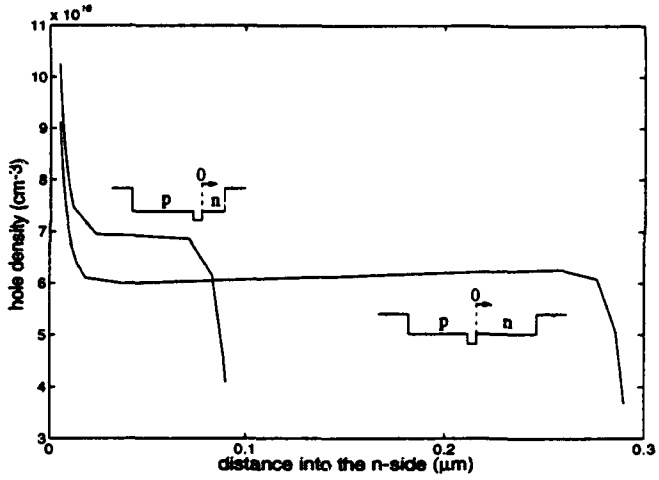


Figure 3: Hole densities on the n-sides of the asymmetric and wide lasers.

IV. Conclusions

An accurate simulation of a quantum well laser diode requires the treatment of carrier and photon dynamics throughout the entire device. The carrier dynamics can be calculated by treating transport appropriately in different regions of the laser. Coupling the quantum well to the surrounding bulk regions is particularly critical because it determines the pumping of the active region. We have described a model which permits the simulation of classical transport in bulk regions as well as size quantization and carrier capture in the quantum well. Furthermore, we have presented a discretization of this model that is suitable for numerical solution.

This work was funded by NSF support through the NCCE.

[1] H. Statz and G. de Mars, in *Quantum Electronics*, edited by C.H. Townes, (Columbia University Press, New York, 1960), p. 530.

[2] G.H. Song, K. Hess, T. Kerkhoven, and U. Ravaioli, Europ. Trans. Telecomm. Related Tech., **1**, 375 (1990).

[3] S. Seki, T. Yamanaka, and K. Yokoyama, J. Appl. Phys., **71**, 3572 (1992).

[4] Z.M. Li, K.M. Dzurko, A. Delâge, and S.P. McAlister, IEEE J. Quantum Electron., **28**, 792 (1992).

[5] T. Ohtoshi *et al.*, Solid-State Electron., **30**, 627 (1987).

[6] M. Grupen, U. Ravaioli, A. Galick, K. Hess, and T. Kerkhoven, in *Proceedings of the SPIE OE/LASE Conference, Los Angeles* (SPIE, Bellingham, WA, 1994).

[7] G. Lasher and F. Stern, Phys. Rev., **133**, A533 (1964).

[8] K. Hess, *Advanced Theory of Semiconductor Devices* (Prentice Hall, Englewood Cliffs, NJ, 1988).

[9] M. Grupen, K. Hess, and G.H. Song, in *Simulation of Semiconductor Devices and Processes, Vol. 4*, edited by V. Fichtner and D. Aemmer (Hartung-Gorre Verlag, Konstanz, Germany, 1991), p. 303.

[10] M. Grupen, G. Kosinovsky, and K. Hess, in *Proceedings of the International Electron Device Meeting, Washington, DC* (IEEE, New York, 1993) p. 23.6.1.

[11] S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer-Verlag, Wien-New York, 1984).

MONTE CARLO SIMULATION OF HIGH FIELD ELECTRON TRANSPORT IN ZnS WITH MODIFIED DENSITY OF STATES

John Fogarty, Weiran Kong and Raj Solanki
Dept. of Electrical Engineering and Applied Physics
Oregon Graduate Institute
P.O. Box 91000 Portland, OR 97291

ABSTRACT

An ensemble Monte Carlo simulation has been developed for electrons in ZnS which includes three nonparabolic valleys in the first conduction band as well as a single nonparabolic valley in the second conduction band. The density of states for the first band was modeled phenomenologically to resemble the density of states obtained by numerical pseudopotential calculations. This density of states was used only in the calculation of the scattering probabilities and no attempt was made to modify the conductivity effective mass. The scattering probabilities included in the simulation are those due to ionized impurities, acoustic phonons, polar optical phonons, intervalley phonons, and impact ionization. It is found that the inclusion of the second band and modification of the density of states effective mass has a dramatic effect on the energy distribution, especially at fields above 1MV/cm. When the field exceeds 1MV/cm, the distribution begins to show a secondary peak associated with the second band that is not found in previous investigations. At lower fields, the second band contributes a high energy tail but does not alter the peak of the energy distribution noticeably. The importance of introducing the second band is that, because the width of the first band is less than the band gap, it is believed that nearly all band to band impact ionization is due to electrons in the second conduction band. This has specific relevance in the operation of ZnS AC thin-film electroluminescent devices.

I. INTRODUCTION

An understanding of the high field transport properties of electrons in ZnS is essential for the design and operation of many AC thin film electroluminescent (ACTFEL) devices. These devices operate at fields ranging from 0.5 MV/cm to 2.0 MV/cm and rely on highly energetic (hot) electrons for both the excitation of the activator as well as the for the process of carrier multiplication via band to band impact ionization. At present very few attempts have been made to investigate the behavior of this material in the high field regime. Previous studies, [1,2] have indicated that the electron distribution in ZnS is relatively cool with the high energy tail of the distribution ending at less than 4.0 eV even for fields up to 2.0 MV/cm. While Brennan [1] included the full structure of both the first and second conduction band, it is not clear what his results imply in terms of impact ionization in ZnS, since his work is mainly a comparison between ZnS and ZnSe, and results for fields in excess of 1.0 MV/cm were not reported. Bhattacharyya et al [2] used a model that included only the first conduction band in a infinite nonparabolic three valley model. Their model is not consistent with the band structure or density of states of ZnS at these high fields because the electrons were found to have energies exceeding the energy depth of the first conduction band. Müller et al [3] have used a single parabolic valley in the first conduction band. This again is unrealistic at these fields and this model leads to an unstable distribution with ballistic, loss free transport.

One method of including higher bands is to use a method suggested by Brunetti et al [4]. This model uses a band structure so chosen as to give the density of states function a good fit with experimental and theoretical determinations. Our model consists of three valleys in the first conduction band and one valley in the second conduction band with the valley parameters of the second conduction band chosen so as to give a good fit with the density of states as determined by numerical pseudopotential calculations [5,6]. Typically, when an effective mass approach is taken, no steps are taken to insure that the density of states of the first conduction band begins to decrease when the top of the band is approached. We have taken the density of states effective mass to be energy dependant so that the density of states does not extend infinitely. This method affects the scattering rate only by not allowing transitions to states at energies within valleys where no states exist. We have also added a second conduction band, with a single valley X2, that has valley parameters similar to the X valley in the first band. The energy of the X2 valley bottom was taken to be 2.2 eV above the bottom of the gamma valley. The conductivity effective mass was not altered

and is energy independent. Other than this change, we have used the same scattering rate calculations as Bhattacharyya et al [7,8]. Since the ACTFEL devices that are of interest to us are polycrystalline, the electric field is not taken in any specific crystal direction.

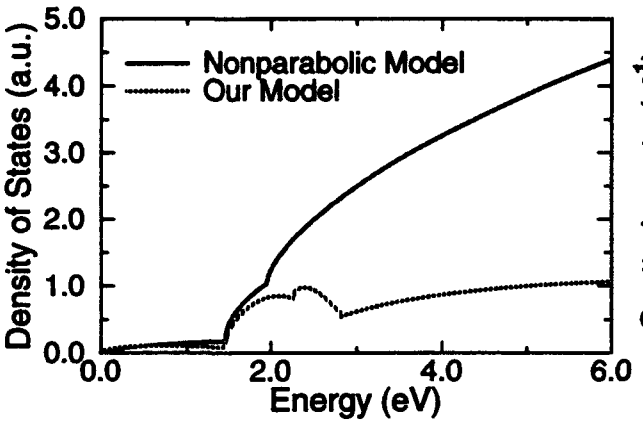


Fig. 1. The electron density of states in the conduction band of ZnS in both the simple infinite valley nonparabolic model, and our approach.

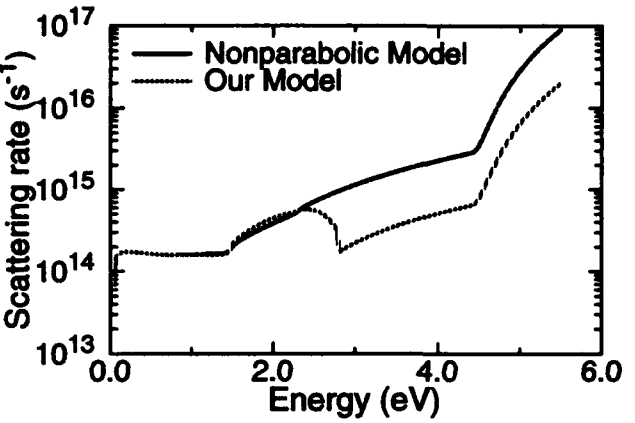


Fig. 2. The electron total scattering rate in the gamma valley for the simple infinite valley nonparabolic model, and our approach.

II. RESULTS

Fig 1. shows the density of states of the conduction band in our approximation as well as the simple, nonparabolic model used by Bhattacharyya. Our approximation more closely resembles the density of states computed by pseudopotential model than the simple infinite nonparabolic valleys yet it also requires much less computational time than that of Brennan's full band approach. Fig 2 shows the total scattering rate for the gamma valley as a function of energy. Because of the way that the density of states mass is defined, the scattering rates represent a sort of convolution of the scattering rate given by Bhattacharyya and the theoretical density of states. The simple infinite valley approach tends to overestimate the scattering rates at high energies, resulting in cooler distributions. For the gamma valley, polar optical scattering dominates at low energies. At 1.5 eV intervalley scattering to first conduction band valleys begins to dominate. When the electrons have acquired enough energy to scatter to the second band this process begins to dominate until the threshold for impact ionization is reached. Impact ionization is the dominant scattering mechanism for electrons in any valley once the total energy exceeds the threshold energy for impact ionization. Acoustic scattering is more important at high energies while ionized impurity scattering is more important at low energies but neither of these scattering mechanisms are ever dominant at any energy. For all of the upper valleys, intervalley scattering is the dominant scattering mechanism at all energies until the onset of impact ionization.

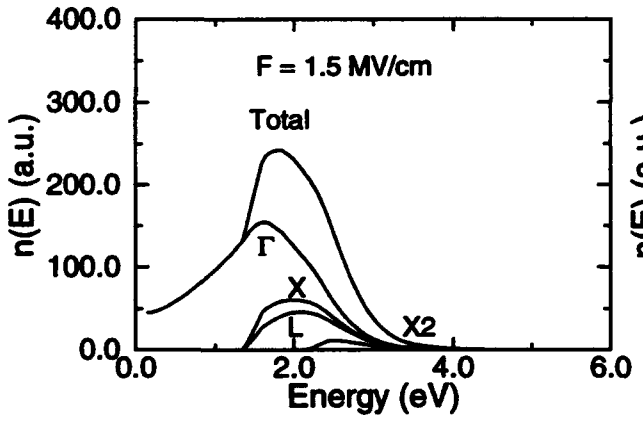


Fig. 3. Four valley, two band electron energy distribution for ZnS at 1.5 MV/cm in the simple infinite valley nonparabolic model.

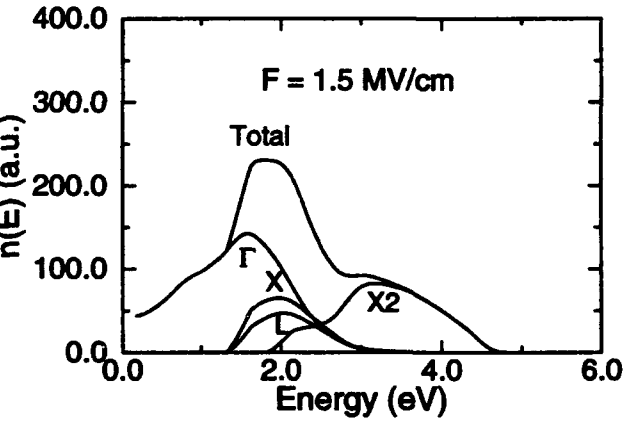


Fig. 4. Four valley, two band electron energy distribution for ZnS at 1.5 MV/cm in our model

Figs. 3 and 4 show the steady state electron energy distribution n(E) plotted as a function of energy for each valley in the first two conduction bands for an electric field of 1.5 MV/cm which is a typical field for an ACTFEL device [9]. Fig. 3 shows the distribution as it is calculated in the infinite valley approach and showing the cool distribution, while Fig. 4 shows the same distribution calculated in our approach showing an enhanced tail which extends past the threshold for impact ionization. At low fields, the total distribution is not that much different from the results of Bhattacharyya et al, the tail of the distribution begins to slope off at less than 4.0 eV. For higher fields, the results are quite new. We predict an enhanced tail and the formation of a secondary peak in the energy distribution at about 3.3 eV. The inclusion of the second conduction band, and lowering of the scattering rates allows the energy distribution to become hotter.

Impact ionization rates were calculated at several different fields, and Fig. 5 shows the impact ionization rate vs. the inverse field plotted on a semi-log scale to show the linear dependence. The data points are obtained from the simulation while the solid line is the impact ionization field dependence we have used to model ZnS ACTFEL devices as shown in Fig. 6. The impact ionization rate dependance supports the work of Shockley [10] which gives an ionization rate that depends on the field as

$$\alpha \sim \exp\left(\frac{B}{|E|}\right)$$

where E is the electric field strength and B is a constant. The results of the simulation determine B to be 3.95×10^6 V/cm

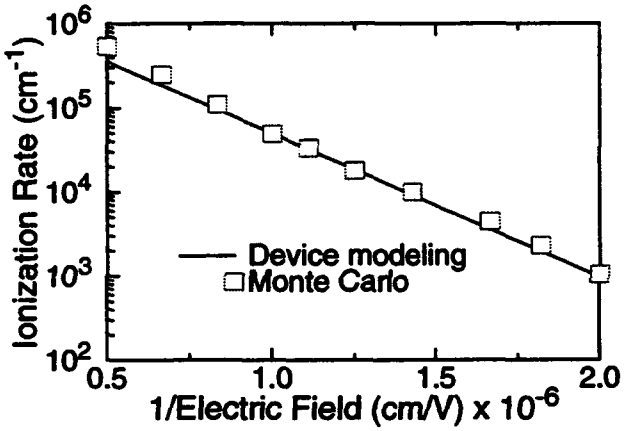


Fig. 5. Impact ionization rate for electrons in ZnS as a function of the inverse applied electric field.

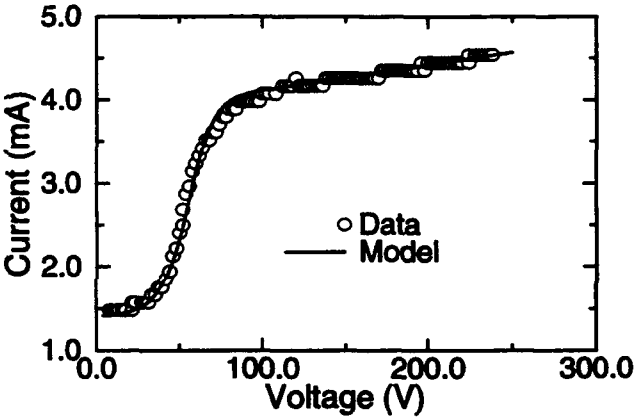


Fig. 6. I-V curve for a ZnS ACTFEL device showing both experimental data and device modeling based on the impact ionization rate derived from the simulation.

III. DISCUSSION

The results of the simulation clearly indicate that because of the high field heating of the electrons, the second conduction band can not be ignored. This simulation was not meant to include the full band structure in the calculations but rather a first approximation to the behavior of the electrons when a second conduction band is included. Just blindly adding a second band will not have much effect on the distribution though, since the infinitely extending valleys lead to an overestimation of the density of state in the first band which in turn leads to unrealistically high intervalley transfer into the first band. At low fields, the results of our simulation agree qualitatively with the results of Bhattacharyya et al [2] and this is evidence of the validity of the approximation in the limit of low fields. The simulation suggests that the second conduction band is sufficiently populated to support impact ionization rates that are high enough to drive the current voltage characteristics of a typical ZnS ACTFEL device. Previously it was thought [1,2,11] that impact ionization was at best a secondary source of electrons with the most important contributions coming from the tunneling of electrons from the interface states. Our simulation suggests that impact ionization is probably the dominant source of the electrons involved in the gain behavior of the devices. The work of Müller et al [3] and Mach [12] was done with a single parabolic valley and ballistic electron transport was reported. In our model, the single valley in the second conduction band acts in a similar manner, and if impact ionization is not included in the simulation, ballistic electron transport is also encountered

with an unstable energy distribution. We believe that the first conduction band supports a stable electron distribution due to the high scattering rates that intervalley scattering introduces. In the second conduction band on the other hand, the electrons in the X2 valley will impact ionize before they have enough energy to scatter to other second band valleys. Only electrons with low kinetic energy can scatter back into the first conduction band because there are no states available at higher energies in the first conduction band for them to transfer to. The first conduction band is still responsible for the supply of electrons that impact excite the luminescent centers since most of these have excitation energies of 2-3 eVs. At high enough fields however, the simulation suggests that there will be a second peak in the distribution. The secondary peak is in the second band and is located around 3.3 eV. This peak could be utilized to excite other luminescent centers if a proper dopant is found.

IV. CONCLUSIONS

Monte Carlo simulation of electrons in ZnS indicates the importance of the inclusion of the second conduction band. The electron population of the second band has a significant impact on the electron energy distribution as well as the impact ionization rate. Because the density of states of the first conduction band was limited so that electrons were not able to scatter into states with energy greater that the top of the band, there is a reduction in the scattering rate in the second band at about 3.0 eV. This reduction of scattering rates leads to a second peak in the conduction band which begins to show up for fields above 1.0 MV/cm. Results show that impact ionization is a reasonable candidate for a source of carrier multiplication in ZnS ACTFEL devices.

ACKNOWLEGEMENTS

This work is supported by ARPA under contract no. N00014-93-0312

REFERENCES

1. K. Brennan, J. Appl. Phys. **64**, 4024 (1988)
2. K. Bhattacharyya, S. M. Goodnick, J. F. Wager, J. Appl. Phys. **73**, 3390 (1993)
3. R. Mach and G. O. Müller, J. Crys.Growth **101**, 976 (1990)
4. R. Brunetti, C. Jacoboni, F. Venturi, E. Sangiorgi, and B. Riccò, Solid State Electron. **32**, 1663 (1989)
5. Landölt-Bornstein, New Series (Springer, Berlin,1987)
6. C. S. Wang and B. M. Klein, Phys Rev B 24, 3393 (1981)
7. C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, (Springer, Berlin,1989)
8. L. V. Keldysh, Sov. Phys JETP **21**, 1135 (1965)
9. W. Kong and R. Solanki, J. Appl Physics **75**, 3311 (1994)
10. W. Shockley, Solid State Electron. **2**, 35 (1961)
11. V. P. Singh, Q. Xu and S. Krishna, *Proceedings of the Sixth International Workshop on Electroluminescence*, El Paso TX, 1992
12. R. Mach, *Proceedings of the Sixth International Workshop on Electroluminescence*, El Paso TX, 1992

SIMULATION OF PERIODICALLY SEGMENTED WAVEGUIDES AS CONCURRENT BRAGG REFLECTORS AND QUASI-PHASE-MATCHED SECOND HARMONIC GENERATORS

Cangsang Zhao and Reinhart Engelmann
Department of Electrical Engineering and Applied Physics
Oregon Graduate Institute of Science & Technology
P. O. Box 91000, Portland, OR 97291-1000

Abstract

A simplified method has been developed to simulate the quasi-phase-matched (QPM) second harmonic generation (SHG) and distributed Bragg reflection (DBR) properties in a periodically segmented (PS) waveguide with graded index profiles for the purpose of designing a concurrent DBR and QPM SHG waveguide device.

I. INTRODUCTION

In the past several years QPM SHG has drawn considerable interest in the development of coherent blue or green light sources. Conversion efficiencies as high as 20% and blue light powers over 20 mW have been achieved in nonlinear crystals like lithium niobate ($LiNbO_3$) [1], lithium tantalate ($LiTaO_3$) [2], and potassium titanyl phosphate (KTP) [3]. Using semiconductor diode lasers as the sources in the QPM scheme has the potential of making compact coherent blue or green sources, and much work has been done to develop this kind of system in various crystals [2], [4], [5]. Either an optical isolator or an extended-cavity configuration using a bulk diffraction grating for feedback has to be used in the frequency doubling system to stabilize the diode laser sources, which will lead to a less compact and high cost device. A more attractive scheme of achieving coherent compact blue or green light sources is to make a concurrent QPM SHG and DBR waveguide [6], [7]. The waveguide itself then functions not only as the QPM SHG device but also as an extended-cavity mirror to stabilize the diode laser source. In the PS waveguide suitable for QPM SHG, in addition to the non-linear optical coefficient the linear optical constant (refractive index) is also periodically modulated along the propagation direction. This allows the waveguide to serve simultaneously as a distributed Bragg reflector. To achieve both QPM SHG and Bragg reflection at the same wavelength λ the QPM condition (order q):

$$(N_{\lambda/2} - N_{\lambda})\,\Lambda = \frac{q\lambda}{2} \qquad (q = 1, 2, \ldots, \infty) \tag{1}$$

and the DBR condition (order m):

$$N_{\lambda}\Lambda = \frac{m\lambda}{2} \qquad (m = 1, 2, \ldots, \infty) \tag{2}$$

have to be concurrently satisfied. Thus it is very critical to accurately estimate the mode indices $N_{\lambda/2}$ and N_{λ} of the periodically segmented structure in order to determine the periods Λ which satisfy both the conditions (1) and (2). It has been verified [8], [9] that in order to calculate in a PS step-index slab waveguide the effective mode index and the field distribution, one can safely replace the PS waveguide with a uniform waveguide such that its refractive index is equal to the weighted average of the high and low indices. This procedure can also be applied to a PS step-index channel waveguide [9]. But the intensity of the Bragg reflection cannot be easily derived in this approach unless a complicated numerical calculation is used [8], [9]. Also, the more realistic graded refractive index distributions in both vertical and lateral directions make it highly complicated or even impossible to find the effective mode indices and the Bragg reflection intensity in PS channel waveguides. In the following a simplified method will be presented to calculate the mode indices and the Bragg reflection intensity in order to design a device which can simultaneously satisfy DBR and QPM SHG conditions in a PS channel waveguide with more realistic graded-index distributions.

II. MODE INDEX CALCULATION CONSIDERATIONS

In principle, a PS waveguide can be viewed as a lens waveguide whose guided modes periodically diffract and refocus with negligible diffraction loss [10]. When a mode is propagating in the PS waveguide, it is locally guided or "refocused" in the high index sections, and unguided or "diffracted" in the substrate index sections and it is the guiding region which leads to the overall waveguiding. So it is reasonable to expect a homogeneous waveguide to have an equivalent guiding behavior as the PS waveguide [8], [9]. As shown in Figs. 1(a) and 1(b) for a PS step-index channel waveguide the effective mode index N can be calculated from an equivalent homogeneous step-index channel waveguide with weighted average of the high and low indices in the two sections [9]. In the high index sections the mode is guided locally as in a homogeneous step-index channel waveguide as shown in Fig. 1(c), and the local mode can be described by a mode index N_{high} calculated therefrom. But in the substrate index sections the mode is actually unguided and cannot be described by a mode index locally. In order to estimate the DBR reflectance in such a waveguide, an average index N_{low}, which describes the mode propagating speed in the substrate index sections, can be calculated from

$$N = DN_{high} + (1 - D)\,N_{low} \tag{3}$$

where D is the duty cycle of the PS waveguide. Thus the reflectance of the DBR can be calculated as in a periodically layered media with index N_{high} and N_{low} as shown in Fig. 1(d).

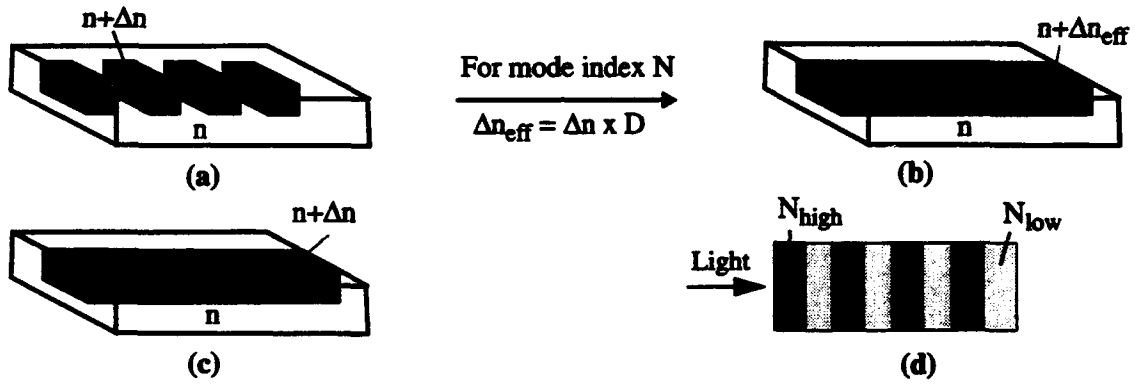


Figure 1 (a) Periodically segmented (PS) step-index channel waveguide; (b) Equivalent waveguide for calculating the mode index N; (c) Equivalent waveguide for calculating mode index N_{high}; (d) Equivalent layered media for calculating DBR properties.

For more realistic index distributions in a PS channel waveguide shown in Fig. 2(a), the mode index N_{high} can always be calculated by using the effective index method when the index profiles are given. As shown in Fig. 2(b), an equivalent step-index waveguide having the same local mode index N_{high} as in Fig. 2(a) can be found to simulate the guiding behavior in the high index sections. Thus the waveguide problem

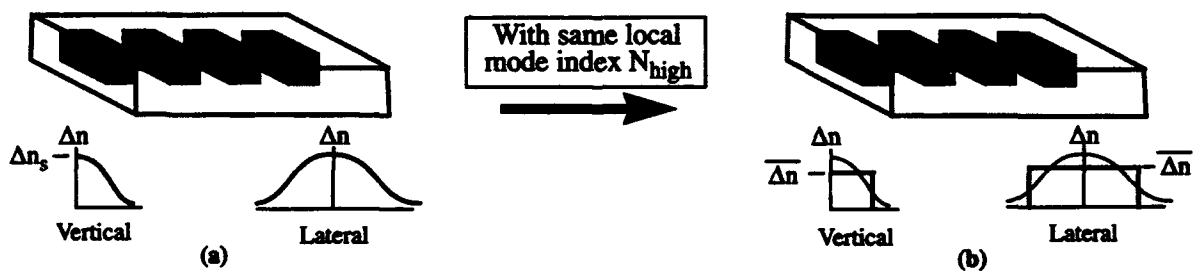


Figure 2 (a) Periodically segmented (PS) graded-index channel waveguide;
(b) Equivalent periodically segmented (PS) step-index channel waveguide.

reduces to a PS step-index waveguide problem, which has been solved previously. In waveguides as shown in Fig. 2(a) the index profiles are mostly Gaussian, squared hyperbolic tangent, or exponential functions, the mode index N_{high} can be calculated analytically (approximately, but accurate enough for mode index calculation). Thus the simulation of the concurrent DBR and QPM SHG device can be accomplished very rapidly and accurately.

III. RESULTS

For a KTP PS waveguide as shown in Fig. 2(a) exhibiting high-index sections with a step-index function in the lateral direction (w = 8 μm; a lateral step-index approximation is good enough when w is not too small), a squared hyperbolic tangent index function in the vertical direction (characteristic depth d = 3 μm, surface index difference $\Delta n_s = 0.016$), and a duty cycle D = 0.52, the equivalent PS step-index channel waveguide has been found with a uniform index difference $\overline{\Delta n} = 0.012$. From this result the concurrent DBR and QPM grating periods and fundamental wavelengths as shown in Fig. 3(a) have been calculated for the first order (q=1) QPM condition and various order DBR conditions. In Rb/Ba-diffused waveguides in KTP, temperature tuning allows one to shift the relative position of λ_{DBR} and λ_{QPM} by 0.051 nm/°C [11]. Consequently if one is constrained to temperature tuning of the KTP waveguide by say ± 10 °C, then, at the center of the temperature band, λ_{DBR} and λ_{QPM} should match within some 0.51 nm. This corresponds to a fabrication allowance of ± 0.005 μm deviation of the grating period from 4.202 μm according to the enlarged plot shown in Fig. 3(b) for m=18 and q=1. At the intersection point, the calculated material index n is 1.840, and the effective mode index N is 1.842, N_{high} and N_{low} are 1.847 and 1.837 respectively, which are used to calculate the DBR reflectance as shown in Fig. 3(c). An interesting behavior of the waveguide especially from a device design point of view is shown in Fig. 3(d). By choosing three different $\Delta n_s = 0.01$, 0.015, and 0.02, and varying the duty cycle from 0.5 to 0.75, one can choose the preferable wavelengths in a relatively wide range of about 10 nm. The refractive index data used above is obtained from previously reported index measurements on hydrothermally grown KTP [12].

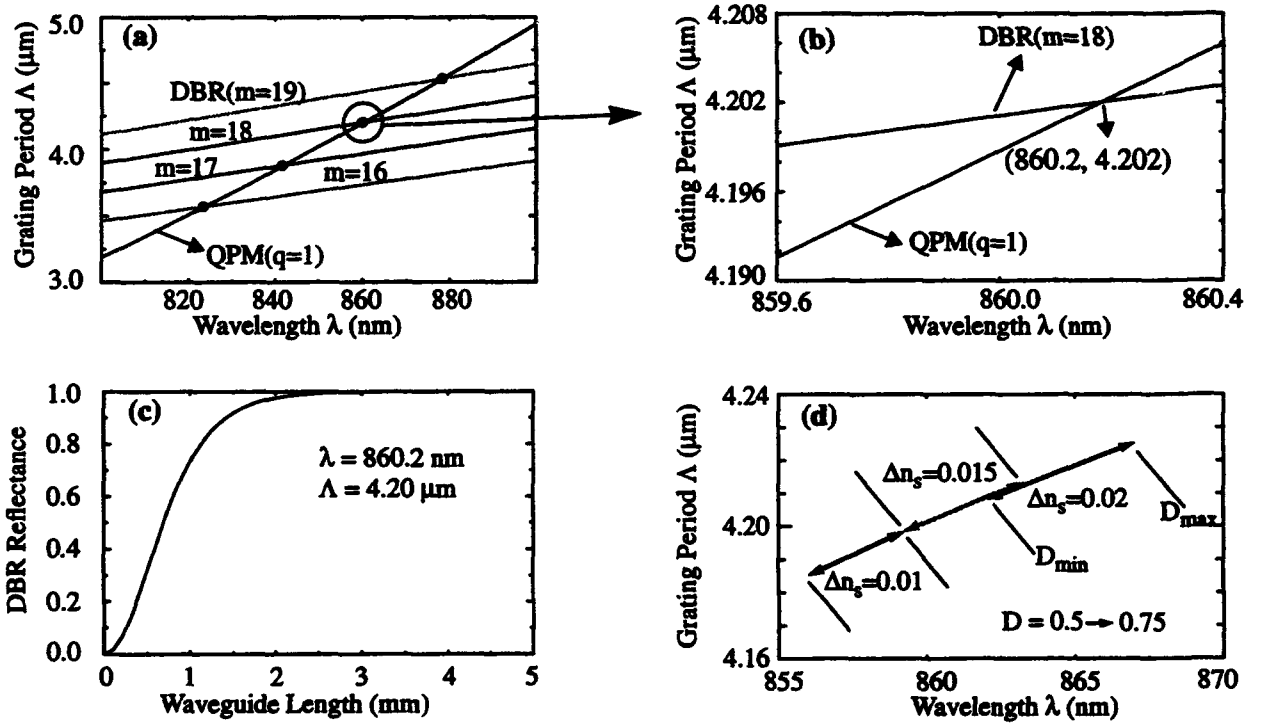


Figure 3 (a) Concurrent DBR and QPM grating periods and wavelengths in a KTP PS waveguide; (b) Enlarged plot of (a) at a certain cross-over point (q=1, m=18); (c) DBR reflectance vs. waveguide length at the concurrent DBR and QPM point of (b); (d) Concurrent DBR and QPM grating periods and wavelengths with various Δn_s and D.

IV. SUMMARY AND CONCLUSIONS

We have presented a simplified method to calculate the mode indices in a periodically segmented graded-index channel waveguide for the practical design of a concurrent DBR and QPM SHG device. The method has been used to model a simple device structure in a KTP crystal in which the fabrication tolerance is extremely tight. Similarly, the method can be used to model the superperiod structure which has been proposed for loosening the fabrication tolerance of a concurrent DBR and QPM SHG waveguide in KTP [7]. We intend to use the method to simulate the DBR and QMP SHG properties for designing a more efficient device structure with practically achievable fabrication tolerances.

ACKNOWLEDGMENT

This work was supported in part by a Motorola-University Partnership in Research Grant.

REFERENCES:

[1] M. Yamada, N. Nada, M. Saitoh, and K. Watanabe, Appl. Phys. Lett. **62**(5), 435-436 (1993).
[2] K Yamamoto, K. Mizuuchi, Y. Kitaoka, and M. Kato, Appl. Phys. Lett. **62**(21), 2599-2601 (1993).
[3] D. Eger, M. Oron, M. Katz, and A. Zussman, Compact Blue-Green Laser Topical Meeting, OSA, New Orleans, 3-4 Feb. 1993.
[4] X. Cao, R. Srivastava, and R. V. Ramaswamy, Optics Lett. **17**(8), 592-594 (1992).
[5] W. P. Risk, W. J. Kozlovsky, S. D. Lau, G. L. Bona, H. Jaeckel, and D. J. Webb, Appl. Phys. Lett. **63**(23), 3134-3136 (1993).
[6] K. Shinozaki, T. Fukunaga, K. Watanabe, and T. Kamijoh, Appl. Phys. Lett. **59**(5), 510-512 (1991).
[7] M. G. Roelofs, W. Bindloss, A. Suna, J. D. Lee, and J. D. Bierlein, OSA, Compact Blue-Green Laser Topical Meeting, New Orleans, 3-4 Feb. 1993.
[8] L. Li and J. J. Burke, Opt. Lett., **17**(17), 1195-1197 (1992).
[9] Z. Weissman and A. Hardy, J. Lightwave Tech., **11**(11), 1831-1838 (1993).
[10] D. Marcuse, *Light Transmission Optics*. New York: Holt, Reinhart and Winston, 1985, Ch. 5.
[11] W. P. Risk, and S. D. Lau, Optics Lett., **18**(4), 272-274 (1993).
[12] J. J. Bierlein, and H. Vanherzeele, J. Opt. Soc. Am. **B6**(4), 622-633 (1989).

NUMERICAL MODELING OF INJECTION INDUCED CARRIER CONFINEMENT IN QUANTUM-WELL (QW) STRUCTURES

Yijun Cai and Reinhart Engelmann
Department of Electrical Engineering and Applied Physics
Oregon Graduate Institute of Science & Technology
Portland, OR 97291-1000

Abstract

A comprehensive band filling model including the injection induced carrier confinement effect has been developed for device simulation of QW lasers, based on self-consistently solving the Schroedinger and Poisson equations. A simplified approach has also been derived for quick evaluation in gain calculation.

I. INTRODUCTION

The Separate Confinement Heterostructure (SCH) Single Quantum Well (SQW) laser is a commonly used structure to achieve low threshold current density operation. In such a structure, as shown in Fig.1, the carriers (electrons and holes) are mainly confined by the band offsets ΔE_c and ΔE_v between QW and barriers (optical guiding layers). In real laser structures, ΔE_c and ΔE_v are often less than 0.2 eV. As the injection level becomes increasingly higher, quasi-Fermi levels are raised and carrier distributions in the barriers will increase. This carrier spill-over may becomes significant, especially in diode lasers operating at high temperature or in laser structures with low band offsets (poor carrier confinement). Generally, due to the structural difference between conduction and valence band, (e.g. electron-to-hole effective mass ratio is about 1/6 in ZnCdSe/ZnSSe based lasers), the spill-over of electrons and holes tend to be different. This will result in an internal electrostatic field which works to reduce the difference of the electron and hole distributions, leading to the modification of the band profile.

An extreme case of this kind of injection-induced carrier confinement can be found in some laser structures where one of the conduction or valence band is completely flat or even slightly type-II. In such a case, it is still possible to get some carrier localization due to the electrostatic attraction generated by space charges as demonstrated in reference[1]. In some more general cases, even with structures that have type-I band alignment, the threshold carrier density for lasing may be very high and the carrier confinement is relatively poor, as in most wide bandgap II-VI compounds based lasers. It is still necessary to evaluate this injection-induced electrostatic confinement effect in gain/ threshold current calculations for device design and optimization. Also, the quantitative modeling of this effect can

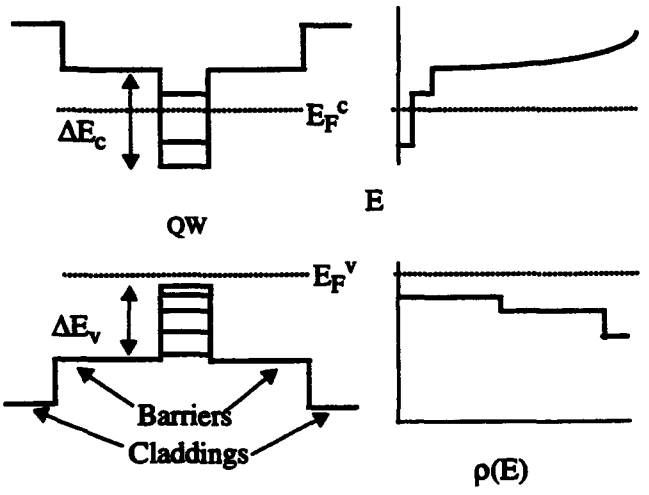


Fig.1 SCH SQW structure and density of states.

help us gain more understanding about the influence on device performance of band offset ratio $Q_c = \Delta E_c / (\Delta E_c + \Delta E_v)$, whose value for the time being is still very difficult to predict. In this paper we will present a comprehensive band filling model for a SCH SQW laser structure which including the spill-over carriers contribution.

The numerical method is based on solving the coupled set of Poisson and Schroedinger equations. An accurate picture of carrier distributions can be obtained from this comprehensive model. Based on these we then developed a simplified approach for quick evaluation of this effect in gain/threshold-current calculation through band offset ratio adjustment. The controversy on band offset ratios can possibly be clarified in this way. As an example, the numerical simulation results on a ZnCdSe/ZnSSe based laser will be presented.

II. BAND FILLING MODEL

Flat quasi-Fermi levels across the hetero-junction between QW and barrier are assumed and the carrier densities in QW and barrier regions are obtained under Fermi-Dirac distributions. The two-dimensional carrier density distribution in the z-direction can be described by

$$n(z) = k_B T \sum_i |F_{e,i}(z)|^2 \times \rho_{c,i} \times ln\,(1 + e^{(E_F^c - E_{e,i})/(k_B T)}),$$
$$p(z) = k_B T \sum_i |F_{h,i}(z)|^2 \times \rho_{v,i} \times ln\,(1 + e^{(E_F^v - E_{h,i})/(k_B T)}) \qquad (1)$$

where k_B is Boltzmann constant and T is the temperature. E_F^c and E_F^v are the quasi-Fermi levels in conduction and valence band, respectively. $E_{e,i}$ and $E_{h,i}$ are the quantized energy levels, and $\rho_{c,i}$ and $\rho_{v,i}$ are the density-of-states functions of electrons and holes of the i-th subband. Here we also include those unconfined (spill-over) states in the barrier with z-direction energy higher than the band off set ΔE_c and ΔE_v. The envelope functions $F_{e,i}(z)$ of electrons and $F_{h,i}(z)$ of holes are described by Schroedinger equations:

$$\left(-\frac{h^2}{2m_e}\frac{d^2}{dz^2} + U_e(z) + V(z)\right)F_{e,i}(z) = E_{e,i}F_{e,i}(z),$$
$$\left(-\frac{h^2}{2m_h}\frac{d^2}{dz^2} + U_h(z) - V(z)\right)F_{h,i}(z) = E_{h,i}F_{h,i}(z) \qquad (2)$$

For simplicity the effective masses m_e and m_h are considered constant. $U_e(z)$ and $U_h(z)$ are the conduction and valence band-edge potentials. V(z) is the electrostatic potential generated by space charge due to the imbalance of electron and hole spatial distribution, which can be described by Poisson's equation:

$$\frac{d^2}{dz}V(z) = \frac{e^2}{\varepsilon}[n(z) - p(z)] \qquad (3)$$

The numerical solutions of the above problem are similar to those applied in HEMT (high electron mobility transistor) simulations, except now both electrons and holes need to be considered at the same time. The solution of Poisson's equation is based on a standard finite difference method. The solution of Schroedinger's equation follows the Numerov process as described by P. C. Chow[2]. The band profile and carrier distributions are then obtained by self-consistent iteration procedures. Based on the detailed simulations, a simplified approach has also been developed for easy incorporation into gain calculation.

III. SIMPLIFIED APPROACH FOR GAIN MODELING

In practice, it is important to have a flexible modeling program available that allows to elucidate trends in a large parameter space rather than a sophisticated procedure that generates highly accurate data for specialized situations. The injection-induced confinement effect can be incorporated into gain modeling in a simplified way. The basic feature of this approach assumes that the injection induced space charge distribution is an interface dipole sheet, and this dipole sheet only changes the built-in potential of the hetero-junction. In other words, we use the band offset Q_c as an adjustable parameter to fulfill the charge neutrality condition in both QW and barrier regions. Thus, an injection level dependent Q_c which makes the carrier densities of electrons and holes in both QW and barrier almost equal, can be derived. Based on this, we can then calculate the gain spectrum and the radiative current including the influence of barrier recombination, in a self-consistent manner. The gain/current calculations are based on the model introduced in Ref.[3] except in our case we also include the barrier recombinations.

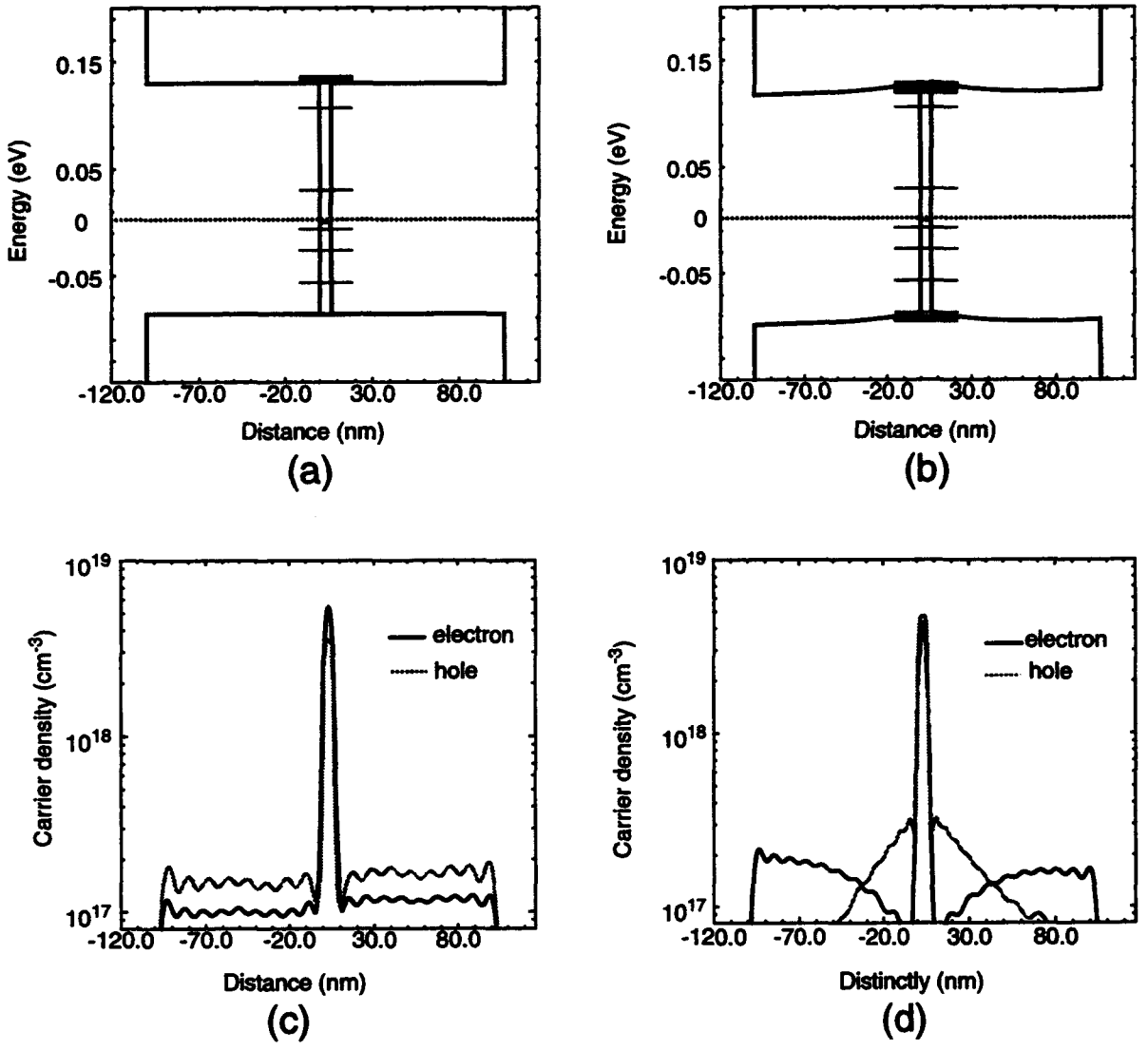


Fig.2 Band profile (without QW band gap) and carrier distribution. (a) Initial band diagram (b) Band diagram including injection induced carrier confinement (c) Unbalanced carrier distribution of configuration (a). (d) balanced carrier distribution of configuration (b).

IV. RESULTS

As an example, we consider a particular device structure (SCH SQW $Cd_{0.2}Zn_{0.8}Se/ZnS_{0.06}Se_{0.94}$ 6.5 nm QW, 100 nm barrier) to perform our numerical calculations. The injection levels is about $5x10^{18}$ cm^{-3} which is close to the threshold carrier density measured in the experiments. The initial band configuration is shown in Fig.2a, as estimated from strain effect and common anion rule. The Q_c value (refers to heavy hole band-edge) in this case is about 0.62. As we keep the total charge neutrality in the combined barrier and QW region, the locally unbalanced carrier distribution at this injection level is shown in Fig.2c, where in this particular case there are more electrons in QW and more holes in barriers. The modification of the band profile due to the locally unbalanced carrier distribution is obtained with the comprehensive model as shown in Fig.2b. The injection induced carrier confinement can be clearly observed when compared to the balance carrier distribution shown in Fig.2d.

Next we use our simplified approach to evaluate gain/threshold current relation. Here Q_c is changed to provide for the same quasi-Fermi levels as those in comprehensive model. Once this condition is satisfied, Q_c = 0.53 is obtained and the carrier distribution are found to be very close to those calculated in the comprehensive model and the charge neutrality condition in both QW and barrier region are then satisfied as shown in Fig.3. The gain and threshold current density is as calculated for Q_c = 0.53 is plotted in Fig.4. The results are very close to the value of 500 A cm^{-2} which has been observed experimentally.

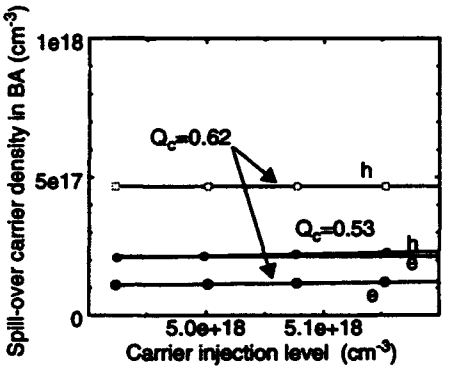


Fig. 3 Carrier spill-over under different Q_c

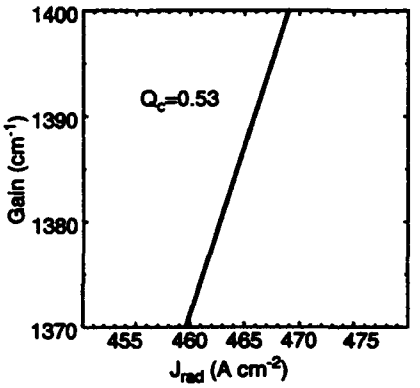


Fig. 4 Gain vs. radiative current density.

V. CONCLUSIONS

In summary, we have developed a comprehensive theoretical approach to analyze the injection induced carrier confinement and its influence on band filling process of SCH SQW lasers. A simplified approach has also been derived for quick evaluation. It has been shown that in modeling of ZnCdSe/ZnSSe based diode laser, the method can give satisfied explanation of observed results.

Acknowledgment: Supported in part by ARPA/ONR though a subcontract with the University of Florida. A HEMT simulator (SPS) courtesy by Claus Fischer, Technical University Vienna, Austria.

REFERENCE

[1]. J. Barrau, B. Brousseau, M. Brousseau, R. J. Simes and L. Goldstein, Electron. Lett. vol. 28 (8), 786, 9 Apr., 1992.

[2]. P. C. Chow, Americ. J. of Phys., vol. 40, 730-734, May 1972.

[3]. S. W. Corzine, R. H. Yan and L. A. Coldren, in *Quantum Well Lasers*, P. S. Zory, Ed. Chap.1, Academic Press, Boston, 1993.

A Self-Consistent Simulation of the Modulation Response of Quantum-Well Lasers Including the Effect of Strain

R. Baca, M. Grupen, and K. Hess

Beckman Institute for Advanced Science and Technology and
Center for Compound Semiconductor Microelectronics and
University of Illinois at Urbana-Champaign, Urbana, IL 61801

A self-consistent simulation of a InGaAsP quantum-well laser is conducted to study the effect strain has on the modulation response. This study is an extension of the work done in [1] where the effect of carrier capture on the modulation response was studied. There are a number of factors determining the modulation response of a quantum well laser. One must consider the majority carrier drift outside the quantum well which produces the low frequency roll-off in the modulation response. In addition, minority carrier diffusion and ballistic emission into the quantum well need to be included. These transport processes are also affected by the carrier capture in the quantum well which sensitively influences the modulation response. The difficulty in the simulation lies in coupling the transport across the classical portion of the device (i.e. separate confinement and optical confinement regions) with the quantum well region where size quantization is considered. This was done by treating the classical regions with drift-diffusion and thermionic emission theories and coupling them to the quantum well regions by including the relaxation of injected carriers into bound states.

Strain is incorporated into our existing model by approximating the nonparabolic valence bands as being anisotropic but parabolic parallel and perpendicular to the growth plane [2]. By including the corresponding effective masses , $m_{\perp}$ and $m_{\parallel}$, in the appropriate equations we present the effect strain has on the modulation response. We also study how the low frequency roll-off is affected as the capture and emission rates are varied while including strain.

References

[1] M. Grupen, G. Kosinovsky, and K. Hess,"The Effect of Carrier Capture on the Modulation Bandwidth of Quantum Well Lasers" IEDM Technical Digest, Washington D.C., Dec. 1993, pp 609-612.

[2] Z.-M. Li, M. Davies, M. Dion, and S. P. McAlister, "A versatile two-dimensional model for InGaAsP quantum-well semiconductor lasers," Can. J. of Phys., vol. 70, no. 10, pp. 937-942, 1992.

ELECTRON-HOLE IMBALANCE IN THE ACTIVE REGION OF QW LASERS, AND ITS EFFECT ON THE THRESHOLD CURRENT.

G. A. Kosinovsky, M. Grupen, K. Hess
Beckman Institute, University of Illinois, Urbana, IL 61801

Abstract

This study predicts that the electron/hole density ratio in thin quantum wells (QW's) of GaAs/AlGaAs laser diodes with intrinsic QW active regions (QW PIN lasers) can be significantly different from unity and depends on the doping density near the active region. These deviations from local charge neutrality can have significant effects on the laser threshold.

I. INTRODUCTION.

Macroscopic 1-D studies of threshold currents in QW lasers, based on the gain threshold condition and radiative rate equations, have frequently been performed (Ref.[1-4]). Later, efforts have been made to formulate self-consistent 2-D simulators for diode lasers (Ref[5-7]). These simulators are more precise for studying semiconductor lasers and obtaining quantitatively meaningful data on laser performance. Here we describe a particular aspect of the physics of threshold current in ideal lasers.

In macroscopic 1-D investigations of the threshold current dependence on QW width (Ref[2-4]), the balance of mobile charge (n=p), i.e local charge neutrality, has been used as one of the constraints in the quantum well regions of PIN diodes. This is an accurate assumption for any sizable active region, since in normal device operation Poisson's equation does not permit a large build up of net charge. However, for the dimensions of quantum wells this is not strictly valid, since the local charge neutrality is not necessarily preserved. It is difficult to incorporate such charge imbalance into rate equation based simulators without making arbitrary assumptions. We show here that the deviation from charge neutrality follows naturally in the solution of our self-consistent simulator MINILASE (Ref.[6,8]). We also show that this has important consequences for laser threshold currents.

II. INVESTIGATION OF THE EFFECT OF CHARGE IMBALANCE USING A RATE EQUATION MODEL.

We derive the effect on the predicted threshold current that is obtained from a simple rate equation simulator when the n/p ratio is varied, rather than set to 1. We have performed a simulation based on solving the laser gain equation and used the fact that spontaneous emission dominates the diode laser current value at the onset of stimulated emission (Ref[2-4]). We took the quantum well width to be 50 $\AA$, and the distributed loss factor of $5200\mathrm{m}^{-1}$, which was taken for consistency with MINILASE simulations, described in more detail later. The relationship between threshold current and the n/p ratio is shown on Fig.1. We see that for this QW structure the nominal threshold current density increases monotonically with the increase in the n/p ratio. The reason for the variation of threshold current with the k=n/p ratio follows logically from the underlying physical model. Consider the following

equation for the rate of stimulated emission (r_{th}^{st}) at threshold:

$$r_{th}^{st} = \sum_i B(E_c^{\nu_0,i}, E_v^{\nu_0,i}) \cdot g(E_c^{\nu_0,i} - E_v^{\nu_0,i}) \cdot (f_e(F_c^{th}, E_c^{\nu_0,i}) + f_h(F_v^{th}, E_v^{\nu_0,i}) - 1), \qquad (1)$$

which is used to calculate the quasi-Fermi levels F_c^{th} and F_v^{th} at threshold. Here B is the Einstein coefficient, g is the reduced density of states (assuming no line broadening), f_e, f_h are the Fermi functions for electrons and holes respectively, and $E_c^{\nu_0,i}$, $E_v^{\nu_0,i}$ are the conduction and valence band levels (with respect to minimum of the i-th subband) contributing to lasing mode ν_0. Our calculations for a 50 $\AA$ well show that (for the given range of k) ν_0 is always the lowest allowable optical mode. Therefore, g is constant and has non-zero value only for i=1, B is constant, and f_e, f_h are functions of F_c^{th} and F_v^{th} only. Hence equation (1) reduces to

$$r_{th}^{st} = A \cdot (f_e(F_c^{th}) + f_h(F_v^{th}) - 1), \qquad (2)$$

were A is a known factor. Since r_{th}^{st} is fully determined by the gain threshold value $G_{th} = 5200m^{-1}$ (see discussion above), $f_e(F_c^{th})$ and $f_h(F_v^{th})$ must vary by equal and opposite amounts as k varies. Because of the effective mass disparity in GaAs (and most other materials), the slope of $f_e(F_c)$ at F_c^{th} is usually small, while the slope of $f_h(F_v)$ at F_v^{th} is large. For example, $\frac{\partial f_h}{\partial F_v^{th}} \approx 3.7 \cdot \frac{\partial f_e}{\partial F_c^{th}}$ for k=.6, and $\frac{\partial f_h}{\partial F_v^{th}} \approx 44 \cdot \frac{\partial f_e}{\partial F_c^{th}}$ for k=1.5. Therefore, as k=n/p increases, the increase in F_c^{th} is much greater than the decrease in F_v^{th}. Since the calculation of the threshold current involves the summation of terms including $f_e(F_c^{th}, E_c^{\nu,i}) \times f_h(F_v^{th}, E_v^{\nu,i})$ over all optical modes ν and subbands i (see Ref.[1]), it is clear that the large increase in F_c^{th} outweighs the much smaller decrease in F_v^{th}, and the threshold current will increase with k.

III. MINILASE SIMULATIONS.

In order to see whether the n-p imbalance and its effects on threshold current are physically meaningful we turned to the self-consistent 2-D simulator MINILASE, originally described in a previous paper (REF[6]). It consists primarily of the coupled discretized solution of Poisson's equation and the electron and hole current continuity equations, iterated with the photon mode rate equations. The 2-D Helmholtz equation is also solved to determine the transverse intensity profile of the lasing mode(s) (Ref[9]).

The system of the continuity and Poisson's equations is solved by the Newton iteration on its Jacobian. The solution variables of this system are the electrostatic potential and the electron and hole quasi-Fermi levels. In this formalism, there is no rigid local charge neutrality constraint. The physical charge neutrality constraint is globally enforced through Poisson's equation. Since the original publication (Ref[6]), one of the key changes made has been the addition of the Schroedinger Equation for the QW active region, solved iteratively with the continuity-Poisson Newton system. Considering the true quantum nature of the active region, this addition was critical for investigation of any physical effects related to the electronic properties of the active region, notably in our case, the charge distribution and the radiative recombination. For more details see Ref.[8]

The structure considered for our example is the quasi-one dimensional buried Separate Confinement Heterostructure (SCH) laser. This structure has a total width of 3μm and has symmetric material structure. There is a 50$\AA$ $GaAs$ QW active region in the middle, then 975$\AA$ $Al_{0.4}Ga_{0.6}As$ light guiding regions to each side, followed by 1.4μm $Al_{0.65}Ga_{0.35}As$ regions bounded by electrodes. The doping profiles on this structure were varied in order

to achieve the n-p imbalance and investigate the resulting effects. We investigated three different structures. Structure A had the 1.4μm region under the top (bias) electrode p-doped at $5.0\times10^{18}cm^{-3}$ and the 1.4μm region above the bottom (ground) electrode n-doped at $5.0\times10^{18}cm^{-3}$. The waveguide region was kept intrinsic. Structure B had the top 1.4μm region and the adjacent half of the waveguide region p-doped at $3.0\times10^{18}cm^{-3}$, while the bottom 1.4μm region was n-doped at $2.0\times10^{18}cm^{-3}$. The bottom half of the waveguide region and the active region were kept intrinsic. Structure C had the top 1.4μm region p-doped at $3.0\times10^{18}cm^{-3}$, while the bottom 1.4μm region and the bottom half of the waveguide region was n-doped at $3.0\times10^{18}cm^{-3}$. The top half of the waveguide region and the active region were kept intrinsic. The doping concentrations were chosen not only to optimize the n-p imbalance in structures B and C, but also to achieve approximate equality of the gain threshold (loss factor) among the three structures. Table I shows the results.

Notice that the threshold current versus k=n/p ratio dependence for these structures follows qualitatively the trend suggested by the simplified calculation that led to Fig.1. Quantitatively, there is about a 25% difference between the threshold current values obtained from MINILASE and listed in Table I and the corresponding points of Fig.1. This difference is acceptable, considering the simplicity of the rate equation model and, in fact, underscores the importance of the self-consistent simulators for accurate quantitative analysis.

Comparing the MINILASE data of Table I for different structures, we note that the lasing threshold current difference between structure B (k=0.66) and structure C (k=1.47) is a very considerable 49.4%. More importantly, we note the difference between structure A, representing a "conventional" PIN laser with n-p neutrality preserved, and structure B, where the heavy p-doping up to the quantum well results in a much higher concentration of holes than electrons in the well. We can see that decreasing the n/p ratio from 1 to 0.66 results in a 15.6% lowering of the threshold current value, which is a significant improvement.

IV. CONCLUSION.

We have shown that, due to a large hole-electron effective mass difference in GaAs, the lowering of the electron concentration in the quantum well relative to that of the holes improves laser threshold performance. This lowering may be achieved by appropriate modulation doping.

We acknowledge financial support from the National Science Foundation through NCCE and from the Office of Naval Research.

References

[1] K. Hess, B. A Vojak, N. Holonyak, Jr., R. Chin and P. D. Dapkus, Solid-St. Electron., **23**, 585 (1980)

[2] Y. Arakawa and A. Yariv, IEEE J. Quantum Electron. **QE-21**, 1666 (1985)

[3] Akira Sugimura, IEEE J. Quantum Electron. **QE-20**, 336 (1984)

[4] S. R. Chinn, P. S. Zory and A. R. Reisinger, IEEE J. Quantum Electron. **24**, 2191 (1988)

[5] D. P. Wilt and A.Yariv, IEEE J. Quantum Electron. **QE-17**, 1941 (1981)

[6] G. H. Song, K. Hess, T. Kerkhoven and U.Ravaioli, European Trans. Telecom. Related Tech. 1, 375 (1990)

[7] S. Seki, M. Tomizawa, K. Yokoyama and Y. Yoshii, Tech. Dig. Int. Electron Device Meetings (1988)

[8] M. Grupen and K. Hess, Proceed. SPIE's OE/LASE Conference (1994)

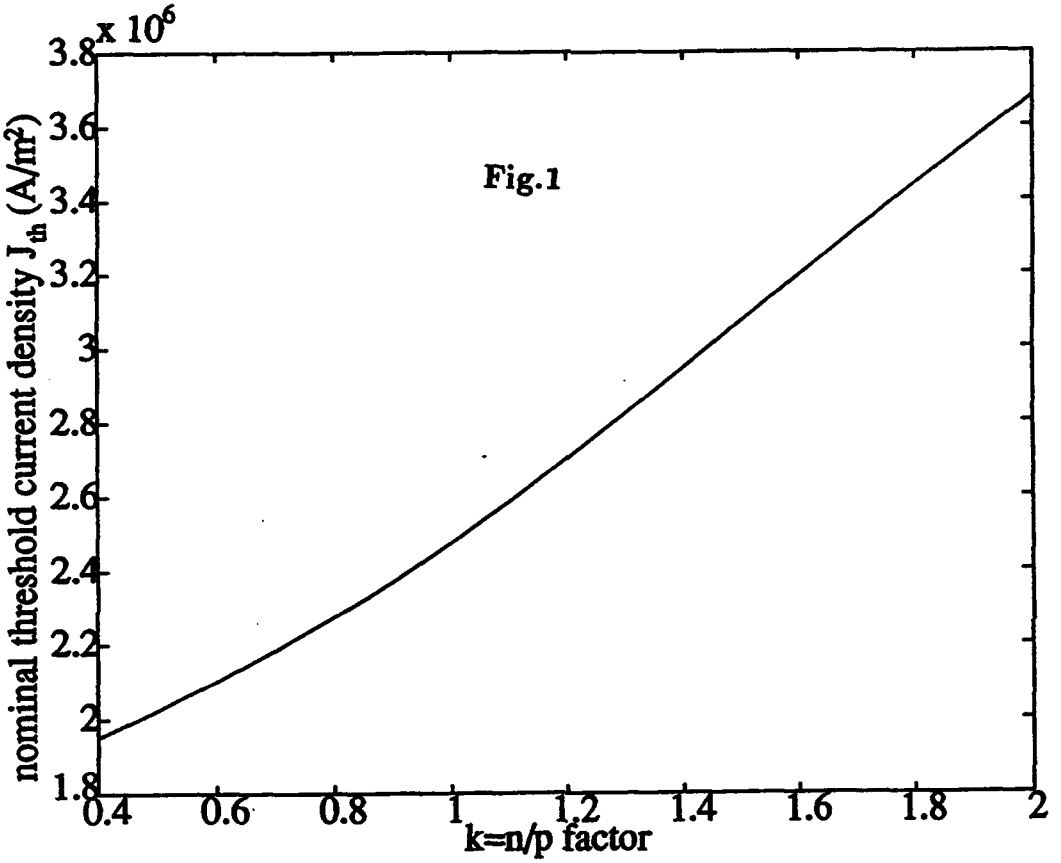


Fig.1

Table I

Structure	k=n/p	$J_{th}(A/m^2)$
A	1.05	$3.38*10^6$
B	0.66	$2.78*10^6$
C	1.47	$4.16*10^6$

Simulation results for structures A, B, and C

A NUMERICAL MODEL FOR COMPUTING THE EMISSION SPECTRUM IN TIME-RESOLVED PHOTOLUMINESCENCE EXPERIMENTS

Steven M. Durbin[1], Dean H. Levi[2] and Jeffery L. Gray[1]
[1]*School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-1285*
[2]*National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401*

Abstract

A detailed numerical model is presented for computing the luminescence spectrum in a semiconductor layer which properly accounts for the effects of photon recycling. The resulting model is incorporated into a 1-D drift-diffusion simulation package, and used to simulate time-resolved photoluminescence (TRPL) experiments on AlGaAs/GaAs/AlGaAs double heterostructures in order to extract carrier recombination and transport information. Results of initial simulations are compared with measured spectra to verify the accuracy of the model.

I. INTRODUCTION

A common method for determining minority carrier lifetimes in compound semiconductors is time-resolved photoluminescence (TRPL), which measures the decay of photoluminescence intensity in response to pulsed laser excitation. The effects of re-absorption of photons emitted via radiative recombination (commonly referred to as "photon recycling") have been shown to significantly affect the measured lifetime in direct-gap semiconductors such as GaAs when radiative recombination is dominant or nearly dominant [1]. In order to permit the investigation of the physical processes underlying the experimental technique, a detailed numerical model for the emission intensity has been incorporated into a complete 1-D numerical semiconductor simulation package developed at Purdue University. Simulations of AlGaAs/GaAs/AlGaAs double heterostructures (DHs) are performed in order to explore the sensitivity of the measurement to various device and material parameters.

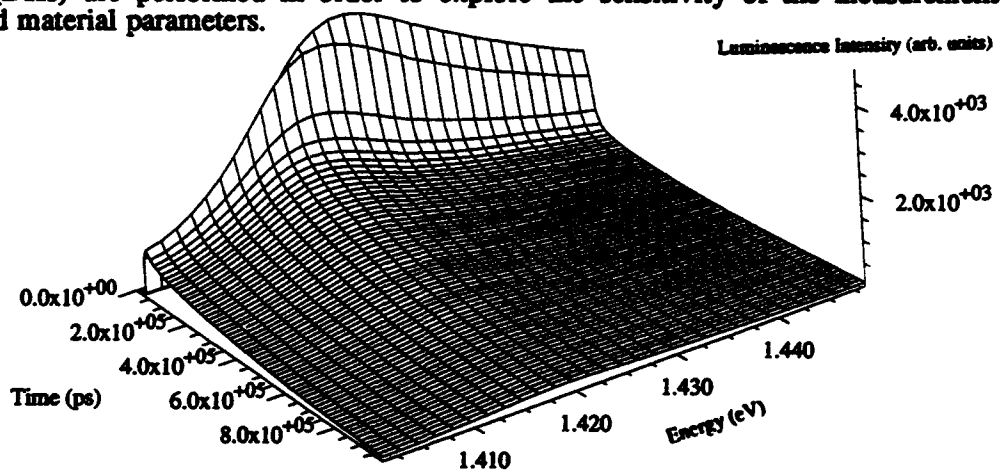


Fig. 1: Experimental time-resolved photoluminescence spectrum of a 10 μm DH.

Because the near-bandedge absorption coefficient increases with photon energy, self-absorption will shift the externally observed emission peak towards lower energies. Thus, as carriers diffuse away from the front surface after the excitation pulse, an increasing fraction of the higher-energy portion of the luminescence will be reabsorbed, resulting in a red shift of the observed spectrum with time. Such an effect has been observed experimentally in AlGaAs/GaAs/AlGaAs double heterostructures (Fig. 1), and simulation results predict similar behavior. Although TRPL is typically used to extract minority carrier lifetimes, the temporal behavior of the peak energy can provide information about the minority carrier diffusivity as well. Thus, an accurate numerical model for TRPL can provide insight into both carrier recombination and transport.

II. NUMERICAL APPROACH

A. Solution of Semiconductor Equations

A one-dimensional finite difference/ finite box discretization technique is applied to the self-consistent solution of the Poisson equation and the electron and hole current continuity equations. Newton-Raphson iteration is employed to solve the resulting set of coupled nonlinear partial differential equations for the electrostatic potential as well as hole and electron concentrations at discrete mesh points within the modeled device. The self-absorption of photons emitted through radiative recombination ("photon recycling") is accounted for by calculation of an effective generation rate after the method of Kuriyama et. al. [2], which is introduced through the generation term in both current continuity equations [3].

B. Numerical Emission Intensity Model

The expression for luminescence intensity begins with assuming rotational symmetry in a cylindrical geometry (r, ϕ, z) having infinite transverse dimensions; the device geometry is taken to vary along the z axis. Assuming infinite transverse dimensions will not affect the accuracy provided that the physical dimensions of the device being modeled are greater than the inverse absorption coefficient corresponding to the emitted photon energy. Photons are assumed to be emitted via radiative recombination throughout the 3-D cylindrical volume, and incident upon a line through the origin (r=0, ϕ, z). A 1-D expression is obtained by assuming that there is no transverse variation in physical variables (i.e. p, n, and V), so that the photon path lengths may be written in terms of a longitudinal component along the z-axis. Considering a differential slab of thickness dz' emitting a flux of $\Phi_0 = 0.5\ B\ [n(z')p(z') - n_i^2]dz'$ towards the front surface (z' = 0), the fraction of photons incident on the interior surface in a differential solid angle $d\Omega$ is given by [4]:

$$d\Phi = \Phi_0 \exp(-\alpha z'/\cos\theta)\, d\Omega/2\pi = \Phi_0 \exp(-\alpha z'/\cos\theta) \sin\theta\, d\theta\, d\phi\, /2\pi \qquad (1)$$

where θ is the angle between the emission ray and the z-axis. The expression includes self-absorption through the exponential decay term, where α is the absorption coefficient. The rotational symmetry of the chosen geometry permits integration over the angular coordinate ϕ, resulting in a factor of 2π. Integrating over the angle θ by making the variable substitution $u = \alpha z'/\cos\theta$, equation 1 becomes:

$$\Phi_{incident} = \Phi_0\, \alpha z' \int_{\alpha z'}^{\infty} \frac{e^{-u}}{u^2} du = \Phi_0\, E_2(\alpha z') \qquad (2)$$

where E_2 is a member of the family of functions known as exponential integrals [5]. The equation differs from the simple exponential expression commonly used to describe absorption ($\Phi = \Phi_0\, e^{-\alpha z'}$) because the majority of photons incident on the interior surface travel further than z' (since the emission is isotropic rather than collimated), resulting in greater absorption of the total flux.

In order to provide the external PL spectrum due to radiative recombination everywhere within the device layer, the contribution from angles greater than the critical angle of reflection (θ_c) must be subtracted from equation 2, and the result multiplied by the polarization-averaged transmissivity of the interface (1 - R). Equation 2 must also be integrated over the entire device width (w), and integrated over energy [6]:

$$\Phi_{emission} = \int_0^{\infty} dE\, \hat{\gamma}(E) \int_0^{w} dz' \frac{B}{2}\left[n(z')\, p(z') - n_i^2\right] \Big\{ (1 - R_f)\left[E_2(\alpha(E)\, z') - \cos\theta_c\, E_2\left(\alpha(E)\, z'/\cos\theta_c\right)\right]$$

$$+ \sum_{k\ even,\ >0} (1 - R_f)\, R_b^{k/2}\, R_f^{k/2}\left[E_2(\alpha(E)\,[kw + z']) - \cos\theta_c\, E_2\left(\alpha(E)\,[kw + z']/\cos\theta_c\right)\right]$$

$$+ \sum_{k\ odd,\ >0} (1 - R_f) R_b^{(k+1)/2}\, R_f^{(k-1)/2}\left[E_2(\alpha(E)\,[(k+1)w - z']) - \cos\theta_c\, E_2\left(\alpha(E)\,[(k+1)w - z']/\cos\theta_c\right)\right] \Big\}$$

where $\hat{\gamma}(E)$ represents the normalized van Roosbroeck- Shockley representation of the spontaneous emission spectrum [2]. In the above equation, k is the number of reflections between interfaces before a photon either escapes or is reabsorbed. To completely account for all spontaneous emission

photons, an infinite number of reflections (k) is required. However, numerically only about 10 reflections are required to obtain a reasonable degree of accuracy, depending on the geometry.

C. Relation to Photon Recycling

In 1-D, the generation rate due to band-to-band absorption of photons is given by:

$$G(z) = -\frac{d\Phi}{dz}.$$

Using this relation on equation 2, an effective photon recycling generation rate can be computed [6]:

$$G_{pr}(z) = -\Phi_0 \frac{d}{dz} E_2(\alpha z) = \alpha \Phi_0 E_1(\alpha z)$$

which is equivalent to the general expression presented by Kuriyama, et. al. [2]. Thus, the total emission (including emission from the front surface, and loss to the substrate) can be expressed in terms of the effective photon recycling generation rate:

$$\Phi_{emission} = \int_{slab} [R_{radiative}(z) - G_{pr}(z)]\, dz$$

In our implementation, self absorption/ photon recycling is incorporated into the solution by computing the effective generation rate and adding it to the external optical generation rate in the current continuity equations. The emission spectrum is then computed as a post-processing step after each iteration so that it is self-consistent with the current solution.

III. SIMULATION RESULTS

In order to determine the quantitative effects of the minority carrier diffusivity on the observed time-dependent PL spectrum, a series of simulations were performed on a 10 μm GaAs DH with a doping of $n \sim 1.3 \times 10^{17}$ cm^{-3}. The absorption coefficients were obtained from transmission measurements on an identical series of structures with various GaAs layer thicknesses [7]. Fig. 2 shows the calculated time-dependent PL spectrum for the structure assuming a bulk nonradiative lifetime of 1.3 μs, a radiative recombination coefficient of 2.45×10^{-10} cm^3/s, and a minority hole

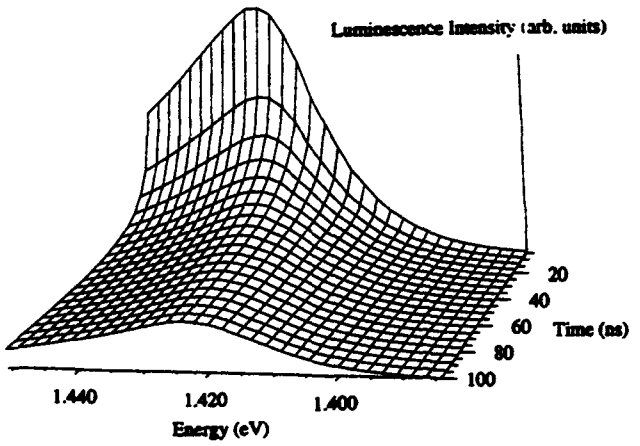


Fig. 2: Calculated time evolution of the PL spectrum of the 10 μm GaAs DH.

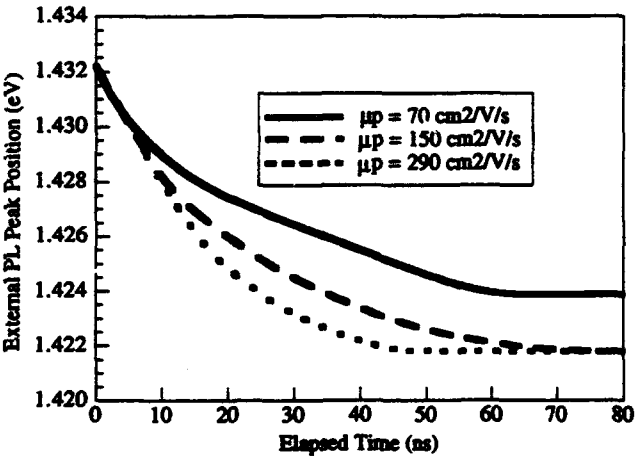


Fig. 3: Comparison of PL spectra for different values of hole mobility.

mobility of 70 cm^2/V/s. As can be seen, the calculated behavior is similar to the experimental results plotted in Fig. 1. The predicted external PL peak as a function of time is plotted in Fig. 3 for minority hole mobilities of 70, 150 and 290 cm^2/V/s. As expected, the peak energy shifts much more quickly with larger values of the mobility, as diffusion moves the excess carriers further away from the front surface before they can recombine.

Initial simulations of the measurement in Fig. 1 did not predict quite as large of a red shift, even with very large values of minority hole mobility. In order to examine this in more detail, steady state PL was performed on the same structure. The results of the best fit are plotted in Fig. 4 (as "predicted PL"), along with the experimental data. As in the TRPL simulations, the predicted external PL peak is approximately 5 meV bluer than the experimental data. However, by assuming a flat excess carrier profile, the E_2 integral expression may be integrated and used to remove the effects of self-absorption on the PL peak. Working backwards in such an approximate fashion results in a reasonable fit to the peak energy, also plotted in Fig. 4 (as "shifted experimental") along with the van Roosbroeck-Shockley (VRS) relation. The disagreement of the "shifted experimental data" curve with the VRS curve away from the peak suggests that either the absorption model or the emission expression may need further refinement. The lack of agreement to the lower energy side of the VRS curve, however, provides a possible explanation for the lack of agreement between the experimental and simulated PL curves, as that energy region overlaps the experimental PL peak.

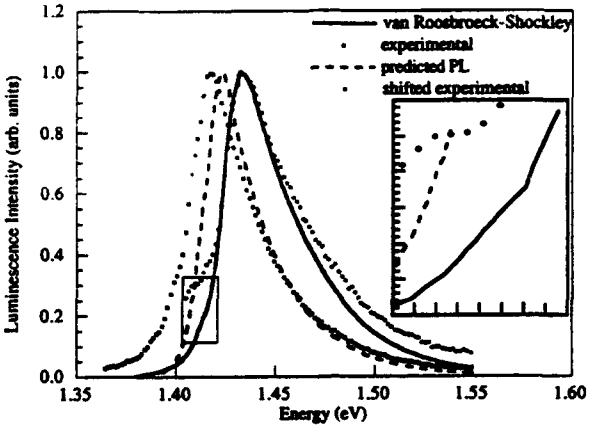


Fig. 4: Experimental steady-state PL of the 10 μm DH. The highlighted region shows where the absorption model accuracy leads to difficulty in predicting external PL.

IV. SUMMARY

A detailed numerical model for computing the emission spectrum in time-resolved photoluminescence experiments has been incorporated into a 1-D semiconductor simulation package. The calculation of the photoluminescence (PL) is performed self-consistently with calculation of the photon recycling generation rate, thereby automatically accounting for the effects of self-absorption on the externally observed emission spectrum. While detailed modeling of the decay of PL intensity provides information with regard to recombination processes, the temporal behavior of the PL spectrum depends on the minority carrier diffusivity as well. Thus, detailed modeling of the TRPL experiments can provide information about both minority carrier transport and recombination.

REFERENCES

[1] R.K. Ahrenkiel, D.J. Dunlavy and T. Hanak, J. of Appl. Phys. 64, 1916 (1988).
[2] T. Kuriyama, T. Kamiya and H. Yanai, Jpn. J. Appl. Phys. 16, 465 (1977).
[3] S.M. Durbin and J.L. Gray, IEEE Trans. Elec. Dev. 41, 239 (1994).
[4] E. Velmre, B. Freiden and A. Udal, Phys. Scr. 24, 468 (1981).
[5] W. Gautschi and W.F. Cahill, in Handbook of Mathematical Functions, M. Abramowitz and I.A. Stegun, eds., Dover Publications, Inc., New York, 227 (1970).
[6] S.M. Durbin, J.L. Gray, M.P. Patkar and M.S. Lundstrom, Conf. Rec. OE/Lase '94, SPIE 2146.
[7] G. B. Lush, Ph.D. Thesis, Purdue University (1992).

The computational development work was supported by the NASA Graduate Research Fellowship Program (author S.M.D.). The research at NREL was supported by the Department of Energy under contract number DE-AC02-83CH10093.

TIME-DEPENDENT SIMULATIONS OF FILAMENT PROPAGATION IN PHOTOCONDUCTING SWITCHES

P. W. Rambo , W. S. Lawson,
*University of California, Lawrence Livermore National Laboratory**
P. O. Box 808, Livermore, CA 94551
C. D. Capps and R. A. Falk
Boeing Defense & Space Group
P. O. Box 3999, Seattle, WA 98124

ABSTRACT

We present a model for investigating filamentary structures observed in laser-triggered photoswitches. Our model simulates electrons and holes in two-dimensional cylindrical (r-z) geometry, with realistic electron and hole mobilities and field dependent impact ionization. Because of the large range of spatial and temporal scales to be resolved, we are using an explicit approach with fast, direct solution of the field equation. A flux limiting scheme is employed to avoid the time-step constraint due to the short time for resistive relaxation in the high density filament. Self-consistent filament propagation with speeds greater than the carrier drift velocity are observed in agreement with experiments.

I. INTRODUCTION

Laser-triggered solid-state switches operating at high fields have promising applications to fast pulse-power technology, to microwave generation, and may play an important role in impulse radar. Experiments with semi-insulating GaAs switches have shown exceptionally fast low-jitter operation in the high-gain regime, where the laser trigger energy is small compared to the switched electrical energy; see e.g. [1]. Connected with this efficient high-gain operation is the experimental observation of filamentary structures, which are suspected of carrying the bulk of the device current [2]. An understanding and characterization of these filaments is important for improving device performance and lifetime, and is the primary goal of our simulations.

Our model simulates electrons and holes in two-dimensional cylindrical (r-z) geometry. The continuity equation for each species s, with density n_s, is advanced in time with the particle flux expressed through a drift-diffusion type relation:

$$\frac{\partial n_s}{\partial t} = -\nabla \bullet (n_s \mathbf{u}_s) + \sum_{s'} \alpha_{s'} |\mathbf{u}_{s'}| n_{s'}, \quad \mathbf{u}_s \equiv \frac{q_s}{e} \mu_s \mathbf{E}$$

Realistic coefficients for mobility μ_s and impact ionization α_s are included, both non-linearly dependent on the electric field. For the time and space scales of interest, both diffusion and recombination are unimportant, but could be easily included. In fact, the physical diffusion would be small relative to numerical diffusion. The electrostatic field is found from the Poisson equation (in CGS units),

$$\nabla^2 \Phi = -\frac{4\pi}{\varepsilon} \sum_s q_s n_s, \quad \mathbf{E} = -\nabla \Phi$$

where ε is the crystal dielectric (assumed to be a constant), and the charge of a species is q_s ($-e$ and $+e$ for electrons and holes respectively).

Our strategy for numerical solution of this model is determined by the necessary time and space scales that must be resolved. Experimentally, these filaments are observed to have a radius on the order of tens of microns, thus we anticipate a typical cell size $\Delta r \approx \Delta z < 1\mu m$. The switches are typically one to several millimeters in dimension, hence our system sizes will be many hundreds of zones on a side and

total number of grid cells $N_g \approx 10^5$-10^6. The relevant time scale to be resolved is the ionization time, $\Delta t < (\mu\alpha)^{-1} \approx 0.5$ ps. Switching times are of the order of 100 ps - 1 ns, so we can anticipate several thousand cycles necessary for a simulation.

II. NUMERICAL IMPLEMENTATION

Apart from accuracy, ie. properly resolving the phenomena of interest, there are stability restrictions on the time step, determined by the choice of numerical algorithms. The Courant limit on the continuity equation requires that the flow not transit a cell in less than one time step (similar arguments would lead to a time step restriction for explicit diffusion). For the spatial zoning typically employed in these simulations, however, neither of these conditions limits the time step beyond what is required for resolving the phenomena of interest. A more limiting constraint is due to the resistive relaxation time. A region of space charge tries to relax under the action of its self-consistent electrostatic field; if the time step is too large, the motion over-compensates generating an oppositely directed field of larger magnitude. The stability requirement for explicit differencing is that the resistive decay time be resolved,

$$\Delta t < \tau_r \quad , \quad \tau_r \equiv \frac{\varepsilon}{4\pi e \sum_s n_s \mu_s}$$

For intrinsic GaAs at 300 K ($\mu \approx 5.3\times10^3$ cm^2/V-s), the resistive decay time is approximately $\tau_r \approx (1300\ \Omega/\text{cm}^{-3})/n_e$. Because of nonlinear mobilities, $\Delta t > \tau_r$ may be acceptable due to nonlinear saturation of the unstable oscillations. However, in the high density, low-field region of the filament this limit can be a serious constraint. Typical carrier densities in the filament are found to be $n \approx 10^{17}$-10^{18} cm^{-3}, and so the resistive decay time is very small, $\tau_r < 10$ fs.

Typical semiconductor device simulation is concerned with steady state behavior and solution schemes use implicit methods, where all the terms on the right-hand side of the continuity equation are written at the advanced time. This requires the solution of large block matrix systems, which are iterated due to non-linearities in the equations. These methods are robust and efficient at finding final steady states, but excessively costly for our application. An alternative is to treat only the most troublesome term by using the advanced electric field to define the drift velocity; substituting into the Poisson equation determines the field equation for a non-iterative scheme. However, this non-separable field equation still precludes the use of very efficient rapid-elliptic-solvers, which are the preferred solution method because of the large system size and necessary number of cycles.

We have implemented a *flux limiting* scheme which still allows use of a direct, rapid solver for the Poisson equation. The drift term is defined by limiting the velocity compared to a simple explicit definition,

$$n_s^{n+1} = n_s' - \Delta t\, \nabla \bullet \left(n_s' \frac{q_s}{e} \frac{\mu_s^n \mathbf{E}^n}{(1+\Delta t/\tau_r)} \right), \quad n_s' = n_s^n + \Delta t \sum_s \alpha_s^n \left|\mathbf{u}_s^n\right| n_s^n$$

where superscripts denote time level and the primed density on the right hand side includes other terms advanced explicitly in time. In regions where the resistive decay time is well resolved (low density, low mobility), $\Delta t << \tau_r$, the flux takes on its usual value. In regions where $\Delta t >> \tau_r$, (high density, high mobility) this flux limiting ensures that only enough density moves to shield out electric field fluctuations.

Currently, spatial differencing of the transport is fully upwind; with the diffusion neglected, this is equivalent to Scharfetter-Gummel [3] differencing in the (appropriate) high field limit. The problem domain is restricted to be "rectangular", ie. $0 < z < Z_{max}$ and $0 < r < R_{max}$. The zoning is logical rectilinear, but can be variable in space, specified at problem set up. We have implemented two choices for the field solution, a fast Fourier transform method (FFT) and a cyclic reduction (CR) package. Although restricted

to uniform zoning in the z direction, the FFT package is substantially faster on vector machines like the CRAY Y-MP. The left and right boundaries for z are electrodes, and so defined to be equipotential surfaces. The system potential may be fixed or determined by an external circuit. At r=0, the radial electric field E_r must vanish in the absence of a line charge on axis; similarly E_r =0 at $r = R_{max}$ to approximate a transition to plane parallel equipotential surfaces.

III. REPRESENTATIVE RESULTS

To demonstrate the utility of our model, we present representative results from simulations of filament propagation. The system is $0<z<Z_{max}$=256 μm, and $0<r<R_{max}$ =100 μm. Zoning in the z-direction is constant, Δz =0.50 μm; there are 51 radial zones, with Δr=0.50 μm at the axis and monotonically increasing to the radial boundary. This small system size and coarse zoning was picked to allow a calculation with small enough time step to avoid instability without flux limiting. Material parameters are typical for GaAs, except that the mobility model used here is monotonic to avoid complications from NDR and Gunn domain formation. A uniform background of electrons and holes representing the photo-generated carriers is initialized with a density of $n_0 = 1\times10^{14}$ cm^{-3}. To initiate the filament, a high density needle is initialized at the left hand side of the system, with radius 10 μm, length 60 μm, and density 1×10^{17} cm^{-3}. The system is initially charge neutral, so that at time t=0, the electric field is uniform with E_0=100 kV/cm (negative z-direction); the system potential is fixed in time. In the first few picoseconds of the simulation, the electric field is excluded from the needle with a large enhancement developing at the tip, which eventually breaks down. The filament then propagates, with the density and tip shape determined self-consistently. A contour plot of electron density at t=140 ps (Fig. 1a) clearly shows the initial perturbation (z<60 μm), a transition region (60 μm<z<80 μm), and the propagating filament (z>80 μm); contours of potential are shown in Fig. 1b. As the filament transits the device, the potential drop occurs over a smaller distance between filament head and anode increasing the field strength which leads to increasing filament radius and density. This qualitative effect of increasing

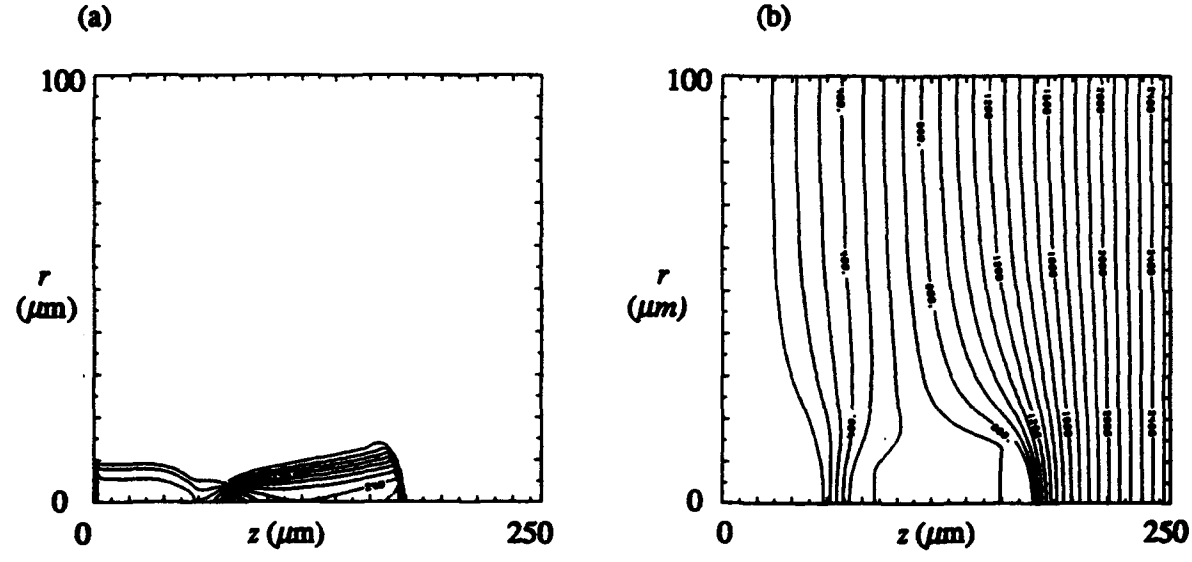


Fig. 1 Flux limited filament calculation at t=140 ps, showing (a) contours of electron density, and (b) contours of electric potential; time step is $\Delta t = 5.0\times10^{-14}$ s.

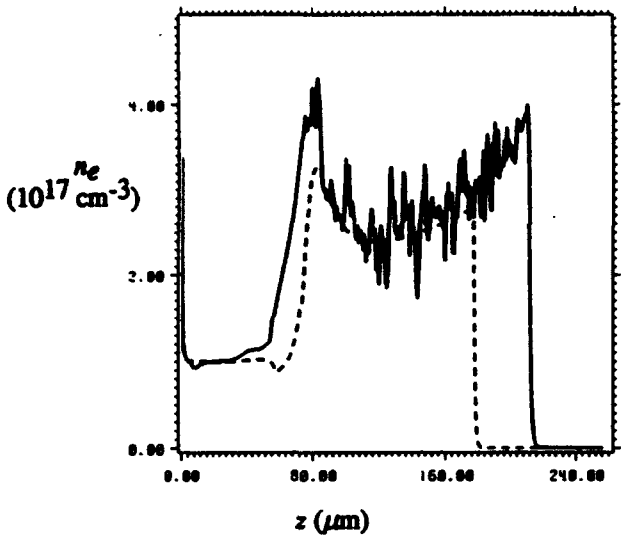


Fig. 2 Axial electron density $n_e(z)$ at time t=140 ps; solid line is without flux limiting, dashed is flux limited (identical to Fig. 1), Δt=5.0×10^{-14} s.

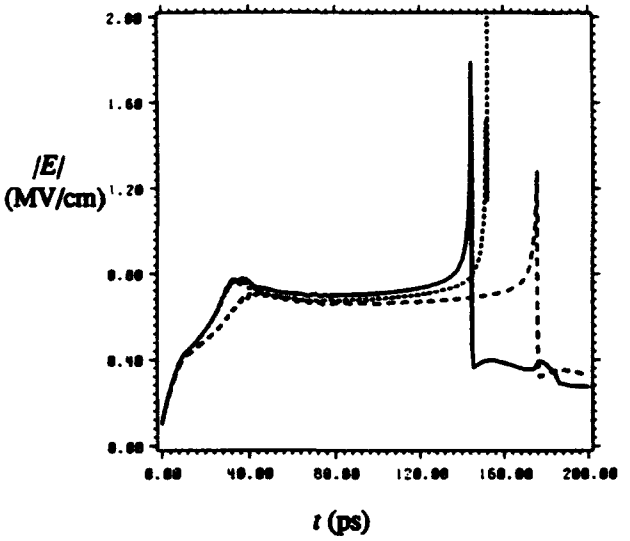


Fig. 3 Magnitude of electric field at filament head versus time: solid, Δt=2.5×10^{-15} (no flux limit); dashed (flux limited) and dotted (no flux limit), Δt=5.0×10^{-14} s.

radius is missing in previous filament simulations which mocked up the field enhancement in one dimensional calculations and picked the filament radius as a parameter [4].

This calculation was performed with flux limiting and a time step of Δt=5.0×10^{-14} s. An otherwise identical calculation but without flux limiting is compared in Fig. 2a, which shows the axial (r=0) profile of electron density n_e at time t=140 ps; unstable oscillations are obvious. The gross features of the two simulations are quite similar, except for a somewhat smaller propagation speed in the flux limited case. The calculation without flux limiting might be considered satisfactory, except that as the filament finally completes its transit across the system, instability terminates the calculation and prevents the subsequent determination of the switched current. A comparison of the previous two calculations with one using a time step Δt=2.5×10^{-15} s and no flux limiting is presented in Fig. 2b, which shows the magnitude of the electric field at the filament head versus time.

* This work was supported by the U. S. Department of Energy at Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

[1] F. J. Zutavern et al., "Photoconductive Semiconductor Switch Experiments for Pulsed Power Applications," IEEE Trans. on Electron Devices ED-37, 2472 (1990).
[2] J. C. Adams, R. A. Falk, C. D. Capps, and G. Bohnhoff-Hlavacek, "Characterization of Current Filamentation in GaAs Photoconductive Switches," Proc. SPIE 1873, 10 (1993).
[3] D. L. Scharfetter and H. K. Gummel, "Large-Signal Analysis of a Silicon Read Diode Oscillator," IEEE Trans. on Electron Devices ED-16, 64 (1969).
[4] C. D. Capps, R. A. Falk, and J. C. Adams, "Time-dependent model of an optically triggered GaAs switch," J. Appl. Phys. 74, 6645 (1993); D. W. Bailey, R. A. Dougal, J. L. Hudgins, "Streamer propagation model for high-gain photoconductive switching," Proc. SPIE 1873, 185 (1993).

AUTHOR INDEX

M. A. Alam, 298

M. G. Ancona, 256

Y. Apanovich, 107

V. Aristov, 191, 199

R. Baca, 318

B. Bagheri, 260

N. Bannov, 191, 199

D. M. Barry, 266, 270

W. Batty, 131

K. Bhattacharyya, 288

P. Blakey, 107

P. Bordone, 175

F. A. Buot, 183

J. Cai, 127

Y. Cai, 314

C. D. Capps, 327

G. F. Carey, 248

M. Chen, 215

L. A. Coldren, 292

S. W. Corzine, 292

R. Cottle, 107

W. M. Coughran, Jr., 247

H. L. Cui, 127

S. Datta, 157

J. Denavit, 33

C. R. DeVore, 256

R. Drury, 78

A. Duncan, 37

S. M. Durbin, 323

R. W. Dutton, 82, 99

R. Engelmann, 310, 314

W. L. Engl, 45

T. Ezaki, 163

M. Fadel, 49

R. A. Falk, 327

E. Fatemi, 135

C. L. Fernando, 235

D. K. Ferry, 74, 175

W. Fichtner, 69

J. Fogarty, 306

W. R. Frensley, 235

K. Furuya, 207

A. Galick, 149

R. Gaška, 25

M. Gault, 207

T. J. Godin, 243

N. Goldsman, 61, 70, 274

T. González, 29

S. M. Goodnick, 231, 288

T. R. Govindan, 179

P. Graf, 45

J. L. Gray, 323

E. Grosse, 247

H. L. Grubin, 103, 179

M. Grupen, 318, 319, 382

V. Gružinskis, 95

K. K. Gullapalli, 171

T. K. Gustafson, 203

S. Haas, 278

C. Hamaguchi, 19, 163

X. Han, 270

H. K. Harbury, 153

R. Haydock, 243

K. Hess, 302, 318, 319

J. Higman, 1

L. Hlou, 15, 49, 53

N. J. M. Horing, 127

P. Houlet, 15, 49, 53

M. J. Howes, 270

C. Huster, 11

M. K. Ieong, 111

A. Ilin, 260

S. Jallepalli, 91

J. W. Jerome, 252

R. P. Joshi, 57

C. Jungemann, 45

Y. Kamakura, 19

E. C. Kan, 82, 99

J. A. Kenrow, 203

T. Kerkhoven, 149

R. Khoie, 119

S. K. Kirby, 145

G. Klimeck, 157

P. Kocevar, 49

W. Kong, 306

G. A. Kosinovsky, 319

J. P. Kreskovsky, 103

A. M. Kriman, 57

S. Krishnamurthy, 91

M. Krowne, 187

T. Kuhn, 29, 278

T. Kunikiyo, 19

R. Lake, 157, 239

W. S. Lawson, 41, 327

C. H. Lee, 37

Y. Lee, 15

X. L. Lei, 127

M. Leng, 139, 223

C. S. Lent, 139, 219, 223, 227

E. Lenzing, 127

D. H. Levi, 323

Q. Lin, 70

P. Lugli, 7

M. S. Lundstrom, 11, 298

E. Lyumkus, 107

M. Macucci, 149

R. Madabhushi, 70

H. Matsuura, 207

P. Mawby, 207

I. D. Mayergoyz, 70

C. M. Maziar, 91

T. C. McGill, 145

M. J. McLennan, 157

R. Mickevičius, 25, 195

V. Mitin, 25, 95, 191, 195, 199

T. K. Monson, 65

N. Mori, 163

M. Morifugi, 19

C. G. Morton, 131

J. Nam, 115

D. P. Neikirk, 171

N. Nintunze, 284

J. P. Nougier, 15, 49, 53

M. A. Osman, 284

A. L. Pardhanani, 248

D. Pardo, 29

S. S. Pennathur, 288

B. S. Perlman, 127

B. Polsky, 107

W. Porod, 153, 215, 219

W. Pötz, 167

D. L. Pulfrey, 123

A. K. Rajagopal, 183

P. W. Rambo, 33, 327

U. Ravaioli, 37, 149, 163

L. Reggiani, 29, 49, 53, 95

A. Rein, 7

C. Ringhofer, 87

F. Rossi, 278

L. Rota, 231

M. Saraniti, 7

J. W. Scott, 292

L. R. Scott, 260

S. Searles, 123

Z. Shao, 219

P. Shiktorov, 53, 95

C. W. Shu, 252

A. Shur, 107

R. K. Smith, 153

C. M. Snowden, 78, 131, 266

L. So, 82

R. Solanki, 306

K. Sonoda, 19

J. P. Stanley, 274

E. Starikov, 53, 95

M. A. Stroscio, 25, 199

S. Subramanian, 211

G. C. Tai, 70

H. Takenaka, 19

T. W. Tang, 111, 115

K. Taniguchi, 19, 163

A. F. Tasch, Jr., 91

A. Tcherniaev, 107

B. J. Thibeault, 292

R. Thoma, 45

D. Z. Y. Ting, 145

P. D. Tougaw, 139, 227

M. S. Towers, 207

C. S. Tsang-Ping, 266

J. C. Vaissiere, 15, 49, 53

L. Varani, 15, 29, 49, 53, 95

D. Vasileska-Kafedziska, 175

J. A. Van Vechten, 65

P. Vogl, 7

J. F. Wager, 288

X. L. Wang, 91

J. Wood, 131

Y. J. Wu, 61

D. Xie, 260

M. Yamaji, 19

C. F. Yeap, 91

Z. Yu, 82

G. Zandler, 7

Q. S. Zhang, 65

C. Zhao, 310

J. R. Zhou, 74

M. Žiger, 167

G. Zylka, 45